



UNIVERSITAT DE
BARCELONA

First principles evaluation of structure and properties of multifunctional materials

Raul Santiago Piera

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Memòria presentada per

Raul Santiago Piera

Per optar al grau de Doctor per la

Universitat de Barcelona

Programa de doctorat en Química Teòrica i

Modelització Computacional

**First principles evaluation of structure and properties of
multifunctional materials**

Dirigida per:

Prof. Mercè Deumal Solé

Universitat de Barcelona

Dr. Jordi Ribas Ariño

Universitat de Barcelona

Tutor:

Dr. Jordi Ribas Ariño

Universitat de Barcelona



UNIVERSITAT DE
BARCELONA

Barcelona, 2023

A la meua família ...

Abstract

Organic chemistry is recognized for its exceptional ability to arrange a limited assortment of atoms in numerous ways, a quality that is crucial in molecular magnetism. This branch leverages the adaptability of organic chemistry to conceive a plethora of radicals, i.e., open-shell molecules with unpaired electrons. These organic radicals are sought for their potential applications in data storage and spintronics, leading the way towards miniaturized, lightweight and environmentally friendly solutions. However, challenges remain, particularly regarding the stability and reliability of these radicals together with their performance and resilience under the effect of temperature. It thus follows that the successful integration of these radicals into technological innovations requires progress in their synthesis, characterization, and manipulation to better suit real-world applications. Yet, predicting the capabilities of these radicals is far from straightforward. Even minor alterations in their molecular structure can have significant impact, affecting both their electronic configuration and their arrangement in crystals. Consequently, there is a high demand for the strategic design of these radicals and a more rapid evaluation of their properties to fully harness their potential in practical applications.

Our research has encompassed three distinct but interrelated topics, all contributing to the understanding of open-shell molecules and materials. Central to this PhD thesis is the exploration of magnetic properties of these materials using computational chemistry techniques. A significant part of the research involves examining how triarylmethyl-based 2D Covalent Organic Radical Frameworks react to external mechanical stimuli. This investigation has demonstrated the

efficiency of these frameworks in transitioning between various electronic states, each exhibiting distinct properties and characteristics. In the second study, we make use of quantum mechanical insights and qualitative models to design and characterize a new set of diradicals displaying a robust triplet ground state. Notably, we employ pentalene-based couplers as building-blocks, helping to pivot the role of pentalene from mere academic curiosities to potentially viable candidates for high throughput applications. Lastly, the third study consists of the development of a new quantum-informed descriptor for Machine Learning applications in chemistry, referred to as Molecular Orbital Decomposition and Aggregation (MODA). In this regard, we prove that MODA excels in predicting magnetic exchange couplings with respect to other well-established descriptors such as Bag of Bonds and Smooth Overlap of Atomic Positions.

Overall, the interplay among the three facets provides viable prospects, underscoring the potential of organic radicals in advancing molecular magnetism and paving the way for innovative applications in data storage and spintronics.

Agraïments

En primer lloc, Mercè, Jordi, moltes gràcies en majúscules. Estic convençut que tots els consells, paciència, coneixements i riures que m'heu aportat durant aquests anys no només m'han permès iniciar-me, conèixer-me i créixer com a científic, sinó que també m'han fet ser millor persona (us he volgut dedicar-ho en una frase ben enrevesada per no perdre el costum de l'escriptura d'aquesta tesi). També m'agradaria agrair a l'Stefan i al Ibério, els vostres punts de vista i les discussions científiques que hem tingut aquests anys han estat crítics per molts dels projectes en els quals hem participat conjuntament. Encara pel departament, Sergi i Maria, gràcies pel suport durant aquesta última etapa. Feu una combinació increïble.

M'agradaria també tenir unes paraules per l'Oriol, el Jordi, i el Genís. Vau ser uns companys de màster i de tesi immillorables! En la mateixa línia, Martí, Lorena, Marc, Berta, Santiago, Laia i Cristina, gràcies per tots els moments i les bones festes que ens hem marcat. També vull destacar alguns dels pilars que han estat essencials d'aquest viatge. Toni, Assut i Samu, no només heu sigut uns companys de pis increïbles, sinó que ja sou uns amics per tota la vida. Alberto i Sergi, sou un parell de personatges, però us estimo fort. Sònia, Júlia i Marc, gràcies per donar-me aquests espais on puc ser jo mateix tan obertament (para bien o para mal). Adrià i Marta, la vostra passió per la vida i riure m'ha inspirat i motivat constantment. Finalment, Pablo i Laura, el vostre suport i la vostra amistat han sigut de les coses més preciades que m'emporto d'aquests anys. Estic segur que us espera un futur junts immillorable.

Vull també agrair als meus amics de tota la vida: Miquel, Ignasi, Xavi, César, Judit i Arnau, sou una preciosa constant a la meua vida, i les nostres històries i moments compartits són un tresor ple de records que guardo amb clau. La vostra influència en mi va més enllà de les paraules, i sou una part fonamental de la meua història. Gràcies per ser-hi sempre. Continuant amb les persones que hi han sigut sempre, Papa, Mama, Mònica, anda que no us ho he posat difícil a vegades, eh... m'heu donat uns bons valors que guien com soc i que són els fonaments de tot el que he aconseguit. Sou la meua font de força. Sense la vostra confiança i amor aquest camí no hauria estat possible. Us estimo moltíssim i us estaré sempre agraït per tot.

Finalment, Judit, encara que quan ens vam conèixer em vas preguntar si "a part de fer la tesi tenia alguna altra feina", la teua capacitat per comprendre les meves inquietuds, les meves alegries i les meves pors han estat un regal que valoro profundament. De fet, em fa mal la cara de somriure quan hi penso. La complicitat mútua i el suport que compartim fan que formem un equipazo. Gràcies per ser-hi i per molts i molts anys més.

Table of Contents

1. Introduction	1
1.1. Gombert's discovery: an accidental leap into modern chemistry	3
1.2. Present and future of radicals: innovation and applications	4
1.3. Some current challenges and overview of the thesis objectives	5
1.3.1. Modulating the properties of 2D Covalent Organic Radical Frameworks through external mechanical stimuli	6
1.3.2. Design of organic diradicals with ferromagnetic interactions	8
1.3.3. Machine Learning acceleration of magnetic coupling predictions	10
1.4. Structure of the upcoming chapters	12
References	13
2. Methodology	17
2.1. Overview of methodological approaches	19
2.2. Calculation of magnetic exchange interactions	20
2.2.1. Heisenberg-Dirac-van Vleck Hamiltonian	20
2.2.2. Difficulties on the evaluation of magnetic exchange coupling J_{AB}	24
2.2.2.1. Broken-symmetry: a mono-determinantal approximation	26
2.2.2.2. Configuration Interaction: a multi-determinantal approach	28
2.3. Assessment of the (anti)aromatic character	31
2.3.1. Nucleus-Independent Chemical Shift	33
2.4. Solid-state calculations	36
2.5. Evaluation of the conductivity via Hubbard model	38
2.5.1. Role of t/U in determining the conducting/insulating character	40
2.5.2. Extraction of t and U from hybrid-DFT band structure	41
2.6. Calculation of Young's modulus to assess the effect of mechanical stimuli on extended systems	43
2.7. Ab Initio Molecular Dynamics	45
2.8. Machine Learning methods	47
2.8.1. Kernel Ridge Regression	48
2.8.2. Leave-p-groups-out Cross-Validation	53
2.9. Other aspects related to the methodology	56
2.10. Codes and methodology scope	56
References	57

3. Tuning the electronic properties of 2D-CORFs by means of external stimuli	63
3.1. Introduction	65
3.2. Results and discussion	71
3.2.1. Out-of-plane compression in TAM-based 2D-CORFs	73
3.2.1.1. Structural distortions upon compressive stress	73
3.2.1.2. Electronic structure response upon compression	81
3.2.2. In-plane strain in TAM-based 2D-CORFs	88
3.2.2.1. Structural distortions upon progressive strain	90
3.2.2.2. Electronic structure response upon strain	93
3.3. Conclusions	97
References	98
Publication #1. 2D Hexagonal Covalent Organic Radical Frameworks as tunable correlated electron systems	103
Publication #2. Controlling pairing of π -conjugated electrons in 2D covalent organic radical frameworks via in-plane strain	115
4. Design of antiaromatic organic diradicals with robust triplet ground state	127
4.1. Introduction	129
4.1.1. Insights from the electronic structure point of view	130
4.1.2. Qualitative models for ground-state properties of diradicals	135
4.1.3. Potential candidates: pentalene-based diradicals	140
4.2. Results and discussion	149
4.2.1. Prediction of the ground-state multiplicity of R ₂ -PI isomers	149
4.2.2. Qualitative prediction of the ΔE_{ST} gap of R ₂ -(DA)DPA	151
4.2.3. Quantitative prediction of the ΔE_{ST} gap of R ₂ -(DA)DPA	155
4.2.4. Assessment of the (anti)aromatic character of R ₂ -(DA)DPA	158
4.2.4.1. NICS and BLA results of My ₂ -(DA)DPA[0]	158
4.2.4.2. Extension of NICS and BLA results on R ₂ -(DA)DPA[<i>n</i>] series	163
4.2.4.3. Assessment of the thermodynamic stabilization energy	167
4.2.5. Properties of dibenzoannulated pentalene diradicals	168
4.3. Conclusions	172
References	174
Publication #3. Rational design of organic diradicals with robust high-spin ground state based on antiaromatic building blocks	181
Supporting Information. Rational design of organic diradicals with robust high-spin ground state based on antiaromatic building blocks	209

5. A new quantum-informed representation for ML applications in chemistry	231
5.1. Introduction	233
5.1.1. Selection of the dataset	237
5.1.2. Fundamental ingredients to learn chemical properties: chemical descriptors	240
5.1.2.1. Bag of Bonds	244
5.1.2.2. Smooth Overlap of Atomic Positions	245
5.2. Results and discussion	247
5.2.1. Preliminary results and model’s transferability	247
5.2.2. Decoupling of intra- and intermolecular components	252
5.2.2.1. Decoupling of BoB	253
5.2.2.2. Decoupling of SOAP	253
5.2.2.3. Automatic detection of moieties in chemical systems	254
5.2.2.4. Modification of the RBF kernel for decoupled descriptors	256
5.2.2.5. Analysis of decoupled SOAP in TTTA dataset	258
5.2.3. Derivation of Molecular Orbital Decomposition and Aggregation	262
5.2.4. Intra-/intermolecular decoupling strategy in MODA	266
5.2.5. Comparison of decoupled CIR and QIR descriptors	267
5.3. Conclusions	272
References	273
Publication #4. Unlocking the predictive power of quantum-inspired representations for intermolecular properties in Machine Learning	279
6. Conclusions	295
7. List of publications	301

Chapter 1
Introduction

1.1. Gomberg's discovery: an accidental leap into modern chemistry

In 1900, Moses Gomberg, through an interplay of curiosity and insightful investigation, discovered the triarylmethyl (TAM) radical,¹ inaugurating a century of fervent research of open-shell molecules. While synthesizing hexaphenylethane, Gomberg discerned a surprising incongruity: the anticipated product was absent, and in its place, a stable open-shell entity was discovered. Gomberg's serendipity not just deviated from synthetic expectations, but also called the standard knowledge on stability and existence of radicals into question: How did a molecule, seemingly defiant of the known paradigms of stability, persist? How might the unpaired electron, characteristically reactive, remain stable within a molecule?

The repercussions of Gomberg's discovery extended through the 20th and 21st centuries, launching a surge of investigations into molecular magnetism. Soon, works pioneered by Thiele² and Chichibabin³ confirmed that TAM was not a mere outlier but a harbinger, prompting a shift in chemical understanding. Key concepts like spin delocalization, resonance stabilization, steric protection or molecular orbital topology emerged as fundamental pieces to further understand, design, and synthesize new stable and persistent organic radicals⁴ like the well-known PTM⁵ (perchloro-triphenylmethyl) or TEMPO⁶ ((2,2,6,6-tetramethylpiperidin-1-yl)oxidanyl). These principles have bolstered modern synthetic strategies^{7,8} and predictive models, guiding researchers in designing radicals that not only bear stability but also benefit from their unique properties toward novel applications.

1.2. Present and future of radicals: innovation and applications

Organic chemistry is indisputably renowned for its versatility in manipulating the arrangement of a limited set of atoms (primarily C, H, N, O, S, and a few others) in a myriad of ways. Molecular magnetism leveraged the versatility of organic chemistry to craft a vast amount of molecules and materials with specifically tailored and fine-tuned modulation of their properties. Moreover, as organic molecular magnetism expanded, it found resonance in the solid state,⁴ where ordered arrays of radicals facilitated a higher degree of control and predictability of magnetic properties compared to the unordered nature of radicals in the gas and liquid solutions. These structured assemblies underscore the importance of order and proximity: in these conditions, radicals are no longer isolated entities, but manifest collective magnetic and conductive properties. In addition, it has been observed that even subtle changes in molecular structure or packing can result in large differences in functionality. This is not only the perfect platform to investigate deep questions in academic settings, but it is certainly an advantage when it comes to new applications.

As such, radicals are envisaged as innovative approaches in data storage that can open new avenues for lightweight, less environmentally harmful and miniaturized solutions.⁹ The potential to toggle between magnetic states through external stimuli offers a means to store binary data in a compact format, displaying capacities previously considered out of reach. Moreover, organic materials exhibiting both multifunctional magnetic and charge transfer capabilities are ushering the development of spin-electronic devices (i.e. spintronics^{10,11}) such as capacitors,¹² spin wires,¹³ spin filters¹⁴ or spin valves.¹⁵ Although in an early stage, molecular magnets can also find relevance in quantum computing, where the control on

preparing and managing quantum spin states is crucial. In this context, radicals can be used as reliable frameworks, offering an alternative avenue for the stabilization, manipulation, tailoring and reading of quantum states.

1.3. Some current challenges and overview of the thesis objectives

Despite the optimistic landscape depicted by the former applications, challenges persist, particularly concerning the stability, reliability, efficacy or time and temperature resilience of molecular magnets. The incorporation of radicals in technology demands a thorough understanding of their behaviors and interactions, and still necessitates development in synthesizing, characterizing, and manipulating methodologies that align with practical applications. Yet, the prediction of such functionality is not straightforward because chemical changes in molecular structure can affect both the molecular electronic structure and the crystal packing of the spin-bearing units. It thus follows that the rational design of new materials with specific technological applications is an arduous task.

In this regard, computational chemistry can play a relevant role¹⁶ by (i) defining and determining structure-property relationships, (ii) anticipating the response of these materials upon external stimuli, (iii) predicting new materials by designing structural units with customized properties, and (iv) accelerating the prediction of properties, among many others. In this thesis dissertation, the former (ii)–(iv) points have been addressed and explored through three relatively different, yet interconnected topics, briefly introduced in the following sections of the present chapter and further developed on the following ones. Note that structure-property relations, i.e. point (i), are sought across all three sections.

1.3.1. Modulating the properties of 2D Covalent Organic Radical Frameworks through external mechanical stimuli

The discovery of graphene¹⁷ (see Figure 1.1a), a single layer of carbon atoms arranged in a two-dimensional honeycomb lattice, marked a breakthrough in materials science. Graphene was extracted using an astonishingly simple and yet effective technique known as mechanical exfoliation, often colloquially referred to as the “Scotch tape method”. In a nutshell, the method consists of adhesive tape to peel away layers of graphite until a single-atom-thick sheet is obtained.¹⁸

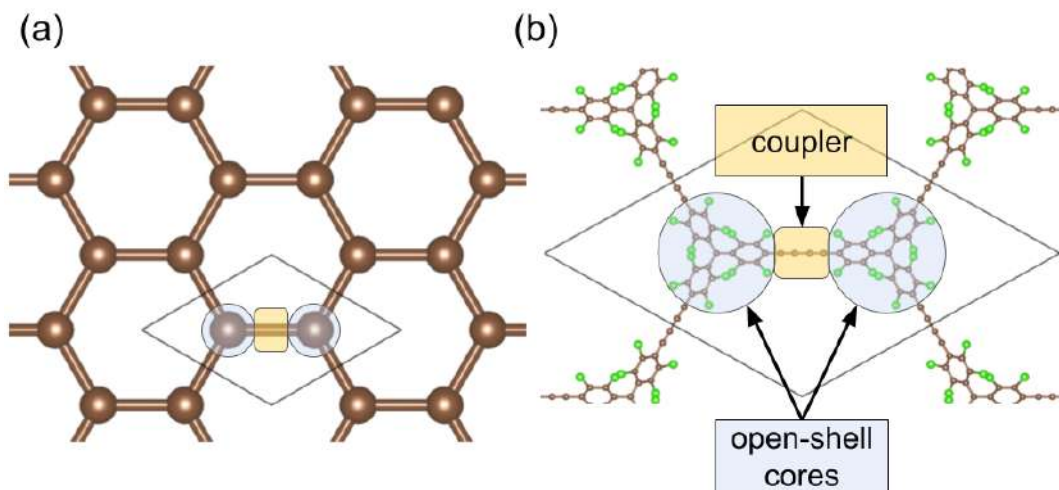


Figure 1.1: Comparative diagrams of the fundamental building blocks of (a) graphene and (b) 2D hexagonal CORFs. Open-shell cores and couplers are highlighted in blue and yellow, respectively.

Hexagonal two-dimensional Covalent Organic Radical Frameworks^{19–21} (2D-CORFs) can be perceived as a modification from the elemental structure of graphene, adopting a higher level of chemical complexity (see Figure 1.1b). Whereas graphene is a two-dimensional sheet composed of sp^2 -hybridized carbon atoms, tightly packed in a hexagonal lattice, 2D-CORFs diverge by replacing these sp^2 carbon centers with open-shell cores. Likewise, the C–C entities in

graphene are substituted by other organic couplers in 2D-CORFs, providing a novel structural paradigm.

The fundamental disparities between graphene and 2D-CORFs extend beyond mere structural differences, transcending into their respective electronic and magnetic properties. It is well-known that graphene exhibits ballistic electron transport,²² whereby electrons can traverse the material plane without scattering, due to its high electron mobility and semi-metallic nature. On the contrary, according to the recent synthesis and characterization conducted by Wu and co-workers,¹⁹ TAM-based 2D-CORFs (TAM-based 2D-CORFs) manifest notably different characteristics. The spin of open-shell TAM cores endow these materials with magnetic behavior, an aspect starkly absent in graphene. Moreover, TAM-based 2D-CORFs have been shown to exhibit only modest conductivity.

Recent studies have delved into the mechanistic explanation unfolding the distinct properties between graphene and TAM-based 2D-CORFs, where theoretical and computational studies from Bromley and co-workers²³ and Bredas and co-workers²⁴ provide pivotal insights. The different properties of graphene and TAM-based 2D-CORFs behaviors are strongly attributed to the strategic structural design of the TAM-based 2D-CORF itself. Here, the sp^2 spin-bearing carbon atoms are shielded by bulky substituents that are not merely accidental, but instrumental in assuring kinetic stability upon aggregations or reactions. As it has been demonstrated, the spin-density on the central carbon atom is critically influenced by the twist angle of the shielding phenyl units.²⁵ This deliberate structural characteristic of TAM-based 2D-CORFs comes with an intrinsic trade-off: the bulky phenyl substituents undergo a dihedral twist with respect to the 2D plane drawn by framework, which disrupts the electronic conjugation along

TAM–coupler–TAM bridges. This results in localized spins with a very small overlap between them, which are, therefore, essentially disconnected.

Strategic approaches focused on modifying this twisting angle can propel the manipulation of the CORFs' magnetic and conductive properties,²³ providing a viable pathway for tuning their electronic and magnetic multifunctional characteristics. In the initial two contributions to the thesis, this hypothesis is rigorously tested, gauging the magnetic and conductive properties of a meticulously chosen subset of TAM-based 2D-CORFs.

1.3.2. Design of organic diradicals with ferromagnetic interactions

Considering that the large majority of organic molecules have closed-shell electronic structures, molecular diradicals become exotic molecular systems in comparison. Formally, molecular diradicals are even-electron entities with two unpaired electrons, in which at least two different electronic states with different multiplicities can be found: triplet (T) and open-shell singlet (S), corresponding to the tendency of such unpaired electrons to align parallel or antiparallel, respectively.²⁶ This situation appears when Coulombic repulsion between electrons in the same molecular orbital lifts the zeroth-order degeneracy and gives rise to singlet and triplet states of different total energy. While in the existing literature many organic diradicals displaying antiferromagnetic, AFM, couplings are found (i.e. the natural tendency of the spin of two electrons to align antiparallel), organic diradicals displaying ferromagnetic, FM, couplings (i.e. the tendency to align parallel) are quite scarce.²⁷ Instead, in other related disciplines of chemistry, like in transition metal complexes, the FM interactions are the rule rather than the exception.²⁷

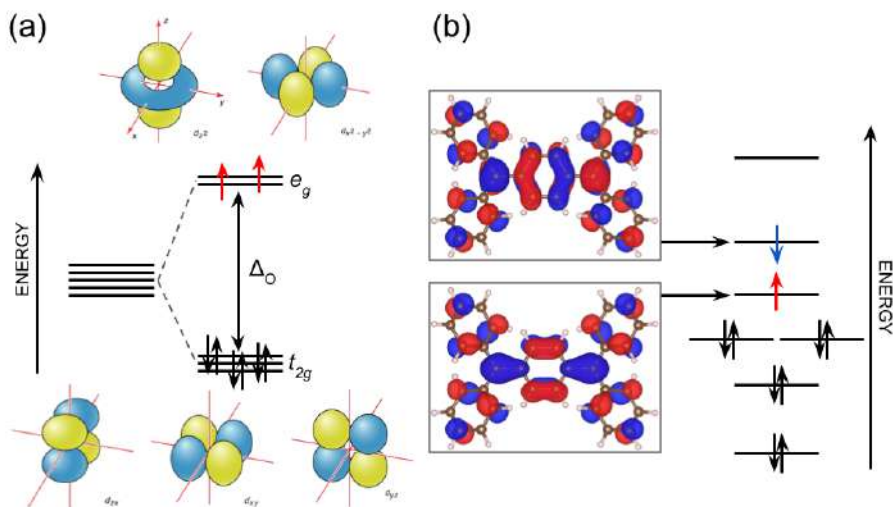


Figure 1.2: Illustration of ferromagnetic (FM) and antiferromagnetic (AFM) interactions in diradicals of (a) transition metal complexes and (b) organic diradicals.

The conditions for having diradicals with FM interactions have been widely studied and discussed.^{28,29} In a first approximation, when two singly-occupied molecular orbitals (SOMOs) are degenerate, the *Aufbau* principle and Hund's rule favors the parallel alignment of spins and, thus, a triplet ground state. Although the degeneracy of SOMOs can accidentally take place, such degeneracy is tightly related with geometrical symmetry and bonding energy between atoms across the molecular scaffold. Hence, it follows that the stark disparity between the presence of diradicals with FM/AFM interactions in transition metal complexes with respect to organic molecules is reasonable. That is, the low bonding energy in metal–ligand complexes and a high (often locally octahedral) symmetry lead to (quasi)doubly-degenerate e_g or (quasi)triply-degenerate t_{2g} orbitals, resulting in FM interactions (see Figure 1.2a). Instead, the high bonding energy and the usual low symmetry of the majority of organic molecules result in AFM interactions more often (see Figure 1.2b).

In this thesis, a key focus is placed on the design of organic diradicals with a stable triplet ground state. Drawing from the fundamental principles outlined above, we have explored strategies to achieve this goal, resulting in a new unprecedented family of organic diradicals displaying a triplet ground state.

1.3.3. Machine Learning acceleration of magnetic coupling predictions

The prediction of macroscopic magnetic properties in solids can be addressed from computational chemistry by, for example, the bottom-up approach outlined by Deumal and co-workers³⁰ (see Figure 1.3), where (i) the different spin-bearing magnetic entities across the crystal structure are identified; then, (ii) the pairwise magnetic interactions between all close A – B open-shell moieties are *euclidated* by means of the magnetic coupling, J_{AB} ; after that, (iii) collective magnetic states are obtained by full diagonalization of the associated matrix representation of an effective hamiltonian; and finally (iv) thermodynamic responses, such as heat capacity or magnetic susceptibility, are determined from statistical mechanics.³¹

However, intermolecular forces packing together molecular units in crystals³² or slight conformational variations present in CORFs²³ are particularly labile and sit below thermal energy (roughly 0.6 kcal/mol at room temperature). Thus, molecular crystals and CORFs usually undergo subtle distortions due to thermal fluctuations, which significantly alter the spin-spin interactions and can affect the observed thermodynamic response.³² Hence, it is crucial to repeatedly assess the time- and temperature-resolved evolution of the thermodynamic quantities across steps (ii)–(iv) of the bottom-up procedure until a stationary regime and/or ergodicity is observed.

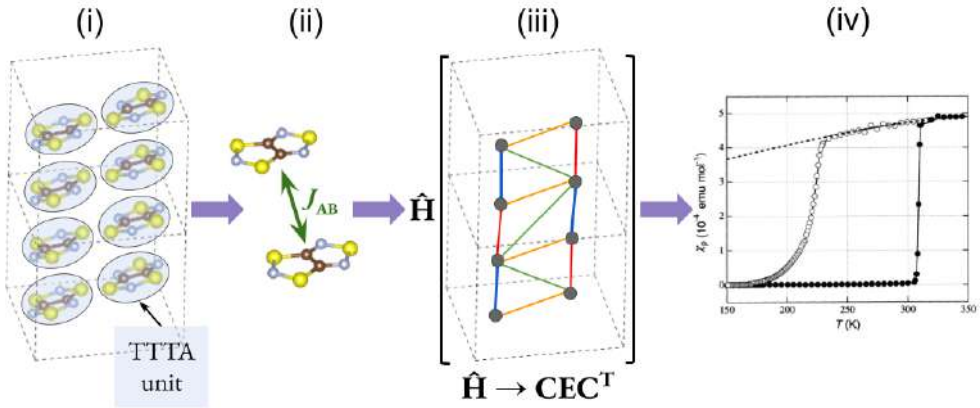


Figure 1.3: Computing thermodynamic quantities for magnetic crystals via a bottom-up approach, exemplified by the TTTA crystal. The steps (i) – (iv) align with the main text.

One of the focal points of computational complexity and bottleneck arises primarily in step (ii) of the bottom-up procedure. The calculation of the magnetic coupling, J_{AB} , is tightly related to the energy gap, $\Delta E_{ST} = E_S - E_T$, between the singlet (E_S) and the triplet (E_T) energy states arising from the coupling of two $S = 1/2$ units. Calculating E_T is typically straightforward, but calculating E_S can entail a substantial challenge, as the open-shell singlet (OSS) state cannot be represented by a single Slater determinant (i.e., the antisymmetrized combination of Hartree products). Therefore, complex multi-determinantal methods, such as Complete Active Space Self-Consistent Field³³ (CASSCF) or broken-symmetry³⁴ (BS) approaches are often required. Although for a single crystal structure such calculations can be readily performed, the whole time-resolved bottom-up process would require several hundreds or thousands of J_{AB} evaluations.

The appearance of Machine Learning models in various domains has markedly expedited computational processes.³⁵ Particularly, in the field of scientific research, technologies like DeepMind’s AlphaFold-2³⁶ exemplify the potent impact

of Machine Learning, significantly mitigating computational demands in predicting the correct protein folding, which have historically been acknowledged to be computationally taxing. This exemplifies the potential of Machine Learning to alleviate computational barriers in processes involving a large amount of data. Moreover, to leverage the full predictive power of Machine Learning, data must be converted into a structured and interpretable format. Accordingly, chemical descriptors are systematic representations that convert complex molecular and material information into a computationally manageable format.³⁷ These data representations, while simplifying the informational content, preserve the essential characteristics related to the phenomena under investigation, such as magnetic interactions.

In relation to the computational challenges outlined here, in particular referring to the calculation of J_{AB} in step (ii) and the potential of the emerging Machine Learning technology, a notable contribution of this thesis is the design of a new descriptor, streamlined to predict J_{AB} interactions. The developed descriptor incorporates critical geometric and electronic characteristics into a coherent structure, enabling precise predictive methodology compared to other existing alternatives.

1.4. Structure of the upcoming chapters

Overall, the present PhD thesis contributes to the field of open-shell organic materials, employing computational chemistry and Machine Learning to understand the electronic properties that govern spin-spin interactions in organic systems and accelerate their computation. This thesis dissertation is organized as a compendium of publications, either already published or in process of publication. It is organized in eight chapters as follows:

First, **Chapter 1** (the current chapter) has presented an overview of organic magnetism and the specific objectives targeted across the thesis. **Chapter 2** consists of the methodology section, where the aspects related to the general electronic structure calculations and analysis of chemical data are reviewed. **Chapters 3 to 5** gather the introduction, specific methodology, discussion of the results, and conclusions of the research accomplished. Specifically, **Chapter 3** encloses two studies describing the tunability of the magnetic and conductive properties in a subset of TAM-based 2D-CORFs by means of external stimuli (anisotropic in-plane strain and isotropic out-of-plane compression, corresponding to **Publications #1** and **#2**, respectively). In **Chapter 4**, the focus is shifted to the design of high-spin organic diradicals (draft of **Publication #3**), where the potential of antiaromatic moieties as building blocks to accomplish robust triplet ground states is explored. **Chapter 5** collects the presentation of the current methods and applications of Machine Learning in the present-day chemical landscape. After that, some limitations of the existing chemical descriptors with respect to the prediction of quantities rooted in molecular magnetism are discussed. Finally, Molecular Orbital Decomposition and Aggregation (MODA) is presented as an alternative descriptor to achieve high-performance predictions (**Publication #4**). In **Chapter 6**, the general conclusions of the thesis are summarized, based on the results presented in the preceding chapters. Finally, **Chapter 7** consists of a brief explanation of the contributions of the PhD candidate to the publications.

References

- (1) M. Gomberg, *Journal of the American Chemical Society*, 1900, **22**, 757–771.

- (2) J. Thiele and H. Balhorn, *Berichte der deutschen chemischen Gesellschaft*, 1904, **37**, 1463–1470.
- (3) A. E. Tschitschibabin, *Berichte der deutschen chemischen Gesellschaft*, 1907, **40**, 1810–1819.
- (4) Z. Chen, Y. Li and F. Huang, *Chem*, 2021, **7**, 288–332.
- (5) V. Diez-Cabanes, G. Seber, C. Franco, F. Bejarano, N. Crivillers, M. Mas-Torrent, J. Veciana, C. Rovira and J. Cornil, *ChemPhysChem*, 2018, **19**, 2572–2578.
- (6) S. Barriga, *Synlett*, 2001, **2001**, 0563.
- (7) R. G. Hicks, *Organic & Biomolecular Chemistry*, 2007, **5**, 1321.
- (8) I. Ratera and J. Veciana, *Chem. Soc. Rev.*, 2012, **41**, 303–349.
- (9) R. Sessoli, *Nature*, 2017, **548**, 400–401.
- (10) *Stable Radicals*, ed. R. G. Hicks, Wiley, 2010.
- (11) J. Linder and J. W. A. Robinson, *Nature Physics*, 2015, **11**, 307–315.
- (12) A. Eftekhari, L. Li and Y. Yang, *Journal of Power Sources*, 2017, **347**, 86–107.
- (13) E. A. Weiss, M. J. Ahrens, L. E. Sinks, A. V. Gusev, M. A. Ratner and M. R. Wasielewski, *Journal of the American Chemical Society*, 2004, **126**, 5577–5584.
- (14) D. Cho and J. Y. Lee, *The Journal of Physical Chemistry C*, 2023, **127**, 8256–8262.
- (15) J. Devkota, R. Geng, R. C. Subedi and T. D. Nguyen, *Advanced Functional Materials*, 2016, **26**, 3881–3898.
- (16) J. J. Novoa, M. Deumal and J. Jornet-Somoza, *Chemical Society Reviews*, 2011, **40**, 3182.
- (17) K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva and A. A. Firsov, *Science*, 2004, **306**, 666–669.
- (18) V. B. Mbayachi, E. Ndayiragije, T. Sammani, S. Taj, E. R. Mbuta and A. ullah khan, *Results in Chemistry*, 2021, **3**, 100163.
- (19) S. Wu, M. Li, H. Phan, D. Wang, T. S. Heng, J. Ding, Z. Lu and J. Wu, *Angewandte Chemie International Edition*, 2018, **57**, 8007–8011.

-
- (20) Y. Yang, C. Liu, X. Xu, Z. Meng, W. Tong, Z. Ma, C. Zhou, Y. Sun and Z. Sheng, *Polymer Chemistry*, 2018, **9**, 5499–5503.
- (21) Y. Jiang, I. Oh, S. H. Joo, O. Buyukcakir, X. Chen, S. H. Lee, M. Huang, W. K. Seong, J. H. Kim, J.-U. Rohde, S. K. Kwak, J.-W. Yoo and R. S. Ruoff, *ACS Nano*, 2019, **13**, 5251–5258.
- (22) X. Du, I. Skachko, A. Barker and E. Y. Andrei, *Nature Nanotechnology*, 2008, **3**, 491–495.
- (23) I. Alcón, D. Reta, I. de P. R. Moreira and S. T. Bromley, *Chemical Science*, 2017, **8**, 1027–1039.
- (24) S. Thomas, H. Li and J.-L. Bredas, *Advanced Materials*, 2019, **31**.
- (25) I. Alcón and S. T. Bromley, *RSC Advances*, 2015, **5**, 98593–98599.
- (26) In *The IUPAC Compendium of Chemical Terminology*, International Union of Pure and Applied Chemistry (IUPAC), 2014.
- (27) H. Iwamura, in *Advances in Physical Organic Chemistry*, Elsevier, 1990, pp. 179–253.
- (28) W. T. Borden and E. R. Davidson, *Journal of the American Chemical Society*, 1977, **99**, 4587–4594.
- (29) W. T. Borden, H. Iwamura and J. A. Berson, *Accounts of Chemical Research*, 1994, **27**, 109–116.
- (30) C. S. Clarke, J. Jornet-Somoza, F. Mota, J. J. Novoa and M. Deumal, *Journal of the American Chemical Society*, 2010, **132**, 17817–17830.
- (31) R. L. Carlin, *Magnetochemistry*, Springer, Berlin, Germany, 2011.
- (32) S. Vela, M. Deumal, M. Shiga, J. J. Novoa and J. Ribas-Arino, *Chemical Science*, 2015, **6**, 2371–2381.
- (33) P. J. Knowles and H.-J. Werner, *Chemical Physics Letters*, 1985, **115**, 259–267.
- (34) L. Noodleman, D. A. Case and A. Aizman, *Journal of the American Chemical Society*, 1988, **110**, 1001–1005.
- (35) P. P. Shinde and S. Shah, 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), IEEE, 2018.

- (36) D. T. Jones and J. M. Thornton, *Nature Methods*, 2022, **19**, 15–20.
- (37) M. F. Langer, A. Goßmann and M. Rupp, *npj Computational Materials*, 2022, **8**.

Chapter 2
Methodology

2.1. Overview of methodological approaches

This chapter provides a review of the methodologies employed across the PhD thesis across **Chapters 3 to 5**. The methodologies are classified based on their application to *molecular properties*, *solid-state properties*, or as *Machine Learning methods*, and can be found in the following order:

Molecular Properties: In **Section 2.2**, the focus is placed on the calculation of magnetic exchange interactions. This section underscores the importance of the Heisenberg-Dirac-van Vleck Hamiltonian (**Section 2.2.1**) and explores the challenges in determining singlet-triplet energy gaps and magnetic exchange couplings. Further elaboration on these topics is presented in **Section 2.2.2**, which is divided into two subsections: **Subsection 2.2.2.1** discusses the Broken Symmetry approach, while **Subsection 2.2.2.2** delves into the Configuration Interaction approach. Additionally, **Section 2.3** examines the assessment of (anti)aromatic character in molecules.

Solid-State Properties: In **Section 2.4**, an overview of methodologies associated to solid-state calculations is presented. Subsequently, **Section 2.5** introduces the Hubbard Model's application in evaluating conductivity, with **Subsections 2.5.1** and **Subsection 2.5.2** delving into specific associated parameters. After that, **Section 2.6** outlines the computational approach employed to calculate the Young's modulus in solid-state. This block concludes with **Section 2.7**, which briefly addresses some aspects related to *Ab Initio* Molecular Dynamics.

Machine Learning Methods: **Section 2.8** details the mathematical details required for the application of Machine Learning techniques to the scope of this thesis. In this regard, **Subsection 2.8.1** elaborates on the details of

the Kernel Ridge Regression model, while **Subsection 2.8.2** covers the leave-p-groups-out cross-validation scheme. Both sections emphasize the significance of these methods in the learning process, particularly utilized in the final results chapter.

Finally, **Section 2.9** outlines other methodologies not detailed in this chapter and **Section 2.10** elaborates on the commercial codes used to carry out the different calculations. This information is discussed and summarized in a table format for an easy identification.

2.2. Calculation of magnetic exchange interactions

There are several models aimed at rationalizing magneto-structural correlations in current used. Among them, the Heisenberg-Dirac-van Vleck (HDvV) model is usually employed not only due to its simplicity and interpretability, but also for its demonstrated accuracy compared with experimental results. The following sections describes the process of calculating magnetic exchange interactions from the first principles to the general utilization, covering the approximations performed during the process and a discussion about its range of applicability.

2.2.1. Heisenberg-Dirac-van Vleck Hamiltonian

The HDvV Hamiltonian, \hat{H}^{HDvV} , is a spin-only model conceived to evaluate the magnetic properties of coupled spin states between well-defined spin-carrying units in a pairwise manner.¹ It is usually written as:

$$\hat{H}^{\text{HDvV}} = -2 \sum_{A,B} J_{AB} \hat{\mathbf{S}}_A \cdot \hat{\mathbf{S}}_B \quad (2.1)$$

where J_{AB} is a scalar quantity referred to as spin exchange coupling that absorbs the complexities of the spin-spin interactions between spin-carrying units A and B . Moreover, $\hat{\mathbf{S}}_k$ in equation 2.1 is the total spin operator associated to the k -th spin-carrying unit, defined as:

$$\hat{\mathbf{S}}_k = \hat{S}_k^x + \hat{S}_k^y + \hat{S}_k^z \quad (2.2)$$

For fermionic sites with spin quantum number $s = 1/2$, e.g. electrons, the former spin-projected operators can be written in the basis of spin up, $|\uparrow\rangle \equiv (1, 0)^T$, and spin down, $|\downarrow\rangle \equiv (0, 1)^T$, states as the following matrix representations in a one-particle Fock space as:

$$\hat{S}_k^x = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \hat{S}_k^y = \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}; \hat{S}_k^z = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (2.3)$$

As could be noted from equation 2.3, the \hat{S}_k^z component is diagonal in the $\{|\uparrow\rangle, |\downarrow\rangle\}$ basis, so the former are eigenstates of \hat{S}_k^z with $m_s = \pm 1/2$. Moreover, as the HDvV model aggregates pairwise spin-spin interactions, it is worth stressing that $\hat{\mathbf{S}}_A \cdot \hat{\mathbf{S}}_B$ components must be in the basis of two-electron states. Thus, $\hat{\mathbf{S}}_A \cdot \hat{\mathbf{S}}_B$ do not correspond to a regular matrix product, but these elements are defined in a direct product space, as the summation of Kronecker products^{2,3}

$$\hat{\mathbf{S}}_A \cdot \hat{\mathbf{S}}_B = \hat{S}_A^x \otimes \hat{S}_B^x + \hat{S}_A^y \otimes \hat{S}_B^y + \hat{S}_A^z \otimes \hat{S}_B^z \quad (2.4)$$

where the Kronecker product (\otimes) between two 2×2 matrices is defined as:

$$A \equiv \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} ; \quad B \equiv \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix}$$

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{21}B \\ a_{12}B & a_{22}B \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{11}b_{21} & a_{21}b_{11} & a_{21}b_{21} \\ a_{11}b_{12} & a_{11}b_{22} & a_{21}b_{12} & a_{21}b_{22} \\ a_{12}b_{11} & a_{12}b_{21} & a_{22}b_{11} & a_{22}b_{21} \\ a_{12}b_{12} & a_{12}b_{22} & a_{22}b_{12} & a_{22}b_{22} \end{pmatrix} \quad (2.5)$$

That is, the 2×2 Hilbert space defining a single $1/2$ -spin particle has been expanded to a 4×4 two-particle space. Accordingly, the 2-vectors basis, $\{|\uparrow\rangle, |\downarrow\rangle\}$, must be expanded to a vector space of cardinality 4 as:

$$\begin{aligned} |\uparrow\uparrow\rangle &= |\uparrow\rangle \otimes |\uparrow\rangle = (1, 0)^T \otimes (1, 0)^T = (1, 0, 0, 0)^T \\ |\uparrow\downarrow\rangle &= |\uparrow\rangle \otimes |\downarrow\rangle = (1, 0)^T \otimes (0, 1)^T = (0, 1, 0, 0)^T \\ |\downarrow\uparrow\rangle &= |\downarrow\rangle \otimes |\uparrow\rangle = (0, 1)^T \otimes (1, 0)^T = (0, 0, 1, 0)^T \\ |\downarrow\downarrow\rangle &= |\downarrow\rangle \otimes |\downarrow\rangle = (0, 1)^T \otimes (0, 1)^T = (0, 0, 0, 1)^T \end{aligned} \quad (2.6)$$

consisting of all the possible arrangements of entangled two $1/2$ -spin states (still being all of them eigenstates of $\hat{S}_{AB}^z = \hat{S}_A^z \otimes \mathbb{I} + \mathbb{I} \otimes \hat{S}_B^z$). According to the previous considerations, it follows that the HDvV Hamiltonian for only two spin sites, A and B , can be written in the following matrix representation:

$$\hat{H}^{\text{HDvV}} = -2J_{AB} \hat{\mathbf{S}}_A \cdot \hat{\mathbf{S}}_B = -2J_{AB} \begin{pmatrix} +1/4 & 0 & 0 & 0 \\ 0 & -1/4 & +1/2 & 0 \\ 0 & +1/2 & -1/4 & 0 \\ 0 & 0 & 0 & +1/4 \end{pmatrix} \quad (2.7)$$

The associated energy and eigenstates of the A - B pair can be obtained analytically from the direct diagonalization of \hat{H}^{HDvV} as $\hat{H}^{\text{HDvV}} \rightarrow \sigma E \sigma^T$, yielding

$$\begin{aligned}
 |\sigma_1\rangle &= |\uparrow\uparrow\rangle & E_{\sigma_1} &= -J_{AB}/2 \\
 |\sigma_2\rangle &= |\downarrow\downarrow\rangle & E_{\sigma_2} &= -J_{AB}/2 \\
 |\sigma_3\rangle &= 1/\sqrt{2}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle) & E_{\sigma_3} &= -J_{AB}/2 \\
 |\sigma_4\rangle &= 1/\sqrt{2}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle) & E_{\sigma_4} &= +3J_{AB}/2
 \end{aligned} \tag{2.8}$$

Notably, the energies of σ_1 , σ_2 and σ_3 eigenstates are three-fold degenerate, forming a triplet state, T , with energy $E_T = E_{\sigma_1} = E_{\sigma_2} = E_{\sigma_3}$, while the σ_4 state is singly degenerate, i.e. a singlet state, S , with energy $E_S = E_{\sigma_4}$. As it can be noted from the results of equation 2.8, the absolute energy of the singlet and the triplet states depends on the sign and magnitude of J_{AB} . Thus, the difference between the triplet and the singlet energies, ΔE_{ST} , is also related to J_{AB} as:

$$\Delta E_{ST} = E_S - E_T = \frac{3}{2}J_{AB} + \frac{1}{2}J_{AB} = 2J_{AB} \tag{2.9}$$

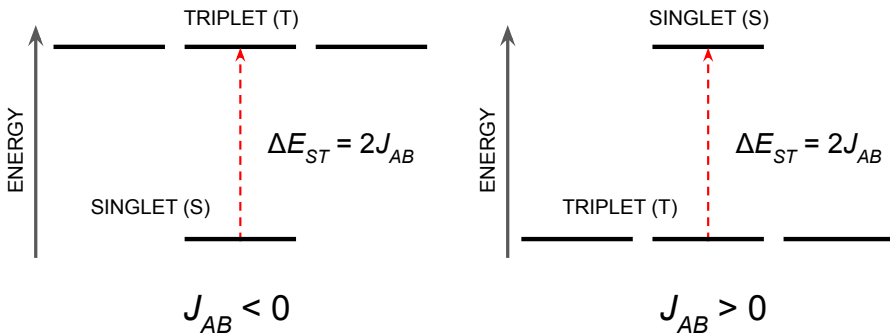


Figure 2.1: Schematic representation of the singlet-triplet energy splitting for a pair of coupled spin $1/2$ particles, ΔE_{ST} , as a function of the sign of the magnetic exchange coupling, J_{AB}

It thus follows that ΔE_{ST} is positive when $J_{AB} > 0$, that is, the triplet is the ground state and it is then said that there is a ferromagnetic (FM) interaction or coupling between the two spin centers A – B . Instead, when $J_{AB} < 0$, the singlet becomes the ground state and it is then said that there is an antiferromagnetic (AFM) interaction or coupling between the two spins (see Figure 2.1 for an schematic representation of both regimes).

2.2.2. Difficulties on the evaluation of magnetic exchange coupling J_{AB}

The utilization of the HDvV model in computational chemistry primarily stems from the results of equation 2.9. Conventional quantum-chemistry packages possess the capability to evaluate the energy of both singlet and triplet states (further discussed below), thereby approximating the total energy by solving the Schrödinger equation for a given chemical species. This capability enables the derivation of spin exchange coupling, J_{AB} , from the energy difference between triplet and singlet states as stated in equation 2.9.

However, the wavefunction of a system of electrons must satisfy the Pauli exclusion principle, which states that two electrons cannot have the same set of quantum numbers and should also adhere to the principle of indistinguishability. In this regard, a Slater determinant provides a way to construct antisymmetrized wavefunctions, meeting both previous requirements.⁴ For example, a system of two electrons, each described by a spatial-orbital, $|\chi_i\rangle$, can be expressed as

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \begin{vmatrix} \chi_1(\mathbf{r}_1) & \chi_2(\mathbf{r}_1) \\ \chi_1(\mathbf{r}_2) & \chi_2(\mathbf{r}_2) \end{vmatrix} = |\chi_1\chi_2| \quad (2.10)$$

where \mathbf{r}_i represent the spatial coordinates of each electron. The spin component can be incorporated by assuming that spin-orbitals can be constructed from the direct product of the spatial orbitals and spin components as $|\psi_i\rangle = |\chi_i\rangle \otimes |\uparrow\rangle$ and $|\bar{\psi}_i\rangle = |\chi_i\rangle \otimes |\downarrow\rangle$ for the spin up and down spin-orbitals, respectively. Thus, four valid Slater determinants for two-electron systems are $|\psi_1\psi_2\rangle$, $|\psi_1\bar{\psi}_2\rangle$, $|\bar{\psi}_1\psi_2\rangle$, $|\bar{\psi}_1\bar{\psi}_2\rangle$ (note the similarities with equation 2.6).

As it could be noted in relation to the results stated in equation 2.8, the eigenstates of $\hat{\mathbf{S}}_A \cdot \hat{\mathbf{S}}_B$ can be written in the current notation as:

$$\begin{aligned}
 |\sigma_1\rangle &= |\psi_1\psi_2\rangle \\
 |\sigma_2\rangle &= |\bar{\psi}_1\bar{\psi}_2\rangle \\
 |\sigma_3\rangle &= 1/\sqrt{2} (|\psi_1\bar{\psi}_2\rangle - |\bar{\psi}_1\psi_2\rangle) \\
 |\sigma_4\rangle &= 1/\sqrt{2} (|\psi_1\bar{\psi}_2\rangle + |\bar{\psi}_1\psi_2\rangle)
 \end{aligned}
 \tag{2.11}$$

Note that the \pm signs of states with sub-index 3 and 4 are non-intuitively inverted with respect to the signs of equation 2.8. The results from the last equation unveils that, while states $|\sigma_1\rangle$ and $|\sigma_2\rangle$ can be described by a single Slater determinant (mono-determinantal states), states $|\sigma_3\rangle$ and $|\sigma_4\rangle$ are multi-determinantal in nature. Therefore, different widely used mono-determinantal methodologies such as Hartree-Fock (HF) and Density Functional Theory (DFT) cannot be directly applied to calculate $|\sigma_3\rangle$ and $|\sigma_4\rangle$ states.

Modern quantum-chemistry packages allows to calculate the triplet state energy from the mono-determinantal $|\sigma_1\rangle$ and $|\sigma_2\rangle$ states,^{5,6} bypassing the limitations in assessing the energy from the remaining $|\sigma_3\rangle$ triplet state. However, the only existing singlet state eigenfunction of $\hat{\mathbf{S}}_A \cdot \hat{\mathbf{S}}_B$ is multi-determinantal, forc-

ing to evaluate the singlet-triplet gap, ΔE_{ST} , from either (i) performing some approximations to the mono-determinantal methods or (ii) make use of multi-determinantal methods. The following subsections cover the details of the so-called Broken-Symmetry approximation and overviews some multi-determinantal methodologies employed across the following chapters.

2.2.2.1. Broken-symmetry: a mono-determinantal approximation

The clever idea to simulate the effects of this multi-determinantal singlet state, σ_4 , without resorting to actual multi-determinantal methods is to start from a single determinant wavefunction that reflects the correct AFM state but is of the ‘wrong’ spin symmetry.^{7,8} This wavefunction is initially written as

$$|\Psi_S^{BS,init}\rangle = |(\text{core}) \dots \psi_a \bar{\psi}_b| \quad (2.12)$$

where ‘(core)’ stands as a shortcut notation to include all the doubly-occupied orbitals, while ψ_a and ψ_b are obtained from the optimized mono-determinantal triplet, i.e. $|\Psi_T\rangle = |(\text{core}) \dots \psi_1 \psi_2|$, as

$$\begin{aligned} |\psi_a\rangle &= \frac{1}{\sqrt{\lambda^2 + \mu^2}} (\lambda |\psi_1\rangle + \mu |\psi_2\rangle) \\ |\psi_b\rangle &= \frac{1}{\sqrt{\lambda^2 + \mu^2}} (\lambda |\psi_1\rangle - \mu |\psi_2\rangle) \end{aligned} \quad (2.13)$$

As accurately illustrated by Neese,⁶ the state written in equation 2.12 (with $\lambda = \mu = 1$, i.e orthonormal linear combination) is not the actual broken-symmetry determinant, but the initial guess. Then, the λ and μ parameters of the broken-symmetry initial guess are variationally and self-consistently optimized to a new set of non-equal parameters as

$$|\Psi_S^{BS}\rangle = \left| (\text{core}) \dots \psi_a^{\text{opt}} \overline{\psi_b^{\text{opt}}} \right| \quad (2.14)$$

where, importantly, ψ_a^{opt} and ψ_b^{opt} are not orthogonal, as the overlap, S_{ab} is

$$S_{ab} = \langle \psi_a^{\text{opt}} | \psi_b^{\text{opt}} \rangle = \frac{\lambda^2 - \mu^2}{\lambda^2 + \mu^2} \neq 0 \quad (2.15)$$

for any other value than $\lambda = \mu$. Now, the evaluation of the broken-symmetry Slater determinant in the spin-only HDvV Hamiltonian leads to

$$\langle \Psi_S^{BS} | \hat{H}^{\text{HDvV}} | \Psi_S^{BS} \rangle = \frac{(1 + 2S_{ab}^2)}{2} J_{AB} \quad (2.16)$$

Thus, the energy difference between the broken symmetry singlet and the triplet, ΔE_{ST}^{BS} , can be obtained

$$\Delta E_{ST}^{BS} = \frac{(1 + 2S_{ab}^2)}{2} J_{AB} + \frac{J_{AB}}{2} = (1 + S_{ab}^2) J_{AB} \quad (2.17)$$

Finally, letting the expression of equation 2.17 to be equal to equation 2.9 (via $J_{AB}^{\text{Eq. 2.17}} = J_{AB}^{\text{Eq. 2.9}}$) the following relation between the singlet-triplet gaps is obtained:

$$\Delta E_{ST} = \frac{2\Delta E_{ST}^{BS}}{1 + S_{ab}^2} \quad (2.18)$$

The result from the last expression allows to relate the mono-determinantal broken-symmetry singlet-triplet gap with the multi-determinantal gap, allowing to calculate J_{AB} via equation 2.9. Although the broken-symmetry is an approximation, its accuracy has been demonstrated in practice when used together with hybrid exchange-correlation functionals such as PBE0⁹ or B3LYP.^{10,11}

The evaluation of ΔE_{ST} and J_{AB} values by means of the Broken-Symmetry approximation has been a core tool in this PhD thesis, providing valuable insights about the magnetic interactions of the target systems across all the studies.

2.2.2.2. Configuration Interaction: a multi-determinantal approach

In this PhD dissertation, multi-determinantal approaches have been used to assess the quality of the broken-symmetry approximation in predicting singlet-triplet energy gaps, ΔE_{ST} , and J_{AB} values. The vast majority of multi-determinantal methods are based on Configuration State Functions (CSFs), which consist on a blend of symmetry and spin-adapted linear combinations of Slater determinants. Multiple CSFs can emerge from a single electronic configuration, maintaining the same overall quantum number for spin and spatial components while varying in their intermediate couplings.

In the Configuration Interaction (CI) methods,¹² the ground state wavefunction is derived from a linear combination of all Slater determinants with the correct symmetry as

$$|\Phi\rangle = C_0 |\Psi_0\rangle + \sum_{a,r} C_a^r |\Psi_a^r\rangle + \sum_{a<b} \sum_{p<q} C_{ab}^{pq} |\Psi_{ab}^{pq}\rangle + \dots \quad (2.19)$$

Here, $|\Psi_a^r\rangle$ symbolizes the singly excited CSFs, distinct from $|\Psi_0\rangle$ as it has the spin-orbital $|\psi_a\rangle$ exchanged by the spin-orbital $|\psi_r\rangle$. Similarly, $|\Psi_{ab}^{pq}\rangle$ denotes the doubly-excited CSFs and this process is carried on to incorporate all possible CSFs. The expansion coefficients (C_0 , C_a^r , C_{ab}^{pq} , etc.) are obtained through the energy minimization subjected to the variational principle.⁴ When all sets of CSFs are employed for a particular basis set, the computation is referred to as a *full*-CI, which is indeed the exact solution of the Schrödinger equation in that

given CSF basis set. However, the full-CI is computationally taxing, making this solution impractical even for relatively small molecular systems. In the following sub-sections two main alternatives to the full-CI are briefly presented: The Complete Active Space Self-Consistent Field (CASSCF) and Difference-Dedicated CI (DDCI).

2.2.2.2.1. Complete Active Space Self-Consistent Field method

The CASSCF method differs from the full-CI method in that it limits the number of Slater determinants considered to strike a balance between computational cost and accuracy.^{13,14} In the CASSCF approach, the molecular orbitals within the $|\Psi_0\rangle$ determinant are categorized into three distinct sub-spaces: (i) Inactive (core orbitals), (ii) Active, and (iii) Virtual (see Figure 2.2a). The inactive space (i) comprises the lowest-energy set of doubly-occupied molecular orbitals. The Active sub-space (ii) encompasses orbitals in which all possible occupations (0, 1 or 2) are allowed. Commonly, the orbitals included in the active space include the frontier molecular orbitals (partially-occupied in diradicals), as well as a few orbitals that lie just below and just above the frontier orbitals. In the CASSCF method, only these active orbitals are permitted to contribute to the CSF as long as the total spin is conserved. Finally, the Virtual sub-space (iii) consists of molecular orbitals that are considerably higher in energy and, as a result, are considered unoccupied in all electronic configurations.

Moreover, while in CI methods only the expansion coefficients are variationally optimized, in CASSCF method the atomic orbital coefficients within each molecular orbital are optimized as well¹⁵ (i.e., c_{ij} in a typical linear combination of atomic orbitals scheme). The optimization of both the atomic orbital and expansion coefficients are proven to result in more accurate total energies. Fi-

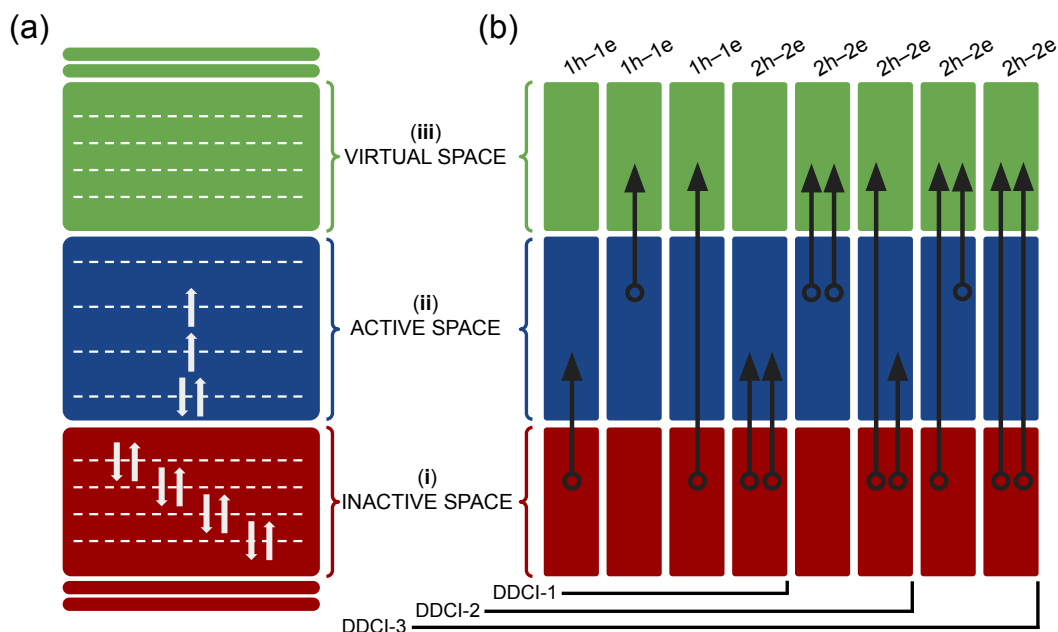


Figure 2.2: (a) Schematic representation of the inactive, active and virtual space in CASSCF calculations (colored in red, blue and green, respectively). (b) Illustration of the different type of excitations permitted in DDCI-1, DDCI-2 and DDCI-3 calculations.

nally, it is worth stressing that the specific orbitals in the active space can vary from one CASSCF calculation to another, and they are determined based on the nature of the problem being investigated. The utilization of CASSCF method in this thesis is restricted to π -conjugated organic diradicals. Accordingly, the active space encompass the highest feasible amount of π -symmetric orbitals in all cases.

2.2.2.2.2. Difference-Dedicated Configuration Interaction method

The Difference-Dedicated Configuration Interaction¹⁶ methodology is specifically tailored to excel in accurately determining energy differences rather than total energies.¹² In the context of DDCI, the CASSCF wavefunction serves as a

reference to generate a set of eight distinct types of excited Slater determinants. These determinants are categorized into classes based on the numbers of holes (h) and electrons (e) in both the occupied and virtual orbitals (as depicted in Figure 2.2b). Generally, one particular class of excitation that involves the double replacement (2h-2e) from the inactive to virtual orbitals is excluded from the CI expansion¹² (see right-most excitation in Figure 2.2b). This exclusion is justified because these double excitations do not significantly contribute to the second-order energy differences necessary for evaluating singlet-triplet energy differences,^{17,18} ΔE_{ST} , as detailed in the work of Malrieu and co-workers.¹⁶

Moreover, the DDCI approach permits three additional levels of calculations, known as DDCI-1, DDCI-2, and DDCI-3,¹⁹ with each level dependent on the number of excitations permitted in the CI. The last one of those, DDCI-3 has been the approach employed herein. However, while the exclusion of (2h-2e)-excitations significantly reduce the amount of computation, it is worth stressing that the DDCI method is still associated with high computational costs due to its reliance on a large set of determinants. In this regard, its use in the current thesis has been limited to a specific set of small size molecules to assess the accuracy of the more-affordable broken-symmetry approximation.

2.3. Assessment of the (anti)aromatic character

The measure of stability in computational chemistry requires a large endeavor, as stability is a multi-fold concept that can shift its meaning in different ways. Examples of such different meanings show up when differentiating kinetic and thermodynamic stability. Moreover, both types of stability can also be altered when setting the compounds under different conditions such as different media, temperature, pressure etc. However, for poly-cyclic and fully-conjugated hydro-

carbons (here referred as Kekulé hydrocarbons, i.e KHCs) a common approach is to assess stability by means of its (anti)aromatic character.²⁰ The understanding of (anti)aromaticity has its roots in the Hückel model. The model revolves around the criterion of the $4n+2$ rule, wherein a cyclic, planar system with $(4n+2)$ pairs of π electrons will exhibit aromatic properties while a system with $4n$ pairs of π electrons displays antiaromatic properties.²¹

Although a direct measure of aromaticity or aromatic energy can only be provided by a few methods that assume certain approximations to the Hamiltonian²² (only π -electrons, only carbon atoms, etc.), the most common way to assess and quantify the aromatic character of KHCs is by means of the molecule response upon the effect of an external magnetic field. When aromatic compounds are subjected to an external magnetic field (B_0) perpendicular to the ring, they induce a *diatropic* ring current perpendicular to the field²³ (see Figure 2.3a). This current induces a magnetic field anti-parallel to the external one (B_{ind}), intensifying its effect in the center of the carbon rings. This phenomenon aligns with the Lenz's law,²⁴ as electrons in a delocalized π -orbital hold certain similarities with a solenoid of a single coil thickness. This shielding effect indicates the molecule's resistance to the external magnetic field resulting in an effective magnetic field (B_{eff}) lower in magnitude compared to B_0 . Conversely, under similar conditions, antiaromatic compounds induce a *paratropic* ring current that yields an induced magnetic field aligning collinearly with the external B_0 field (see Figure 2.3b). This escalates the molecule's exposure to the external magnetic field, leading to a deshielding effect ($B_{\text{eff}} > B_0$).

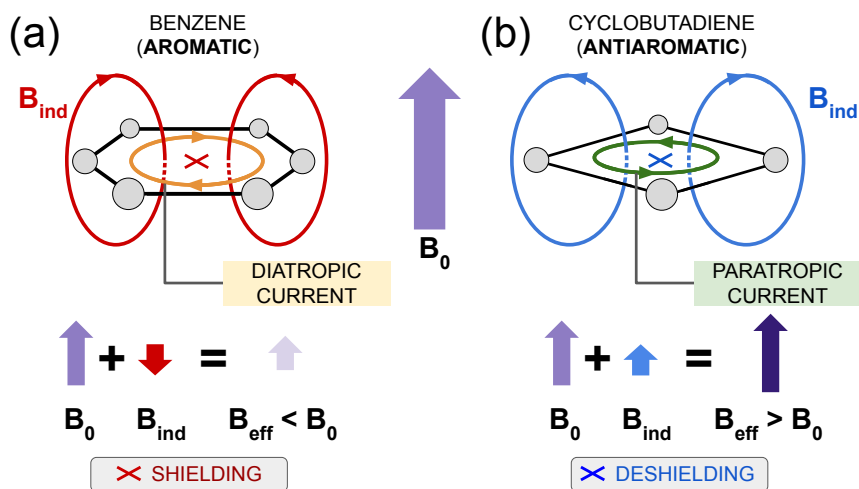


Figure 2.3: Schematic illustrating the impact of a perpendicular magnetic field (in purple, B_0) on aromatic molecules like benzene (a) and antiaromatic ones like cyclobutadiene (b). The circular currents in line with the molecular planes indicate diatropic (yellow) and paratropic (green) induced currents. The resulting induced magnetic flux, B_{ind} , is shown in red and blue based on the electron response. The lower panels of (a) and (b) depict the net magnetic field, B_{eff} , at the ring's center, indicated by the symbol \times in the molecular representations.

2.3.1. Nucleus-Independent Chemical Shift

Nucleus-Independent Chemical Shift²⁵ (NICS) stands out as a benchmark technique in computational chemistry to measure the response of chemical systems to an external magnetic field without explicitly including it in the calculation. While NICS had originally been formulated to compare the chemical shifts in an NMR spectra, it has also been proven to accomplish good results in assessing the shielding or deshielding effects in KHCs and, by extension, the (anti)aromatic nature of such compounds.²³

NICS operates by calculating the anisotropic magnetic shielding tensor²⁶ (Ω), often referred to as the chemical shielding tensor, at various positions around/within a molecule. The components of Ω , at the nucleus N (i.e. σ_{ji}^N) are defined as:

$$\sigma_{ij}^N = \left(\frac{\partial^2 E}{\partial B_i \partial m_{Nj}} \right)_{\mathbf{B}=0} \quad (2.20)$$

where E is the HF or DFT energy, B_i is the i -th component of the external field, and m_{Nj} is the j -th component of the magnetic moment of the nucleus N . From the treatment of the different components of Ω several NICS variants emerge. For example, isotropic NICS,²⁵ often referred to as NICS_{iso} , is defined as

$$\text{NICS}_{\text{iso}} = -\frac{1}{3} \text{Tr } \Omega = -\frac{1}{3} \sum_i \sigma_{ii}^N \quad (2.21)$$

That is, the mean value of the diagonal (isotropic) components. However, since NICS_{iso} considers the xx and yy components of Ω , it is not especially well-aligned with applications in assessing the aromatic character,²⁷ as the magnetic response of π -electrons mostly steams from the external field component perpendicular to the molecular plane (B_z by convention). In this regard, NICS_{ZZ} is a variant of NICS that considers only the zz component of the tensor,²⁸ thus it is simply defined as

$$\text{NICS}_{\text{ZZ}} = -\sigma_{zz}^N \quad (2.22)$$

Note that the sign in both NICS definitions is taken negative with respect to the Ω components. This convention has been mainly adopted to compare NICS with experimental NMR spectra. While NICS can be evaluated in any position, the usual convention for assessing the ring's tropicity is to calculate the magnetic tensor on the center of carbon rings. However, there are other variants consisting on evaluating NICS 1 Å above the molecular plane: $\text{NICS}_{\text{iso}}(1)$, and $\text{NICS}_{\text{ZZ}}(1)$. By moving away from the molecular plane, the direct influence of atomic nuclei is reduced,²⁹ which can give a clearer picture of the magnetic response due to the π -electron system alone.³⁰ Both $\text{NICS}_{\text{iso}}(1)$, and $\text{NICS}_{\text{ZZ}}(1)$ approaches have

been adopted in **Chapter 4** to evaluate the (anti)aromatic character of diradicals derived from antiaromatic couplers.

In a lesser note, it is worth mentioning that the magnetic field \mathbf{B} appears on the electronic Hamiltonian as a vector potential, whose origin (i.e. gauge origin) is not specifically set. While for the complete Hilbert space the results are proven independent from such origin, this is no longer true for calculations with truncated basis sets.³¹ Several solutions to the gauge origin problem have been proposed, with Continuous Set of Gauge Transformations³² (CSGT) and Gauge-Including Atomic Orbitals³³ (GIAO) being the most employed ones.

The NICS calculations reported across the present thesis have been conducted under the GIAO framework. Briefly, The GIAO *ansatz* for calculating chemical shifts by means of truncated basis utilizes the following explicitly modified field-dependent atomic orbitals as basis functions³¹

$$|\psi_i(\mathbf{B})\rangle = e^{-\frac{i}{2c}(\mathbf{B} \times \mathbf{R}_i) \cdot \mathbf{r}} |\psi_i(\mathbf{0})\rangle \quad (2.23)$$

where $|\psi_i(\mathbf{B})\rangle$ represents the i -th atomic orbital with explicit field-dependent component, $|\psi_i(\mathbf{0})\rangle$ corresponds to the usual basis set at $\mathbf{B} = \mathbf{0}$, \mathbf{R}_i is the nucleus position and, finally, i and c are the usual representations for the complex unit and the speed of light, respectively. From this modification, the partial derivatives with respect to the magnetic field can be taken in each molecular orbital constructed as a linear combination of GIAOs after the usual SCF procedure.

2.4. Solid-state calculations

The examination of 2D-CORFs (referenced in **Chapter 3**) requires solid-state calculations adopting a hybrid-DFT approach to investigate the magnetic and conducting properties of these materials. On solids, electrons experience a periodic potential due to the repeating lattice of atoms. Accordingly, one of the central aspects in electronic structure calculations in the solid state is the Bloch's theorem,³⁴ which states that the wavefunction of an electron in a crystal can be expressed as a product of a plane-wave and a function with the same periodicity as the lattice. More precisely, the Bloch's theorem can be formalized as:

$$\Psi_k(\mathbf{r} + \mathbf{g}) = \Psi_{\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{g}} \quad (2.24)$$

Here, Ψ_k represents the *crystalline orbital*, \mathbf{r} stands for a spatial position, and \mathbf{k} is the vector associated with the group of crystal translations, $\{\mathbf{g}\}$, belonging to the irreducible representations of the crystal group symmetry. In short, Bloch's theorem underscores that the *crystalline orbital*, when evaluated at position $\mathbf{r} + \mathbf{g}$, differs from $\Psi_{\mathbf{k}}(\mathbf{r})$ by only a phase factor: $e^{i\mathbf{k}\cdot\mathbf{g}}$.

Just as in non-periodic calculations, the choice of the basis for spanning the Hilbert space is critical in solid state, i.e., choosing the set of periodic functions that obeys the Bloch's theorem. Two primary methodologies emerge based on the criteria for spanning this space: (i) plane-waves combined with pseudo-potentials,^{35–37} and (ii) atom-centered orbitals.^{38–40}

Regarding methodology (i), plane-waves are inherently periodic, making them ideal for crystalline solids with a periodic lattice. They form a complete basis set, meaning that any function can be represented as a superposition of plane-

waves. Furthermore, another advantage of plane-waves is that typical solid-state operations are mathematically simpler in the reciprocal Fourier space. Despite its renowned potential, the plane-wave approach is not without limitations. For example, plane-waves oscillate rapidly near atomic nuclei.³⁶ Hence, to accurately represent wavefunctions in these regions, many plane-waves with very short wavelengths (or equivalently, high energies) are required. In practice, not all possible plane-waves can be used. Instead, a cutoff energy, E_{cut} , is set, yielding to a truncated basis set that includes plane-waves possessing kinetic energy below such threshold. To counteract this limitation, DFT calculations employing plane-waves typically use pseudo-potentials, which substitute the "real" atomic potential near the nuclei, allowing efficient wavefunction representation without requiring high-energy plane-waves.

For 2D materials, like slabs or TAM-based 2D-CORFs, methodology (ii) using atom-centered basis functions (e.g., Gaussian-type orbitals, GTOs or numerically-tabulated atom-centered orbitals, NAOs) is often more adequate. These materials are fundamentally non-periodic in one direction, typically assigned to the z-axis. Thus, to model these systems, in addition to describe their atoms, a vacuum region is introduced to prevent interactions between periodic images. Note that, in plane-wave methods, representing this vacuum escalates the computational costs without yielding additional insights. Instead, atom-centered basis functions, such as GTOs or NAOs, are localized around atoms. This implies that they naturally represent the electron density in a localized manner as well as its decay across the vacuum direction. It must be noticed that the decay of the electron density in plane-wave frameworks must be simulated, that is, it is consequential instead of instrumental.

Moreover, specific correlated materials with notable electron-electron interactions such as CORFs,^{41,42} the commonly-used Generalized Gradient Approximations⁴³ (GGA) and Local-Density Approximation⁴⁴ (LDA) result insufficient to approximate the unknown exchange-correlation (XC) term of the energy functional in DFT. Instead, a hybrid-DFT scheme⁴⁵ is essential, complementing the LDA or GGA exchange term with a percentage of Hartree-Fock exchange (HF-XC) correlation energy. In this regard, the calculation of HF-XC in localized basis schemes do not significantly add computational complexity to the calculations. However, in plane-wave schemes, calculating the HF-XC is a computationally heavy task.⁴⁶ The problem stems from the delocalized nature of the plane-waves, as every pair of plane-waves contribute significantly to the HF-XC.⁴⁷ Nowadays, the usage of plane-waves within a hybrid-DFT scheme is limited to systems described by unit cells that do not exceed a few tens of atoms.⁴⁷

Based on the earlier discussion, we have conducted solid-state calculations on TAM-based 2D-CORFs using hybrid-DFT methods with a localized basis set. We chose this approach because it is particularly suited for the 2D structure and correlated electronic states found in TAM-based 2D-CORFs.⁴¹

2.5. Evaluation of the conductivity via Hubbard model

In some of the studies involving extended systems we have evaluated the conductor/insulator character of the material by means of the Hubbard phenomenological model,⁴⁸ which is usually written in second quantization formalism⁴⁹ as

$$\hat{H} = -t \sum_{i,j} \sum_{\sigma \in \{\uparrow, \downarrow\}} \left(\hat{a}_{i,\sigma}^\dagger \hat{a}_{j,\sigma} + \hat{a}_{j,\sigma}^\dagger \hat{a}_{i,\sigma} \right) + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \quad (2.25)$$

where $\hat{a}_{i,\sigma}^\dagger$ and $\hat{a}_{i,\sigma}$ are the creation and annihilation operators acting on site i and spin state σ and $\hat{n}_{i\sigma} = \hat{a}_{i,\sigma}^\dagger \hat{a}_{i,\sigma}$ is defined as the spin density operator that measure the number of spins in state σ and in site i .

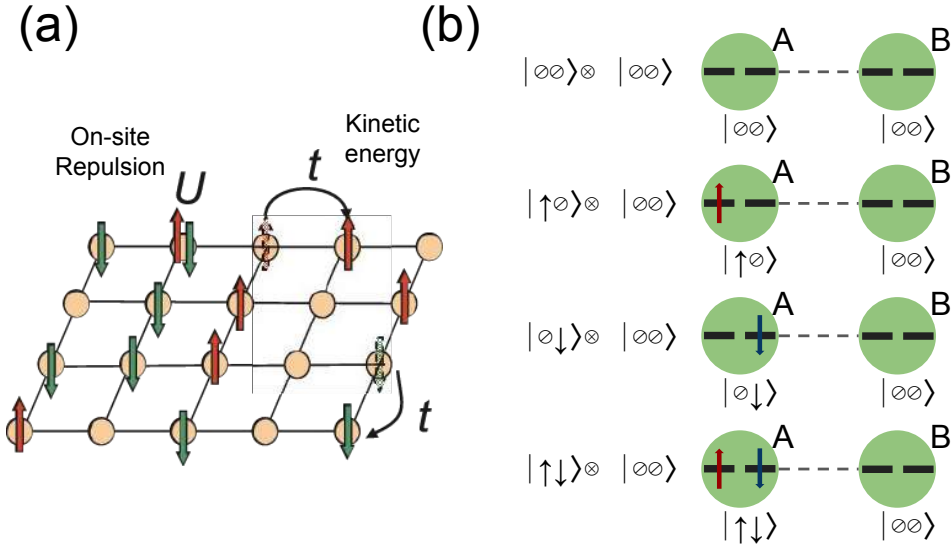


Figure 2.4: (a) Example of the Hubbard picture of a squared lattice, illustrating situations where U (on-site Coulomb repulsion energy) and t (hopping kinetic energy) parameters contribute to the total energy. (b) Example of four possible states: while the site A changes from unoccupied, $|\emptyset\emptyset\rangle$, to doubly occupied, $|\uparrow\downarrow\rangle$, the site B remains unoccupied in all the shown states.

Essentially, the Hubbard model balance the effect of two terms. The first term considers the kinetic energy of the electrons from hopping to one site to another, and it is parameterized by the hopping (or transfer) integral t representing the electron mobility defined as:

$$t = \langle \psi_i | \hat{T} | \psi_j \rangle = \int \psi_i(\mathbf{r}) \left(-\frac{\hbar^2}{2m} \right) \psi_j(\mathbf{r}) d\mathbf{r} \quad (2.26)$$

Likewise, the second term is parameterized by U , representing the on-site Coulomb repulsion, gauging the repulsion between two electrons at the same site.

$$U = \langle \psi_i \psi_i | \psi_j \psi_j \rangle = \iint \psi_i(\mathbf{r})^2 \frac{e^2}{|\mathbf{r} - \mathbf{r}'|^2} \psi_j(\mathbf{r}')^2 d\mathbf{r} d\mathbf{r}' \quad (2.27)$$

While the previously introduced HDvV model can be represented by a $2^N \times 2^N$ matrix (where N stands for the number of spins), the Hubbard model expands a $4^N \times 4^N$ Hilbert space.⁵⁰ As represented in Figure 2.4, for each pair of sites, 4 different states can emerge.

2.5.1. Role of t/U in determining the conducting/insulating character

The t/U ratio in the Hubbard model is critical in determining the conducting or insulating character of a material, providing insight into the competition between kinetic energy and interaction energy in the system.⁵¹⁻⁵³ Essentially three limit regimes exist:

1. **Band insulator or Metal** ($t/U \gg 1$): If t is much greater than U , the kinetic term dominates, implying that electrons can freely hop between lattice sites, reducing the likelihood of double occupancy and thus minimizing the importance of the Coulomb repulsion. However, if the lattice possesses two electrons per site, a given electron will not be able to hop due to Pauli's exclusion principle, giving rise to a "band insulator". On the contrary, if every site is half-filled (i.e. formally one electron per site), a metallic state is possible due to a good electron mobility.
2. **Mott Insulator** ($t/U \ll 1$): If U is much greater than t , on-site electron-electron interactions dominate. Electrons tend to avoid double occupancy, minimizing kinetic energy by localizing and thereby increasing correlation

effects. If one electron per site is present and the coupling between neighbor electrons is of AFM type, a Mott insulating phase is achieved.⁵⁴ Accordingly, electrons are localized due to strong repulsion despite a partially filled band, which would typically imply a metallic state.

- 3. Correlated Metal or Insulator ($t/U \approx 1$):** For values of t/U that are neither very large nor very small, the system might exhibit a mix of metallic and insulating behaviors, depending on additional parameters and conditions (e.g., temperature, pressure).⁵⁵ Various exotic states, like unconventional superconductivity or other metal phases,^{56,57} may arise in certain conditions.

In essence, the t/U ratio allows us to gauge whether the physical properties of a system are governed by either its kinetic or on-site Coulomb potential energy, which in turn is an indication of the material behaving as a conductor or an insulator under certain circumstances. Experimental and computational studies seek to explore these regimes and understand the rich physics that arises, especially in the correlated and strongly correlated electron systems.

2.5.2. Extraction of t and U from hybrid-DFT band structure

To determine the t/U ratio for a material with a half-filled band, one can derive it from hybrid-DFT calculations by projecting the band structure onto a tight-binding scheme and subsequently onto a Hubbard model.⁵⁸ To illustrate this process, we begin by considering a single lattice site with on-site U repulsion energy, as depicted in Figure 2.5a. Next, we examine a spin dimer within the periodic lattice, which consists of a pair of sites in the Hubbard model, each with one orbital and one unpaired electron (see Figure 2.5b). This spin dimer can adopt either a local ferromagnetic (FM) or a local antiferromagnetic (AFM)

arrangement. In the FM arrangement, both the spin-up and spin-down unmixed sites are energetically degenerate, allowing them to mix and resulting in bonding and antibonding levels separated by t . In contrast, in the AFM arrangement, the energy splitting is not simply t but is proportional to t^2/U due to the non-degeneracy of the up and down spins in the unmixed sites.⁵⁹ Finally, we can extend our analysis to continuous bands by considering a periodic lattice of spin sites, as shown in Figure 2.5c. In the FM and AFM arrangements, the valence and conduction bands exhibit widths of $2zt$ and $2zt^2/U$, respectively. Here, it is important to note that z refers to the number of nearest neighbors, which is thus determined by the lattice's topology.

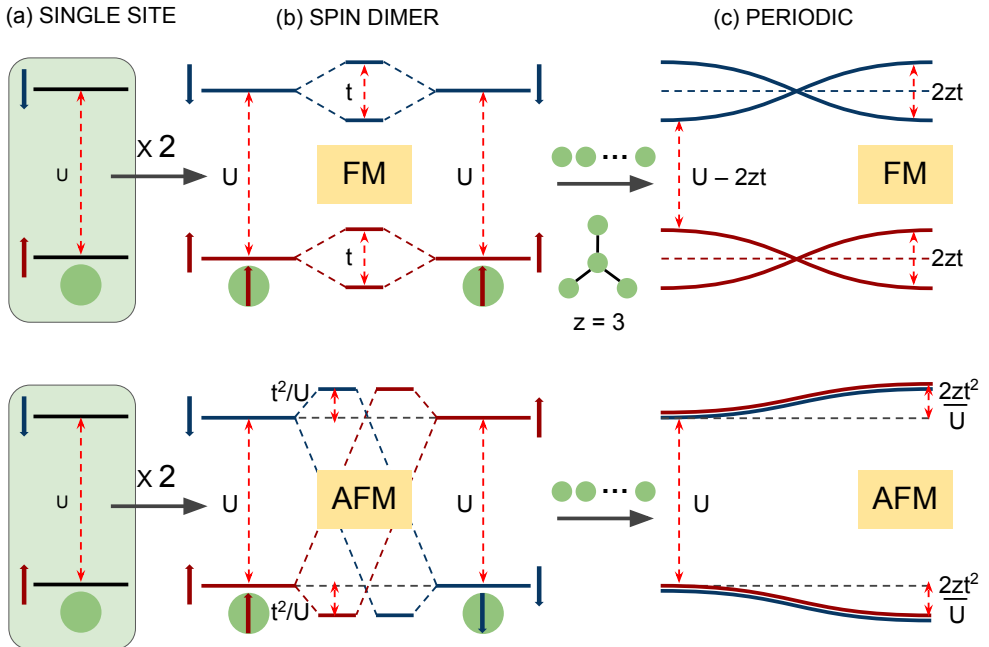


Figure 2.5: Rationalization process to extract t and U parameters from band structure (a) A single site in the lattice with on-site repulsion, U . (b) Orbital mixing in the FM and AFM arrangement of two sites in a lattice, i.e. a spin dimer. (c) Extension of the spin-dimer results to a periodic system.

Building upon our earlier observations, we can now detail the process of extracting the crucial Hubbard parameter U and the hopping parameter t from electronic structure calculations. In the context of the AFM arrangement, the band gap (E_{gap}^{AFM}) corresponds directly to U . Instead, the width of the conduction band (W^{AFM}), as previously discussed, is given by $W^{AFM} = 2zt^2/U$. Thus, we can succinctly express the extraction process as follows:

$$\begin{aligned} U &= E_{\text{gap}}^{AFM} \\ t &= \sqrt{\frac{E_{\text{gap}}^{AFM} \cdot W^{AFM}}{2z}} \end{aligned} \quad (2.28)$$

In the FM arrangement a similar process can be followed, leading to the following mapping:

$$\begin{aligned} U &= E_{\text{gap}}^{FM} + W^{FM} \\ t &= \frac{W^{FM}}{2z} \end{aligned} \quad (2.29)$$

The former procedure stated for the AFM arrangement (equation 2.28) has been employed in the present thesis to extract t , U and t/U for the TAM-based 2D-CORFs studied in **Chapter 3**.

2.6. Calculation of Young's modulus to assess the effect of mechanical stimuli on extended systems

Young's modulus, also known as the elastic modulus, is a fundamental property of materials that indicates their ability to resist deformation under tension or compression. Essentially, a higher Young's modulus means the material is

more rigid and less likely to deform under mechanical stress. This concept has played a crucial role in some aspects of this thesis, particularly in **Chapter 3**, where we assessed the practicality of applying mechanical stress to TAM-based 2D-CORFs.

Generally, the Young's modulus of a material can be derived directly from the stress (ε) vs. strain (σ) curve. In the elastic regime, i.e. the range of stress where a material returns to its original shape after the stress is removed, the stress is linearly proportional to the strain ($\varepsilon = Y\sigma$), where the slope, Y , is the targeted Young's modulus. Noting that F is the force applied normal to the plane of the material, A is the cross-section area, L_0 is the original thickness and ΔL is the thickness change of the material under mechanical strain (see Figure 2.6a for a visual illustration of these quantities), stress and strain can be defined as:

$$\begin{aligned}\varepsilon &= F/A \\ \sigma &= \Delta L/L_0\end{aligned}\tag{2.30}$$

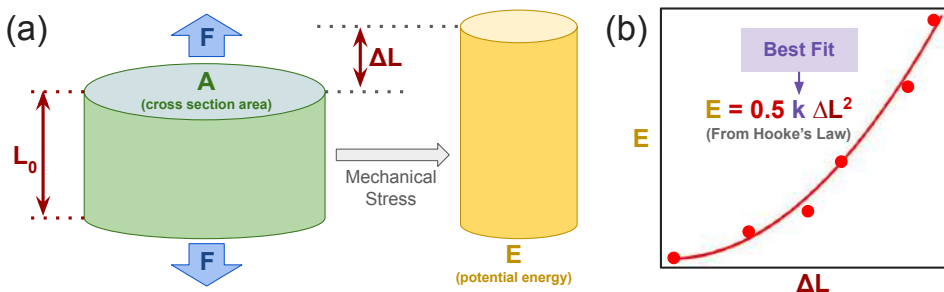


Figure 2.6: Illustration of the mechanical properties of materials. (a) A cylindrical sample subjected to a force, F , (in blue) leading to a change in length, ΔL . Note that A represents the cross-section area and L_0 is the original length. (b) Change in elastic potential energy, E , as a function of ΔL , where the curve represents the quadratic fit to data derived from Hooke's Law, and k is the spring constant to be determined.

Thus, the expression for Y simply reduces to:

$$Y = \frac{\varepsilon}{\sigma} = \frac{F/A}{\Delta L/L_0} = \frac{FL_0}{A\Delta L} \quad (2.31)$$

Our methodology for determining Young's modulus through DFT solid-state calculations relies on Hooke's Law, mathematically expressed as

$$F = k\Delta L \quad (2.32)$$

where k is the spring constant. Moreover, its associated elastic potential energy, E , can be expressed as:

$$E = \int F d\Delta L \rightarrow E = \frac{k}{2}\Delta L^2 \quad (2.33)$$

By fitting the quadratic DFT energy data as a function of length (ΔL), we can approximate k by ordinary least squares criteria (as illustrated in Figure 2.6b). Finally, considering equations 2.31 and 2.32, the Young's modulus can be approximated as:

$$Y = \frac{k_{\text{opt}}L_0}{A} \quad (2.34)$$

where k_{opt} corresponds to the optimal value after fitting k .

2.7. Ab Initio Molecular Dynamics

In certain scenarios, the static representation of molecular structure proves inadequate for simulating the physical and chemical processes occurring under real experimental conditions. This limitation arises because molecular structures experience thermal vibrations due to the effects of temperature. To capture these dynamic effects, it is necessary to integrate Hamilton's equations of motion

using Molecular Dynamics. That is, given a Hamiltonian, $\hat{H}(\mathbf{p}, \mathbf{q})$, expressed as a function of generalized coordinates \mathbf{q} and their conjugate momenta \mathbf{p} , the equations of motion can be expressed as

$$\dot{\mathbf{p}} = -\frac{\partial \hat{H}}{\partial \mathbf{q}} \quad \text{and} \quad \dot{\mathbf{q}} = \frac{\partial \hat{H}}{\partial \mathbf{p}} \quad (2.35)$$

These equations depict the evolution of the system in phase space and are rooted in the principle of stationary action, which dictates that the physical trajectory of the system minimizes the action S , defined as the integral of the Lagrangian L .⁶⁰ In turn, L is defined as the difference between kinetic T and potential V energies ($L = T - V$), integrated over time t :

$$S(\mathbf{q}(t), \mathbf{p}(t)) = \int_0^t L(\mathbf{q}(t), \mathbf{p}(t)) dt \quad (2.36)$$

In the context of the canonical ensemble, where the system is assumed to be in thermal equilibrium with a heat bath at a given temperature T , solving these equations mandates the thoughtful utilization of numerical techniques that can faithfully conserve the temperature throughout the integration by means of a simulated thermostat (see Andersen,⁶¹ Berendsen⁶², Nose-Hoover^{63,64} or the velocity re-scaling stochastic⁶⁵ thermostats), with algorithms like the Verlet or the Velocity Verlet often being preferred for this purpose.⁶⁶

The characterization of the Hamiltonian is critical, as assessing the energy and its gradient constitutes the most computationally demanding steps of the simulation. Nowadays, there are not Molecular Force Fields that work well for radical species (to the best of my knowledge). Hence, classical Molecular Mechanics cannot be utilized to solve the equations of motion of open-shell compounds. The

only available possibility is to solve the decoupled nuclear and electronic Hamiltonians, as obtained through the Schrödinger equation and Born-Oppenheimer approximation.⁶⁷ However, due to the computationally taxing equations that must be solved when using quantum mechanics, the integration of the equations of motion can only be performed on the picosecond scale. When thermal fluctuations give rise to conformational events occurring at this scale, Ab Initio Molecular Dynamics⁶⁸ (AIMD) emerges as a viable option.

In the present thesis, AIMD have been employed in two scenarios. First, to assess the thermal resilience of magnetic properties in TAM-based 2D-CORFs under anisotropic strain and, second, to generate the dataset of some specific molecular crystals. Both scenarios are discussed in **Chapter 3** and **Chapter 5**, together with annotations on the practical contribution of the author to the AIMD calculations.

2.8. Machine Learning methods

This section is dedicated to discuss the Machine Learning (ML) methodologies employed in our research, specifically in those outlined in the **Chapter 5**. Initially, we delve into the core principles and derivation of Kernel Ridge Regression (KRR), emphasizing its pivotal role in accurately predicting molecular-related quantities. Following this, the Leave-P-Groups-Out Cross-Validation (LPGO-CV) scheme is presented, serving as an essential tool for validating the KRR model during the training process. Specifically, this part elaborates on how LPGO-CV contributes to the robustness and reliability of our models across various datasets. Both KRR and LPGO-CV have been extensively employed in a plethora of applications, ranging from stock price prediction to tasks closer to chemical science. Within the scope of this thesis, both KRR and LPGO-CV have

been employed in predicting magnetic properties, particularly the magnetic exchange coupling, J_{AB} , in the systems studied. Notably, KRR requires a processed form of data, usually structured as vectors. The ability to represent complex data in such a vector format can be a challenging task, specially in disciplines such as chemistry. The specific details on how the information of a given molecule can be mapped into a vector representation are provided in **Chapter 5**, as this process corresponds to one of the major milestones discussed there. For the sake of the present discussion, let us assume that a molecular structure, \mathcal{M} , can be mapped to a vector, \mathbf{x} , by means of a set of operations, \mathcal{O} , such that: $\mathcal{O}(\mathcal{M}) = \mathbf{x}$. From this point on, the details of KRR and LPGO-CV are provided such that \mathbf{x} can be any quantity.

2.8.1. Kernel Ridge Regression

Multi-linear regression is a fundamental technique used to model the relationship between a dependent variable and multiple independent variables. At its core, it seeks to fit a linear equation to observed data. The general form of a multi-linear regression model can be expressed as

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_N x_N \quad (2.37)$$

where y is the *target* dependent variable, and $\{x_j\}_{\forall j}$ are a collection of N independent variables (or *features*). Likewise, $\{\beta_j\}_{\forall j}$ correspond to the regression coefficients, linearly mapping \mathbf{x} to y . The Ordinary Least Squares (OLS) criterion is employed to find the best-fitting by minimizing the Mean Squared Error (MSE) between multiple observations of target values, y_i , and multiple $T > N$ data records (or data *samples*) of features, $\mathbf{x}_i = \{x_{ij}\}_{\forall j}$ associated to each y_i .

Mathematically, this involves solving the following optimization problem:

$$\text{MSE} = \min_{\beta} \left[\sum_i^N \left(y_i - \sum_j \beta_j x_{ij} \right)^2 \right] \quad (2.38)$$

which, in short, can be compacted and solved in the following matrix format as

$$\mathbf{y} = \mathbf{X}\beta \xrightarrow{OLS} \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2.39)$$

where $\hat{\beta}$ corresponds to the optimal values minimizing equation 2.38. Thus, for a new unseen set of samples \mathbf{X}^* , the target properties, \mathbf{y}^* , can be readily obtained as:

$$\mathbf{y}^* = \mathbf{X}^*\hat{\beta} \quad (2.40)$$

While OLS is effective in many scenarios, it can lead to overfitting issues, particularly when the features present a high level of noise, multi-collinearity or when the number of features is similar to the number of samples. To address this, *ridge regularization*, a form of shrinkage method, is introduced. This technique adds a penalty term to the MSE objective function, which constrains the magnitude of the coefficients and, thereby, reduces model complexity. Accordingly, the ridge-regularized MSE metric can be represented in matrix notation as:

$$\text{MSE}_{\text{ridge}} = \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda\|\beta\|^2 \quad (2.41)$$

where the penalty parameter $\lambda > 0$ is nothing but a Lagrange multiplier constraining the optimization of the ridge-regularized minimization problem. Importantly, λ is usually a *user-defined parameter*. As λ increases, the impact of the penalty grows, leading to smaller coefficient values and thus a simpler, less

flexible model. Instead, when $\lambda \rightarrow 0$ the problem reduces to the ordinary, non-regularized multi-linear regression. It thus follows that the choice of λ is crucial, as it balances the trade-off between fitting the training data well and keeping the model simple to generalize to new data. It is thus usually referred to as an *hyperparameter* of the model. Overall, the constrained optimization leads to

$$\mathbf{y} = \mathbf{X}\beta \xrightarrow[\text{ridge}]{OLS} \hat{\beta} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} \quad (2.42)$$

While Ridge Regression assumes a linear relationship between the target property, y , and the feature vector, \mathbf{x} , the relationships in real-world data can be non-linear in general. In this regard, Kernel Ridge Regression (KRR) extends the previous setup to accommodate a non-linear scheme.

Consider a general non-linear mapping $\varphi : \mathbb{R}^N \rightarrow \mathbb{R}^L$, which transforms each feature vector of cardinality N to a new set of features of cardinality L . In this mapping, the original N -dimensional vector \mathbf{x} is transformed into a L -dimensional vector \mathbf{z} , such that the latter is linearly related to $\varphi(\mathbf{x})$. Representing these transformed vectors in matrix format as \mathbf{Z} , the regression model now becomes:

$$\mathbf{y} = \mathbf{Z}\sigma \quad (2.43)$$

where $\sigma \equiv \{\sigma_j\}_{\forall j}$ corresponds to the regression coefficients. Thus, equation 2.43 can be solved by ridge-regularized OLS as $\hat{\sigma} = (\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{I})^{-1}\mathbf{Z}'\mathbf{y}$.

However, a fundamental problem arises with this approach, known as the *curse of dimensionality*. For illustrative purposes, consider a simple non-linear transformation of a two-component feature vector $\mathbf{x} = (x_1, x_2)$ into:

$$\varphi(\mathbf{x}) = \mathbf{z} = (x_1, x_2, x_1^2, x_2^2, x_1 \cdot x_2) \quad (2.44)$$

Here, the original variables are expanded to include their squares and cross-product. This seemingly modest transformation leads to a substantial increase in the dimensionality of the feature space. When generalized to feature vectors of cardinality N and allowing all the possible k cross products, the transformed space grows combinatorially as:

$$L = 2N + \sum_k \frac{N!}{k!(N-k)!} \quad (2.45)$$

The exponential increase in cardinality resulting from non-linear transformations, poses significant computational challenges, particularly when calculating the inverse of the large $(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{I})$ matrix, which has a dimension of $L \times L$. This complexity requires a better approach capable of managing non-linear relationships in high-dimensional spaces effectively.

In this regard, the *kernel trick*⁶⁹ offers an efficient solution to this *curse of dimensionality*. The fundamental principle of this method is that, when the number of transformed features, L , significantly exceeds the number of data samples, T , it is computationally advantageous to work with quantities in the T -dimensional space rather than in the L -dimensional one. This reduction in dimensionality is achieved through algebraic manipulation of the non-linear ridge prediction equation, which is succinctly captured by the kernel trick as follows:

$$\mathbf{y}^* = \mathbf{Z}\sigma \xrightarrow[\text{kernel trick}]{OLS} \mathbf{y}^* = \mathbf{K}^*(\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{y} \quad (2.46)$$

where $\mathbf{K} = \mathbf{Z}\mathbf{Z}'$ (known as the *kernel matrix*) has a more manageable size of $T \times T$. Similarly, $\mathbf{K}^* = \mathbf{Z}\mathbf{Z}^*$ represents the kernel matrix for unseen data samples, with dimensions $T \times T'$, being T' the number of samples to predict. To realize further computational savings, it is essential that both \mathbf{K} and \mathbf{K}^* are computable in a straightforward manner. In this regard, noting that the (a, b) element of \mathbf{K} is $\mathbf{K}_{ab} = \mathbf{z}_a^T \mathbf{z}_b = \varphi(\mathbf{x}_a)^T \varphi(\mathbf{x}_b)$, the speed to compute each element would be drastically increased if the former inner product could be calculated without the explicit evaluation of $\varphi(\mathbf{x}_a)$ and $\varphi(\mathbf{x}_b)$ separately.

A *kernel function*, often denoted as $\kappa(\mathbf{x}_a, \mathbf{x}_b)$, provides a solution to the challenge of calculating the inner products in the transformed space. A prominent example of such a kernel function is the Radial Basis Function (RBF) kernel, also referred to as the Gaussian kernel. The RBF kernel is defined as follows:

$$\kappa(\mathbf{x}_a, \mathbf{x}_b) = \varphi(\mathbf{x}_a)^T \varphi(\mathbf{x}_b) = e^{-\gamma \|\mathbf{x}_a - \mathbf{x}_b\|^2} \quad (2.47)$$

which can be understood as the projection of the inner product to the $\exp(\cdot)$ operator. Importantly, in the equation 2.47, γ represents another *hyperparameter*, inherent to the RBF kernel, that controls the spread or decay of the kernel function. This hyperparameter effectively determines how the similarity between data points in the feature space influences their representation in the higher-dimensional space. The choice of γ is significant, as it impacts the ability of the model to generalize the correlations between data samples in the feature space. Accordingly, a small γ value leads to a RBF function with a rapid decay, implying that only data points very close to each other in the feature space, i.e., their inner product is close to 1, are considered similar in the transformed space (i.e., are also mapped to values close to 1). Conversely, a large γ value results

in a slowly decaying kernel, where data points further apart in the feature space can still influence each other significantly in the higher-dimensional space (i.e., all the data samples appear to be very similar).

Overall, the KRR approach elegantly reduces the computational complexity and efficiently handles the challenges associated with the prediction of high-dimensional data within a non-linear scheme. Although it has not been the only model employed through our investigation, it has been the principal one. Notably, KRR requires the selection of λ (the ridge regularization hyperparameter) and γ (the RBF hyperparameter), which needs careful consideration and is often determined through model validation techniques such as brute-force exploration across a limited hyperparameter space together with the subsequently explained LPGO-CV scheme.

2.8.2. Leave-p-groups-out Cross-Validation

Cross-validation is a crucial method for evaluating and optimizing the performance of a Machine Learning (ML) model. This technique aids in understanding and applying predictive modeling by providing a less biased estimate of a model's abilities. Among various cross-validation schemes, Leave-P-Out (LPO) and Leave-P-Groups-Out (LPGO) are frequently used in model evaluation and selection.

LPO cross-validation involves generating all possible combinations of the shuffled dataset of size N , by leaving out p samples at a time, as illustrated in Fig 2.7a. The total number of combinations, or folds, is determined by the binomial coefficient $K_{\text{folds}} = N!/(p!(N-p)!)$. For each fold, the model is trained on the remaining $N - p$ samples, and hyperparameters, such as λ or γ , are optimized.

The model is then tested on the p samples left out. The error metric is computed for each combination, and the overall model performance is assessed by averaging these error metrics, resulting in the cross-validated error.

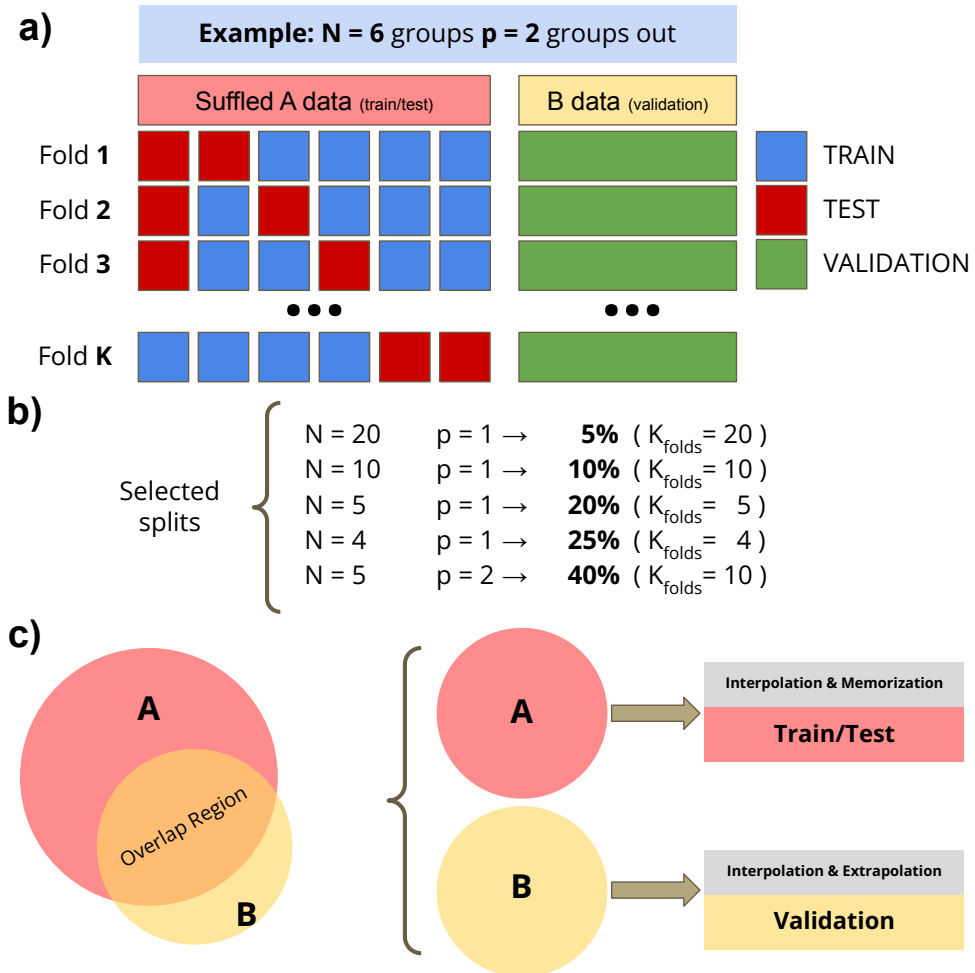


Figure 2.7: (a) Schematic representation of the possible combinations to generate all the cross-validation folds within the leave- p -groups out procedure. The scheme indicates that train (blue) and test (red) groups correspond to A dataset while validation (green) correspond to the B one. (b) Specific values for total groups and groups out (N and p , respectively) and the percentage of training and number of folds that implies each combination. (c) Schematic representation of the overlap between A and B datasets.

However, the factorial nature of the number of folds in LPO can render it computationally intensive for large datasets, being impractical for big data applications. The LPGO cross-validation scheme addresses this challenge by treating groups of samples, rather than individual samples, as the fundamental unit for cross-validation. In LPGO, N refers to the total number of these groups. The data is partitioned into N groups, with p groups being left out in each iteration to test the model, as shown in Fig 2.7b. Adjusting the number of groups and the groups left out allows us to create a learning curve, shedding light on the model's performance as the amount of training data varies.

In both LPO and LPGO schemes, samples in the test (p) and training ($N - p$) sets are routinely rotated to ensure robust evaluation. Furthermore, to enhance the rigor of the validation process, a third set known as the validation set can be incorporated. This additional set is used exclusively for evaluating the model, adding an extra layer of robustness to the model evaluation process.

In general applications of the LPGO-CV strategy, the incorporation of datasets from diverse sources or conditions can be particularly beneficial. For instance, consider two datasets, labeled as A and B, originating from different conditions or parameters. In such a scenario, Dataset A can be used within the LPGO scheme, while Dataset B serves as a separate validation set. As depicted in Fig. 2.7c, A and B may contain overlapping samples, but each dataset also includes unique samples not present in the other. This setup allows to evaluate the performance of the model further. Specifically, the error analysis on Dataset A provides insights into the interpolation capabilities of the model, while the validation on Dataset B can reveal the ability of the model to both interpolate and extrapolate.

2.9. Other aspects related to the methodology

It is worth mentioning that the scope of methodologies discussed in this section is not exhaustive of all approaches employed in the research carried out by the PhD candidate. Several other methodologies have been applied at different stages of the study and have not been detailed herein. These additional methods are deeply integrated into specific parts of the research process and, as such, they are discussed in detail within the respective chapters where they have been employed. The decision to present them in this manner was taken to ensure contextual relevance to the readership and to provide a seamless understanding of their application within the specific research framework.

2.10. Codes and methodology scope

The methodologies described previously have been employed along **Chapters 3–5**. As illustrated in Table 2.1, in **Chapter 3** we have employed **BS-DFT** (Broken-Symmetry DFT), **SS-DFT** (Solid-State DFT), **t/U** (Hubbard Model), and **AIMD** (Ab Initio Molecular Dynamics) methods, using **G09** (Gaussian09⁷⁰), **AIMS** (FHI-AIMS⁷¹), **QE** (Quantum Espresso⁷²), and **CC** (Python Custom Codes) codes. Likewise, calculations in **Chapter 4** resorted to **MRWF** (Multi-Referential Wave Function methods, as CASSCF and DDCI) and **NICS** (Nuclear Independent Chemical Shift) methods along with **BS-DFT**, primarily with **ORCA** (version 4.2.1⁷³) and **G09**. Finally, in **Chapter 5**, we utilized **SS-DFT**, **AIMD**, and **ML** (Machine Learning) techniques are applied, utilizing **AIMS**, **QE**, **SKlearn** (Sci-Kit Learn Python Package⁷⁴), and **CC** facilities. This table efficiently maps specific computational techniques to the corresponding chapters and software, illustrating the diverse methodologies in theoretical chemistry research across these chapters.

Table 2.1: Summary of the computational methods applied across *Chapters 3-5* and the associated software codes. An 'X' in red indicates the use of a method in the respective chapter. The last column lists the specific commercial codes utilized for each method.

	Chapter 3	Chapter 4	Chapter 5	CODES
BS-DFT	X	X	X	G09
MRWF		X		ORCA
NICS		X		G09
SS-DFT	X		X	AIMS, QE
t/U	X			AIMS, CC
AIMD	X		X	AIMS, QE
ML			X	SKlearn, CC

References

- (1) W. Heisenberg, *Zur theorie des ferromagnetismus*, Springer, 1985.
- (2) E. Manousakis, *Reviews of Modern Physics*, 1991, **63**, 1.
- (3) S. Choi, W. Kim and J. Kim, *AIP Advances*, 2016, **6**.
- (4) A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*, Courier Corporation, 2012.
- (5) C. R. Jacob and M. Reiher, *International Journal of Quantum Chemistry*, 2012, **112**, 3661–3684.
- (6) F. Neese, *Coordination Chemistry Reviews*, 2009, **253**, 526–563.
- (7) L. Noodleman, *The Journal of Chemical Physics*, 1981, **74**, 5737–5743.
- (8) A. Ginsberg, *Journal of the American Chemical Society*, 1980, **102**, 111–117.
- (9) M. Ernzerhof and G. E. Scuseria, *The Journal of Chemical Physics*, 1999, **110**, 5029–5036.

- (10) A. D. Becke, *The Journal of chemical physics*, 1992, **96**, 2155–2160.
- (11) C. Lee, W. Yang and R. G. Parr, *Physical Review B*, 1988, **37**, 785–789.
- (12) C. De Graaf and R. Broer, *Magnetic interactions in molecules and solids*, Springer, 2016.
- (13) M. Klene, M. A. Robb, M. J. Frisch and P. Celani, *The Journal of Chemical Physics*, 2000, **113**, 5653–5665.
- (14) N. Yamamoto, T. Vreven, M. A. Robb, M. J. Frisch and H. B. Schlegel, *Chemical physics letters*, 1996, **250**, 373–378.
- (15) P. W. Atkins and R. S. Friedman, *Molecular quantum mechanics*, Oxford university press, 2011.
- (16) P. De Loth, P. Cassoux, J. Daudey and J. Malrieu, *Journal of the American Chemical Society*, 1981, **103**, 4007–4016.
- (17) J. Miralles, J.-P. Daudey and R. Caballol, *Chemical physics letters*, 1992, **198**, 555–562.
- (18) J. Miralles, O. Castell, R. Caballol and J.-P. Malrieu, *Chemical physics*, 1993, **172**, 33–43.
- (19) J. P. Malrieu, R. Caballol, C. J. Calzado, C. De Graaf and N. Guihery, *Chemical reviews*, 2014, **114**, 429–492.
- (20) V. I. Minkin, M. N. Glukhovtsev and B. Y. Simkin, *Aromaticity and antiaromaticity*, John Wiley & Sons, Incorporated, 1994.
- (21) Q. B. zum Bezolproblem and Q. B. zum Problem, *Zeitschrift Fuer Physikalische Chemie*, 1931, **70**, 204–286.
- (22) L. J. Schaad and B. A. Hess, *Chemical Reviews*, 2001, **101**, 1465–1476.
- (23) Z. Chen, C. S. Wannere, C. Corminboeuf, R. Puchta and P. von Ragué Schleyer, *Chemical Reviews*, 2005, **105**, 3842–3888.
- (24) A. Waygood, *An Introduction to Electrical Science, 2nd ed*, CRC Press, Boca Raton, FL, 2nd edn., 2018.
- (25) P. v. R. Schleyer, C. Maerker, A. Dransfeld, H. Jiao and N. J. van Eikema Hommes, *Journal of the American Chemical Society*, 1996, **118**, 6317–6318.

-
- (26) H. Eschrig, G. Seifert and P. Ziesche, *Solid state communications*, 1985, **56**, 777–780.
- (27) G. Merino, T. Heine and G. Seifert, *Chemistry–A European Journal*, 2004, **10**, 4367–4371.
- (28) C. Corminboeuf, T. Heine and J. Weber, *Physical Chemistry Chemical Physics*, 2002, **5**, 246–251.
- (29) A. Stanger, *ChemPhysChem*, 2023, **24**.
- (30) P. v. R. Schleyer, H. Jiao, N. J. v. E. Hommes, V. G. Malkin and O. L. Malkina, *Journal of the American Chemical Society*, 1997, **119**, 12669–12670.
- (31) J. Gauss, *The Journal of Chemical Physics*, 1993, **99**, 3629–3643.
- (32) T. A. Keith and R. F. Bader, *Chemical Physics Letters*, 1993, **210**, 223–231.
- (33) J. R. Cheeseman, G. W. Trucks, T. A. Keith and M. J. Frisch, *The Journal of Chemical Physics*, 1996, **104**, 5497–5509.
- (34) F. Bloch, *Zeitschrift für physik*, 1929, **52**, 555–600.
- (35) D. Hamann, M. Schlüter and C. Chiang, *Physical Review Letters*, 1979, **43**, 1494.
- (36) D. Vanderbilt, *Physical review B*, 1990, **41**, 7892.
- (37) G. Kresse and D. Joubert, *Physical review b*, 1999, **59**, 1758.
- (38) M. F. Peintinger, D. V. Oliveira and T. Bredow, *Journal of Computational Chemistry*, 2013, **34**, 451–459.
- (39) I. Y. Zhang, X. Ren, P. Rinke, V. Blum and M. Scheffler, *New Journal of Physics*, 2013, **15**, 123033.
- (40) M. D. Towler, A. Zupan and M. Causà, *Computer Physics Communications*, 1996, **98**, 181–205.
- (41) I. Alcón, F. Viñes, I. de P. R. Moreira and S. T. Bromley, *Nature Communications*, 2017, **8**.
- (42) S. Thomas, H. Li and J.-L. Bredas, *Advanced Materials*, 2019, **31**.
- (43) J. P. Perdew, K. Burke and M. Ernzerhof, *Physical Review Letters*, 1996, **77**, 3865–3868.
- (44) W. Kohn and L. J. Sham, *Physical review*, 1965, **140**, A1133.

-
- (45) A. D. Becke, *Physical Review A*, 1988, **38**, 3098–3100.
- (46) S. Chawla and G. A. Voth, *The Journal of chemical physics*, 1998, **108**, 4697–4700.
- (47) I. Carnimeo, S. Baroni and P. Giannozzi, *Electronic Structure*, 2019, **1**, 015009.
- (48) J. Hubbard, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 1964, **277**, 237–259.
- (49) A. Beggi, I. Siloi, C. Benedetti, E. Piccinini, L. Razzoli, P. Bordone and M. G. A. Paris, *European Journal of Physics*, 2018, **39**, 065401.
- (50) H. Łukasz, S. Radosław and D. Artur, *CMST*, 2015, **21**, 181–189.
- (51) F. F. Assaad and I. F. Herbut, *Physical Review X*, 2013, **3**.
- (52) Y. Otsuka, S. Yunoki and S. Sorella, *Physical Review X*, 2016, **6**.
- (53) K. K. Gomes, W. Mar, W. Ko, F. Guinea and H. C. Manoharan, *Nature*, 2012, **483**, 306–310.
- (54) P. Phillips, *Annals of Physics*, 2007, **321**, 1634–1650.
- (55) M. Yankowitz, J. Jung, E. Laksono, N. Leconte, B. L. Chittari, K. Watanabe, T. Taniguchi, S. Adam, D. Graf and C. R. Dean, *Nature*, 2018, **557**, 404–408.
- (56) S.-H. Lee, S. Kim and K. Kim, *Physical Review B*, 2012, **86**.
- (57) Y. Cao, V. Fatemi, A. Demir, S. Fang, S. L. Tomarken, J. Y. Luo, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras, R. C. Ashoori and P. Jarillo-Herrero, *Nature*, 2018, **556**, 80–84.
- (58) H. Xiang, C. Lee, H.-J. Koo, X. Gong and M.-H. Whangbo, *Dalton Trans.*, 2013, **42**, 823–853.
- (59) P. W. Anderson, *Physical Review*, 1961, **124**, 41–53.
- (60) F. Gay-Balmaz and H. Yoshimura, *Entropy*, 2018, **21**, 8.
- (61) H. C. Andersen, *The Journal of chemical physics*, 1980, **72**, 2384–2393.
- (62) H. J. Berendsen, J. v. Postma, W. F. Van Gunsteren, A. DiNola and J. R. Haak, *The Journal of chemical physics*, 1984, **81**, 3684–3690.
- (63) S. Nosé, *The Journal of chemical physics*, 1984, **81**, 511–519.

- (64) W. G. Hoover, *Physical review A*, 1985, **31**, 1695.
- (65) G. Bussi, D. Donadio and M. Parrinello, *The Journal of Chemical Physics*, 2007, **126**.
- (66) D. Frenkel and B. Smit, *Molecular simulation: from algorithms to applications*, 2000.
- (67) M. Born and R. Oppenheimer, *Annalen der Physik*, 1927, **389**, 457–484.
- (68) D. Marx and J. Hutter, *Ab initio molecular dynamics: basic theory and advanced methods*, Cambridge University Press, 2009.
- (69) B. E. Boser, I. M. Guyon and V. N. Vapnik, Proceedings of the fifth annual workshop on Computational learning theory, ACM, 1992.
- (70) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian09 Revision E.01*, Gaussian Inc. Wallingford CT 2009.
- (71) V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler, *Computer Physics Communications*, 2009, **180**, 2175–2196.
- (72) P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari and R. M. Wentzcovitch, *Journal of Physics: Condensed Matter*, 2009, **21**, 395502.

-
- (73) T. Hočevar and J. Demšar, *Journal of Statistical Software*, 2016, **71**, 1–24.
- (74) F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *Journal of Machine Learning Research*, 2011, **12**, 2825–2830.

Chapter 3

Tuning the electronic properties of 2D-CORFs by means of external stimuli

3.1. Introduction

Organic aromatic hydrocarbons with diradical character have experienced a growing interest in recent years, largely due to their versatile applications across chemistry,¹ biochemistry,² and materials science.³ These entities play a crucial role in organic electronics and spintronics,⁴ particularly in developing molecular switches.^{5,6} Moreover, their potential in advancing battery technology,⁷ nonlinear optics⁸ and enhancing solar energy conversion through singlet fission processes^{9–12} are equally promising.

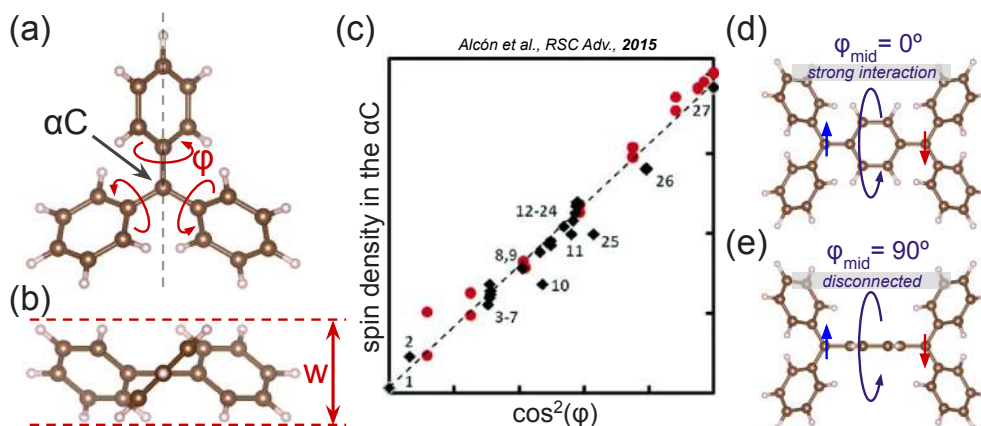


Figure 3.1: (a) Top and (b) side views of the TAM radical, where the αC , the phenyl twisting angle, φ , and the width of the TAM, w , are indicated. (c) Atom-partitioned spin density population in the αC as a function of φ (adapted from Alcón et al.¹³) Two configurations of Thiele's diradical with (d) 0° and (e) 90° twisting angle, φ_{mid} , of the shared benzene, as an extension of the TAM characteristics.

Among the diverse set of organic compounds with diradical character, those based on triarylmethyl¹⁴ (TAM) mono-radical building-blocks stand out for their historical significance and their resilience to time and temperature variations (see Figure 3.1a,b). Central to the structure of TAM is its αC atom bonded to three aryl rings. Interestingly, these aryl rings in TAMs are not planar, but twisted

relative to the sp^2 -hybridized plane of the central carbon¹⁵ (see φ and w in Figure 3.1a,b). This twisting is a crucial structural adaptation that mitigates steric repulsion between the aryl units. Notably, the aryl twist in TAMs is not just a mere structural curiosity, but provides a bulky chemical environment that protects the structure upon dimerization, disproportionation and reduction processes.¹⁶ Moreover, the aryl twist is also the principal driving force modulating the π -conjugated delocalization of the unpaired electron. Indeed, the degree of spin confinement in the αC has been found to be linearly proportional to the cosine squared of the dihedral angle with the sp^2 plane,¹³ as shown in Figure 3.1c. This feature directly affects a variety of properties such as magnetic interactions, optical absorption, and magnetoresistance of TAMs.^{17,18} It thus follows that the capacity to adjust the aryl ring twist is sought, as it provides an interesting mechanism for fine-tuning the properties of TAM-based compounds at molecular level.

One of the simplest diradical derivatives based on TAMs is Thiele's diradical¹⁹ (THIL), as showcased in Figure 3.1d,e. In relation to TAMs, the THIL diradical can be conceptualized as two triphenylmethyl open-shell (TPM) units covalently linked through a shared benzene ring. Paralleling the earlier discussion on TAMs, the interaction between the two unpaired electrons in THIL is significantly influenced by the dihedral angle of the benzene rings,²⁰⁻²⁴ specially the central benzene ring, denoted as φ_{mid} in Figure 3.1c,d. When this benzene moiety aligns in-plane (illustrated as $\varphi_{\text{mid}} = 0^\circ$ in Figure 3.1d), the π -conjugation through this ring maximizes. This, in turn, results in a scenario where the unpaired electrons either remain unpaired and present strong antiferromagnetic interaction or adopt a closed-shell quinoidal electronic structure. Contrarily, when the benzene ring

arranges perpendicular to the plane (represented at $\varphi_{\text{mid}} = 90^\circ$ in Figure 3.1e), the conjugation is effectively disrupted, leading to weak magnetic interactions between the unpaired electrons, becoming thus essentially disconnected.^{24–26}

In the intermediate structural conformations, THIL-based diradicals display an interplay of characteristics between the two energetically similar but structurally and electronically distinct configurations, i.e. the open-shell antiferromagnetic (AFM) and the closed-shell quinoidal (CSQ) states. This bistability allows for interesting possibilities, especially when considering the impact of temperature, such as bond vibrations and aryl ring rotations. These thermal effects make THIL-based diradicals undergo transitions between the two electronic states,²⁷ giving rise to an interplay between magnetic and conducting properties.

A notable limitation of the application of TAM-based diradicals, however, lies in its form of discrete molecular entities. That is, broader device-level applications of the phenomena derived from TAM-based diradicals require scaling the bistable properties in solid-state, i.e., achieving a degree of π -conjugation that extends through the solid. As these molecular diradicals arrange together resulting in molecular crystals, driven primarily by intermolecular cohesive forces, it is hard to devise functional materials derived from TAM units which present properties emerging from the π -conjugation mechanism discussed so far.

Bypassing the difficulties of manipulating molecular units in molecular crystals, several theoretical and computational studies have proposed the integration of diradicals into 2D covalent organic radical frameworks^{28–33} (2D-CORFs), as schematically shown in Figure 3.2a. Specifically, 2D-CORFs built upon TAM units usually exhibit a hexagonal, honeycomb, covalent pattern akin to graphene,³⁴

resulting from a potential three-fold polymerization of the TAM building-blocks. Notably, advanced theoretical studies using hybrid-DFT calculations in the solid state^{29,35} have shed light on the electronic structures of TAM-based 2D-CORFs. These exhibit three close in energy, yet electronically different, states (illustrated in Figure 3.2b, framed in blue). First, the ground state is found to be an antiferromagnetic insulator (AFMI) state, closely followed in energy by a CSQ state. Accordingly, both can be seen as an extended version of the states previously discussed before for THIL diradical. Additionally, a graphene-like semimetallic state³⁶ (SM) is also present, displaying slightly higher energy compared to both the AFMI and CSQ states (see Figure 3.2b). As in graphene, the SM state in TAM-based 2D-CORFs is characterized by a Dirac cone,³⁷ where the valence and conduction band energies near the Fermi level take the shape of the upper and lower halves of a conical surface, meeting at what are referred to as Dirac points. This band structure gives rise to ballistic electron transport,³⁸ characterized by both a high electron mobility and the ability of the conducting electrons to move through the material without scattering. Remarkably, the electronic structure properties of TAM-based 2D-CORFs extend beyond the (already interesting) theoretical constructs. In this regard, later studies conducted in parallel by Wu³¹ and Yang³² not only demonstrate the realistic experimental viability of synthesizing 2D-CORFs, but also corroborate the accuracy of the computational studies in determining their properties, showcasing a situation where theory anticipates the experimental evidence.

In this chapter, we exploit the unique covalent, fully-conjugated characteristics of TAM-based 2D-CORFs to induce controlled rotations of aryl rings by means of external mechanical stimuli, thereby manipulating the relative stability of the

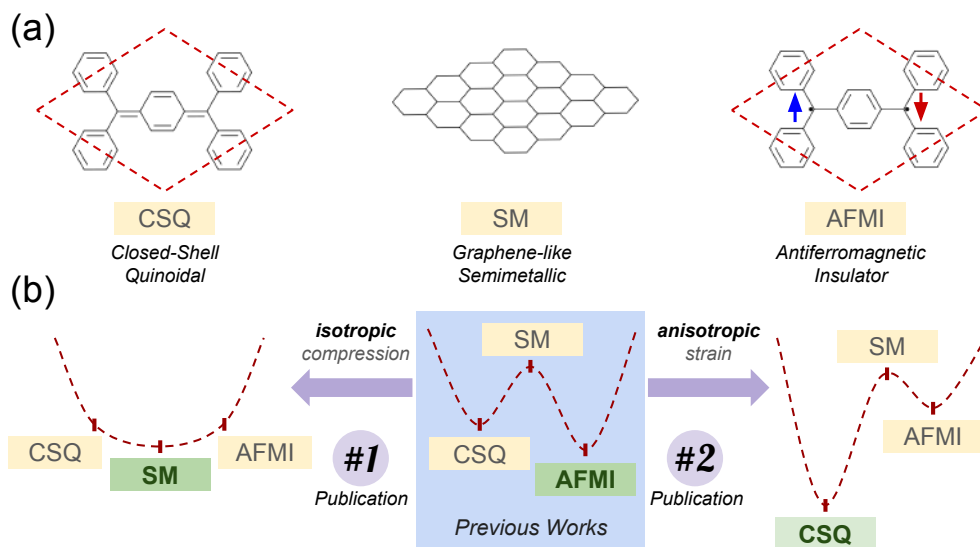


Figure 3.2: (a) Schematic representation of the accessible electronic structures in 2D-TAM-CORFs: In CSQ and AFMI the hexagonal cell of a possible CORF is marked in dashed-red. The SM electronic structure is illustrated as widely known for Graphene. (b) Diagram of the work conducted and discussed in this Chapter, where the energy difference between the CSQ, SM and AFMI electronic structures are shown for the ground-state geometry of the CORF (i.e., the geometry without strain, framed in blue), when an isotropic compression is applied (covered in **Publication #1**) and when an anisotropic strain is applied (covered in **Publication #2**). Note that the scheme only represents a slice of the actual Potential Energy Surface since the dashed line joining the energy levels is only included as a guide to the eye and the SM state is not demonstrated to be a saddle point.

electronic states of the material. Our approach aims at destabilizing the AFMI state, i.e., the ground state, in favor of stabilizing either the semimetallic (SM) or closed-shell quinoidal (CSQ) electronic structures. The first study, detailed in **Publication #1** (Figure 3.2b, left), investigates the application of isotropic out-of-plane compression to a set of TAM-based 2D-CORFs. In this regard, our main purpose is to flatten the aryl ring twisting angles, easing the electron communication between α C centers, thus promoting the stabilization of the SM

electronic structure. In contrast, the second study, presented in **Publication #2** (Figure 3.2b, right), focuses on the effects of anisotropic in-plane tensile stress. This form of stress is anticipated to selectively alter the spatial arrangement of the aryl moieties. Specifically, we expect to locally create more free space for one of the three aryl rings, while increasing steric hindrance among the others. Accordingly, this could lead to one aryl ring twisting back towards the plane, while the remaining two aryl rings twisting predominantly out-of-plane. Overall, such a change in the molecular conformation is anticipated to favor the formation of local quinoids, thereby preferentially stabilizing the CSQ state in the solid state.

In both studies, we have considered at least three examples of TAM-based 2D-CORFs (see in Figure 3.3a-c), which are based on: (a) ring-sharing (rs) triphenylmethyl (TPM) radicals, \mathbf{TPM}_{rs} , (b) ring-sharing perchlorotriphenylmethyl (PTM) radicals, \mathbf{PTM}_{rs} , and (c) acetylenic linked PTM radicals, as found in a recently synthesized and experimentally characterized material^{31,32} ($\mathbf{PTM}_{acetylenic}$). Please, note that although \mathbf{TPM}_{rs} and \mathbf{PTM}_{rs} have not yet been synthesized as extended materials, recently, it has been experimentally demonstrated that \mathbf{TPM}_{rs} can be connected via ring-sharing in molecules containing a ring of six αC centers³⁹ (i.e., a \mathbf{TPM}_{rs} flake). Similarly, a 2D framework based on a perchlorated Chichibabin's diradical has also been prepared and characterized.⁴⁰ These breakthroughs open the door to the possibility of synthesizing extended 2D-CORFs based on ring-sharing TAMs. On top of that, \mathbf{TPM}_{rs} and \mathbf{PTM}_{rs} are the simplest realizable examples based on TAM building blocks, thus serving also as proof of concept.

The studies related to both contributions are presented separately in the upcoming Results and Discussion section, i.e., **Section 3.2**. First, **Section 3.2.1** focuses on the out-of-plane compression of 2D-CORFs. Here, we focus on the effect of compression applied perpendicular to the 2D plane of the 2D-CORFs (see Figure 3.3). We examine the evolution of their electronic properties under this specific stepwise compression. Then, in **Section 3.2.2**, the focus is shifted to the second study, which assesses the effects of in-plane stretching of the 2D-CORFs and, likewise, the resultant changes in their electronic structure. After that, a detailed examination and main results extracted from each study is conducted in the Conclusions section, i.e., **Section 3.3**. Finally, **Section 3.4** presents the formatted texts belonging to **Publication #1** and **Publication #2**.

3.2. Results and discussion

In general terms, our theoretical exploration of TAM-based 2D-CORFs upon out-of-plane compression and in-plane tensile stress has been carried out by means of all-electron density functional theory (DFT) calculations in the solid-state (see **Chapter 2** for more details). In this regard, hybrid functionals provide a computationally effective approach that can be applied in periodic schemes, which gives an accurate description of magnetic properties for many classes of compounds, otherwise uncaptured by means of pure GGA or LDA approaches.³⁶ Specifically, studies conducted by means of pure LDA functional show that, for many 2D-CORFs, only the graphene-like SM electronic state is found.³⁶ Fundamental aspects of electron correlation in 2D-CORFs and its appropriate description of the ground state properties are thus overlooked.

Notably, while the layered stacking pattern in **PTM_{acetylenic}** is established as an A–B arrangement,³¹ the potential 3D structural arrangement of **TPM_{rs}** and

PTM_{rs} CORFs remain uncharted. However, considering that the π -conjugation phenomenon under examination is highly anisotropic, predominantly occurring within the individual layers of the 2D-CORFs, our study has been mainly centered on analyzing single layers (SL) of these materials. Moreover, to mitigate self-interaction errors in our analysis, we defined the unit cell of the calculations in such a way that a separation of 40 Å of vacuum is set between a SL and its periodic image. This approach, combined with the use of atom-centered numerical orbitals (as implemented in FHI-AIMS package,^{41,42} see **Chapter 2**),

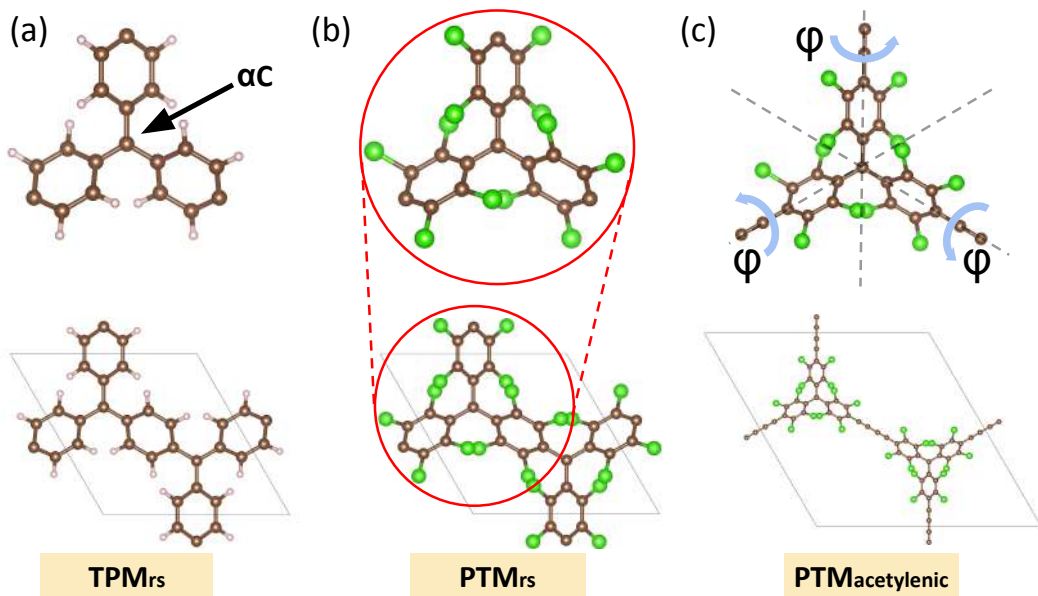


Figure 3.3: Structures of the three considered 2D-CORFs: (a) ring-sharing triphenylmethyl (TPM_{rs}) radicals, (b) ring-sharing perchlorotriphenylmethyl (PTM_{rs}) radicals, and (c) acetylenic coupled PTM radicals ($\text{PTM}_{\text{acetylenic}}$). Top: Local chemical environment of the corresponding protected triconnected sp^2 carbon sites (αC) for each structure. Bottom: Cuts from each extended 2D-CORF structure with the unit cell delimited by a gray line. In (c), we highlight the aryl ring dihedral angles (φ) that are affected by potential structural deformations in all structures. Atom color key: C: dark brown; H: light brown; Cl: green.

allows us to integrate this vacuum space effectively into our calculations without significantly increasing computational cost. These methodological choices streamline our post-DFT analyses, particularly the examination of the band structures of these materials and the mapping of the electronic structure properties to distinct model hamiltonians (discussed below). Drawing from previous research employing this computational setup,^{28,29,35} we have further verified that the characteristics of a single sheet of the studied TAM-based 2D-CORF are accurately captured by DFT/PBE0 calculations. Overall, this approach enables us to demonstrate how the electronic states of these materials can be modulated through the application of both compression and tensile stress.

3.2.1. Out-of-plane compression in TAM-based 2D-CORFs

The results related to the out-of-plane compression are divided into two sections. Initially, we review some additional aspects related to the computational methodology employed to simulate compressive stress, focusing on the specific procedures and parameters used. This section also examines the structural and conformational changes induced onto the 2D-CORFs resulting from the simulated compression scheme. Following this, the second section delves into a detailed discussion of the electronic structure. Here, we explore how structural modifications induced by compression impact the electronic properties of the materials.

3.2.1.1. Structural distortions upon compressive stress

For bulk, 3D materials, computational simulation of compression often involves a stepwise, *ad hoc* reduction of one or more cell parameters, allowing the optimization of the remaining cell parameters and atomic positions. This approach has been effectively used in studies such as the compression of black

phosphorus, where it helped to assess the transition from an insulating to a metallic phase.⁴³ However, the TAM-based CORFs considered in our study consist of single layers separated by a simulated vacuum. Thus, it follows that the process of simply reducing the cell parameter will not be employed since it would only diminish the vacuum width, keeping the structure of the CORFs essentially unaltered.

To accurately simulate out-of-plane compression on 2D single layer systems, we consider two distinct approaches: (i) embedding the SL in a heterostructure and applying compression to the composite system, i.e., a method highly inspired on the experimental procedure,^{44,45} and (ii) applying specific *ad hoc* constraints to the atomic coordinates to directly alter the main internal degrees of freedom involved in a compression process. These methods are designed to overcome the unique challenges posed by the 2D nature of our materials and seek a realistic depiction of the effects of compression.

To implement approach (i), we prepared a 2D layered heterostructure consisting of five different layers, in which **TPM_{rs}** was sandwiched between two slabs of graphite consisting of two layers of graphene (see Figure 3.4). After that, the following steps were undertaken to simulate the compression process:

- We first carried out an unconstrained geometry and cell relaxation for the entire heterostructure, aiming to establish its ground state geometry.
- We reduced and fixed the distance between the outermost graphene layers (denoted as $d(\text{G}_{\text{fix}} \cdots \text{G}_{\text{fix}})$, highlighted by the red arrow in Figure 3.4) to simulate a compression process. With this distance held constant, we performed a geometry and cell relaxation for the entire system, ensuring

that the z -coordinate of the atoms constituting the outermost graphene layers was kept fixed.

- Subsequent to each compression step, we measured the average twisting angle of the aryl rings within TPM_{rs} and we calculated the associated compression, expressed as the percentage reduction in thickness.
- This process was repeated iteratively for a range of $d(\text{G}_{\text{fix}} \cdots \text{G}_{\text{fix}})$ values, allowing us to assess the effects of varying compression levels on the properties of the heterostructure (as illustrated in Figure 3.4).

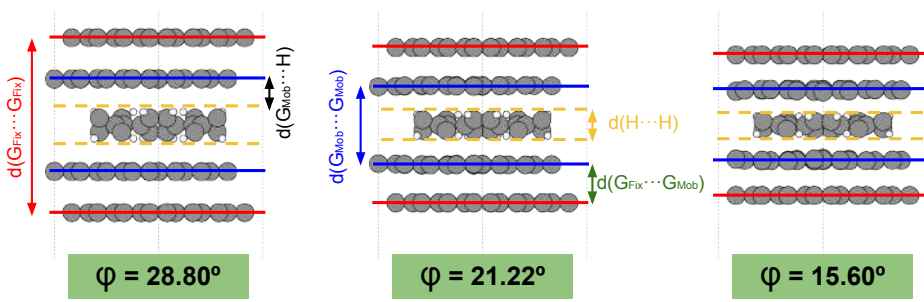


Figure 3.4: *2D layered heterostructure model used to apply compression to a SL of TPM_{rs} by graphene layers. Blue and red lines indicate the xy -plane location for the mobile and fixed graphene layers, respectively. Likewise, the dashed yellow line indicates the upper and lower limits of the sandwiched TPM_{rs} 2D-CORF, marked by the position of the H atoms. The different arrows indicate the relative distances between different layers along the heterostructure. Note how these vary upon increasing compression from left (no compression) to right (large compression). The text framed in green highlights the phenyl ring twisting angle, φ , for all the amounts of compressions.*

It is important to note that, while this approach of simulating compression in a layered heterostructure aims at mirroring the experimental setup, we found that there are inherent technical limitations within the computational periodic schemes. Specifically, both the compressed SL of the 2D-CORF and the graphene slabs are required to share the same in-plane cell parameters dimensions to be

modeled. This constraint means that any epitaxial mismatch between the different layers cannot be simultaneously accommodated. Consequently, the in-plane cell parameters of the entire multi-layered heterostructure are predominantly determined by the layer with the highest stiffness. In our case, this is graphene, which possesses the highest Young's modulus among all 2D materials (approximately 2 TPa upon biaxial strain⁴⁶), leaving the 2D-CORF in a forced strained configuration even at the ground state of the composite system. Note that, while epitaxy is an observed phenomenon taking place in surfaces or junctions,^{47,48} epitaxial stress in these kind of systems is clearly a computational artifact, because in real experimental conditions the epitaxial effects would barely play any roll, as the interaction between 2D-CORFs and graphene is largely labile.³⁶

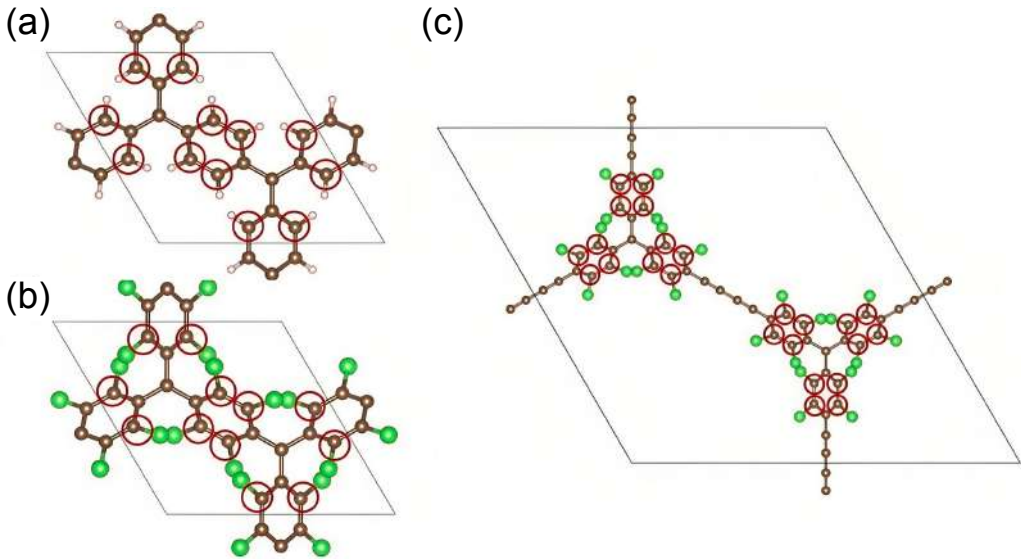


Figure 3.5: Unit cell of the crystal structures of (a) TPM_{rs} , (b) PTM_{rs} , (c) $PTM_{acetylenic}$ as viewed along the z -axis. The sets of atoms whose z -axis (i.e. out-of-plane vacuum direction) were fixed during geometric optimization are circled in red (12 for PTM_{rs} and TPM_{rs} and 24 for $PTM_{acetylenic}$).

In order to solve the epitaxy issues detected in the former method, approach (ii) involves applying compressive stress to the SL of the 2D-CORF by introducing structural constraints to the z-component of specific atoms within the scaffold. As in approach (i), this method allows for the full relaxation of the 2D lattice, along with the unconstrained atoms and the x and y coordinates of the constrained atoms. The rationale behind this approach emerged from our observations in approach (i), where we detected that the most significantly altered structural parameter under compression was, indeed, the twisting angle of the aryl rings (see Figure 3.4). It is thus reasonable to simulate the compression process as a constrained rotation of the aryl rings. To minimize alterations to the structural parameters, we strategically limited the number of constrained atoms to the smallest possible group that could realistically undergo modifications under experimentally realizable compression scenarios, i.e., the carbon atoms within the aryl moieties that lay out of the 2D-CORFs' plane (see the atoms circled in red in Figure 3.5). This selection was aimed at closely mimicking the structural evolution observed in approach (i), focusing specifically on replicating the changes in the twisting angle.

In all subsequent studies, we thus employed the approach (ii). Moreover, the compression applied to each 2D-CORF system was calibrated to reach up to 60% of its original out-of-plane thickness, effectively reducing it to 40% of its initial measurement (as depicted in the x-axis of Figure 3.6a,b). To assess the extent of the mechanical forces involved, we calculated the Young's modulus perpendicular to each 2D-CORF.⁴⁹ This analysis allowed us not only to approximate the out-of-plane pressure necessary to achieve varying degrees of compression for each material, but also to guarantee that the materials do not fail structurally across

the 0–60% range (see the implementation details in the methodology chapter, i.e., **Chapter 2**). In this regard, our calculations revealed that the perpendicular Young’s moduli for TPM_{rs} , PTM_{rs} , and $\text{PTM}_{\text{acetylenic}}$ are 5.1 GPa, 13.8 GPa, and 0.74 GPa, respectively. Accordingly, this suggests that relatively modest pressures of approximately 0.90 GPa for TPM_{rs} , 1.26 GPa for PTM_{rs} , and 0.07 GPa for $\text{PTM}_{\text{acetylenic}}$ would be sufficient to realize the maximum compression levels explored in our study (further illustrated as insets in Figure 3.6b).

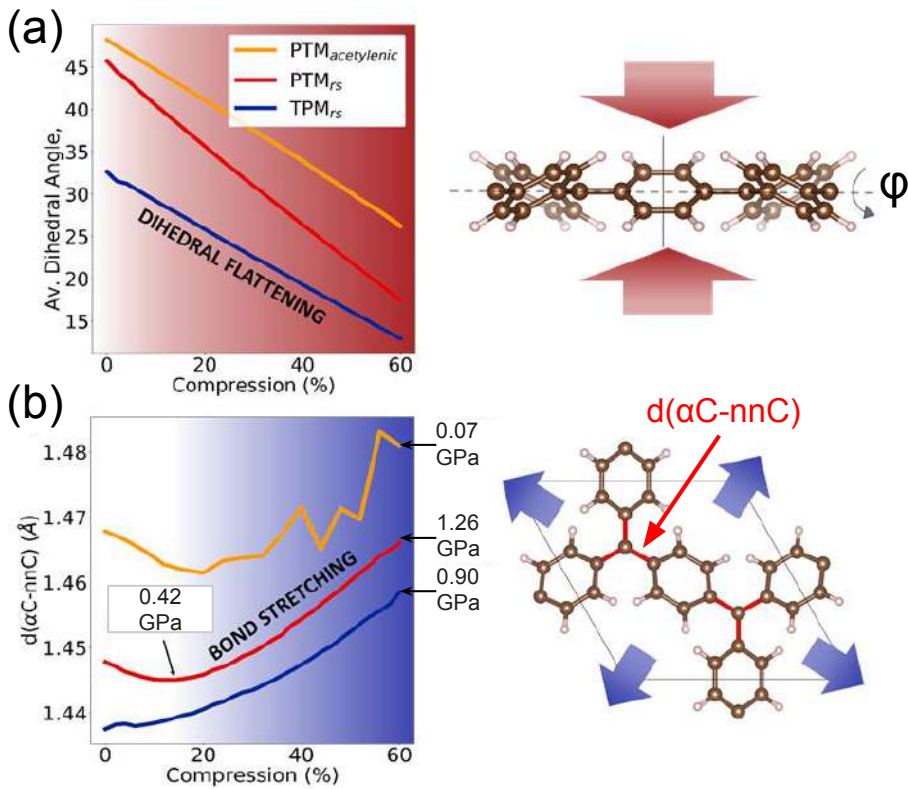


Figure 3.6: Response of TAM-based 2D-CORFs upon out-of-plane compression. The contributions of ring flattening and in-plane bond stretching in the AFM state are indicated by plots of: (a) average ring dihedral angle $\langle\varphi\rangle$, and (b) average of the bond length between the αC and its nearest-neighbor carbon ($\alpha\text{C-nnC}$, highlighted in red), respectively. The pressures in GPa required for selected compressions in each 2D-CORF are indicated in (b).

The key structural changes observed in each 2D-CORF as a result of the compression process are illustrated in Figure 3.6. When the unstrained materials are subjected to compressive stress, we observe a linear decrease in the dihedral angles of the aryl rings relative to the plane delineated by the αC centers (i.e. ring flattening), aligning with the *ad hoc* constraints we applied. The response of the in-plane bond distance between the αC atoms and their nearest neighbors (denoted by $\alpha\text{C}\text{--nnC}$ in Figure 3.6b) to the out-of-plane compression, however, presents a more complex, nonlinear pattern. In general terms, as compression intensifies, these bond distances tend to increase (indicative of bond stretching) due to the increasing steric hindrance among the aryl rings. Essentially, the process of external out-of-plane compression initiates a two-fold interplay: on the one hand, the $\alpha\text{C}\text{--nnC}$ bond strength intensifies owing to ring flattening and the resulting enhancement in $\pi\text{--}\pi$ orbital overlap interactions and, on the other, the growing steric hindrance among aryl rings leads to an expansion of the in-plane $\alpha\text{C}\text{--nnC}$ bonds to attenuate the former. It must be emphasized that this bond stretching would have been unnoticed by the aforementioned approach (i), as the graphene slabs would have prevented the cell parameters to stretch.

The initial average dihedral angle of the aryl rings for TPM_{rs} , denoted as $\langle\varphi\rangle$, is relatively small (approximately 32.6°), a characteristic attributed to the minimal steric hindrance from the H-functionalized aryl rings (see blue line in Figure 3.6a). Under the maximal compression of 60%, this angle reduces significantly, reaching a value of 12.9° due to ring flattening. Accordingly, the in-plane $\alpha\text{C}\text{--nnC}$ bond lengths in TPM_{rs} , which start at slightly below 1.44 \AA , undergo a steady increase with compression, reaching up to about 1.46 \AA at 60% (see blue line in Figure 3.6b).

In the case of **PTM_{rs}**, the initial dihedral angles of the aryl rings, which are influenced by the sterically bulky Cl-functionalized groups, start at a higher value of 45.7° due to the larger steric hindrance among the Cl atoms. Upon maximum compression, this angle is notably reduced to 17.5°. Interestingly, while **PTM_{rs}** shares a similar carbon network topology with **TPM_{rs}**, its higher initial $\langle\varphi\rangle$ values lead to diminished π - π orbital overlap in its uncompressed state, resulting in longer average α C- nnC bond lengths compared to **TPM_{rs}**. Related to this fact, in the initial compression range of 0–25%, **PTM_{rs}** exhibits a unique response: the α C- nnC bond lengths initially contract before expanding as compression progresses and aryl rings flatten more. This distinctive pattern arises from the initially lower π - π overlap in **PTM_{rs}**, which, due to the longer and more strained α C- nnC bonds in its uncompressed form, responds to ring flattening by enhancing π - π overlap.

Similarly, the aryl rings in **PTM_{acetylenic}** are functionalized by Cl atoms, resulting in an initial dihedral angle $\langle\varphi\rangle$ close to that of **PTM_{rs}** (*ca.* 48.2°). However, the π -conjugation in **PTM_{acetylenic}** is mediated by lengthier acetylenic linkers, which results in a lesser magnitude of π - π overlap compared to **PTM_{rs}**. This leads to notably longer initial α C- nnC bond lengths, measuring approximately 1.468 Å. Compression impacts **PTM_{acetylenic}** in a way akin to **PTM_{rs}**, initially enhancing the π - π orbital overlap and causing a reduction in the average α C- nnC bond lengths. Importantly, as compression increases, **PTM_{acetylenic}** exhibits notable variability in these bond lengths, a phenomenon evident across its conjugated in-plane carbon network. This variability is attributed to the existence of energetically competing spiral conjugated π - π overlaps in acetylenic chains.⁵⁰ Despite these fluctuations, there is no indication of structural failure

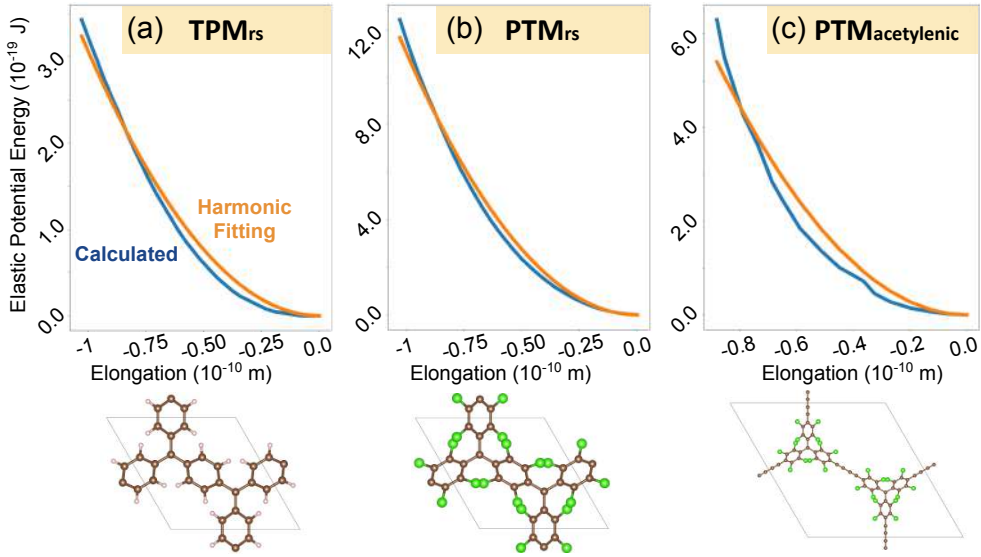


Figure 3.7: *Elongation vs. total energy curves. The results of the DFT/PBE0 calculations (blue) and the prediction assuming the elastic regime (orange) are shown for all the materials studied during this work: (a) TPM_{rs} , (b) PTM_{rs} , (c) $PTM_{acylenic}$. Note that a high degree of similarity can be observed between the actual data and the harmonic fitting.*

in $PTM_{acylenic}$, as evidenced by the minimal impact of these changes on its energy-compression relationship, indicating a clear harmonic behavior ($\sim x^2$) within this range of compression (see Figure 3.7a-c).

3.2.1.2. Electronic structure response upon compression

We observed that the discussed structural changes arising from compression lead to notable modification of the electronic structure in all three types of 2D-CORFs under examination. According to our hybrid-DFT calculations, a gapped AFMI ground state is initially present in each case. Yet, as compression intensifies, this AFMI state becomes increasingly less stable energetically when compared to the SM closed-shell state, a trend clearly depicted in Figure 3.8a. This shift in stability has been further quantified by examining several related

parameters: the average absolute spin moment per αC atom, the AFM coupling between αC centers (referred to as J), and the electronic bandgap, represented by ΔE . As illustrated in Figures 3.8b-d, all these metrics display a general decline with the progression of compression in each of the 2D-CORFs.

Analyzing the data depicted in Figure 3.8a-d, it becomes clear that \mathbf{TPM}_{rs} exhibits the least pronounced changes under compression among the three 2D-CORFs considered. Throughout most of the compression range, \mathbf{TPM}_{rs} remains energetically closer to a SM state, also characterized by lower spin densities and a smaller band-gap compared to \mathbf{PTM}_{rs} and $\mathbf{PTM}_{\text{acetylenic}}$. Remarkably, the magnitude of the AFM interaction in \mathbf{TPM}_{rs} is higher compared to \mathbf{PTM}_{rs} and $\mathbf{PTM}_{\text{acetylenic}}$, demonstrating a higher π - π overlap between the αC centers. In this regard, DFT/PBE0 calculations for unstrained \mathbf{TPM}_{rs} yield a J value of -318 meV, which intensifies to -386 meV upon 60% compression (see Table 3.1 for quantitative J values for each 2D-CORF at various compression levels).

For $\mathbf{PTM}_{\text{acetylenic}}$, the gap between its AFMI state and the SM state is significantly larger than in the other two cases along all the compression range and, moreover, the AFM coupling remains much lower in magnitude. In this regard, the DFT/PBE0 analysis reveals a J value of -15 meV for the initial unstrained $\mathbf{PTM}_{\text{acetylenic}}$ structure. While the predicted J value significantly differs from experimental observations (being -46.5 meV), it is important to note that the $\mathbf{PTM}_{\text{acetylenic}}$ experimental characterization is conducted on several finite layers of material,³¹ each one of them with a mix of spin-carrying and non-spin-carrying sites, together with the likely occurrence of structural defects. Overall, we believe that these differences with respect to our extended model can potentially explain the discrepancies in the predicted J with respect to the higher

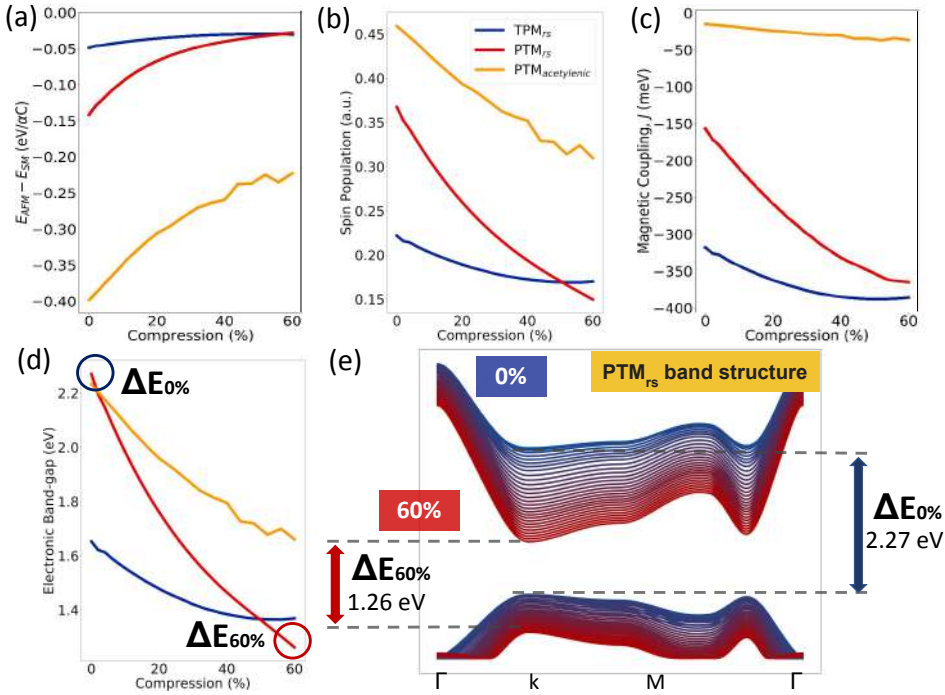


Figure 3.8: Directly extracted results from DFT/PBE0 calculations for out-of-plane compressions ranging from 0% to 60% for each 2D-CORF, namely TPM_{rs} in blue, PTM_{rs} in red and $PTM_{acetylenic}$ in orange. (a) energy difference between the AFMI and semimetallic state (eV per αC), (b) average absolute spin moment per αC site, (c) AFM coupling, J (in meV), (d) AFM electronic bandgap (in eV), and (e) evolution of the AFM band structure of PTM_{rs} from 0% (blue) to 60% (red) compression.

experimental value. Interestingly, with the maximum compression applied in our model, the energy of the AFMI state in $PTM_{acetylenic}$ decreases by 0.18 eV per αC with respect to the SM state (see orange line in Figure 3.8a). Thus, a slight increase in out-of-plane pressure, ranging from 0 to 0.07 GPa (namely 0–60% of compressive stress), applied to the as-synthesized $PTM_{acetylenic}$ could enhance the AFM magnetic coupling strength (potentially up to a J value of -37 meV) and also significantly reduce the electronic band gap by more than 0.5 eV (from 2.2 to 1.7 eV in Figure 3.8d).

Table 3.1: *Extracted values of J , U , t , and U/t for $\mathbf{PTM}_{\text{acetylenic}}$, \mathbf{PTM}_{rs} , and \mathbf{TPM}_{rs} for selected compressions.*

	J (meV)			U (eV)			t (eV)			U/t (adim.)		
	0%	20%	60%	0%	20%	60%	0%	20%	60%	0%	20%	60%
$\mathbf{PTM}_{\text{acetylenic}}$	-15	-24	-37	2.23	1.96	1.66	0.17	0.20	0.22	12.8	9.63	7.39
\mathbf{PTM}_{rs}	-157	-259	-365	2.27	1.76	1.26	0.50	0.50	0.28	4.52	3.50	4.53
\mathbf{TPM}_{rs}	-318	-362	-386	1.65	1.48	1.37	0.60	0.62	0.64	2.75	2.37	2.15

On the other hand, the \mathbf{PTM}_{rs} 2D-CORF exhibits the most pronounced electronic changes when subjected to compression among the three 2D-CORFs we analyzed. In the absence of strain, our DFT/PBE0 calculations indicate that \mathbf{PTM}_{rs} initially displays an AFMI state with an energy gap slightly wider than that of $\mathbf{PTM}_{\text{acetylenic}}$. As compression is applied, the energy gap in \mathbf{PTM}_{rs} diminishes more rapidly, becoming smaller than those observed in both \mathbf{TPM}_{rs} and $\mathbf{PTM}_{\text{acetylenic}}$ when compression exceeds 50%. At such a high level of compression, \mathbf{PTM}_{rs} not only shows the lowest spin density per atom among the three materials, but also closely approaches the energy of the SM state. Similar to \mathbf{TPM}_{rs} , \mathbf{PTM}_{rs} displays substantial AFM coupling, reaching a J value of -365 meV under maximum compression. This significant change in the electronic band structure of \mathbf{PTM}_{rs} with increasing compression is evident in Figure 3.8e, where a noticeable narrowing of the energy gap and an increment of band dispersion (note the local slope near the direct gap) are observed in transitioning from uncompressed (blue curves) to maximum compression (red curves).

Overall, for both \mathbf{TPM}_{rs} and \mathbf{PTM}_{rs} , our results show that at high compression the energy gap between the AFMI and SM states narrows to approximately 40 meV per αC (see 60% blue and red values in Figure 3.8a). Given these small

energy differences between distinct electronic states, caution is warranted when interpreting such results within the framework of a single determinantal method like DFT/PBE0. Our findings imply that under compression, the semimetallic and AFMI states in 2D-CORFs can become energetically degenerate, both potentially contributing to the true multi-configurational ground state.

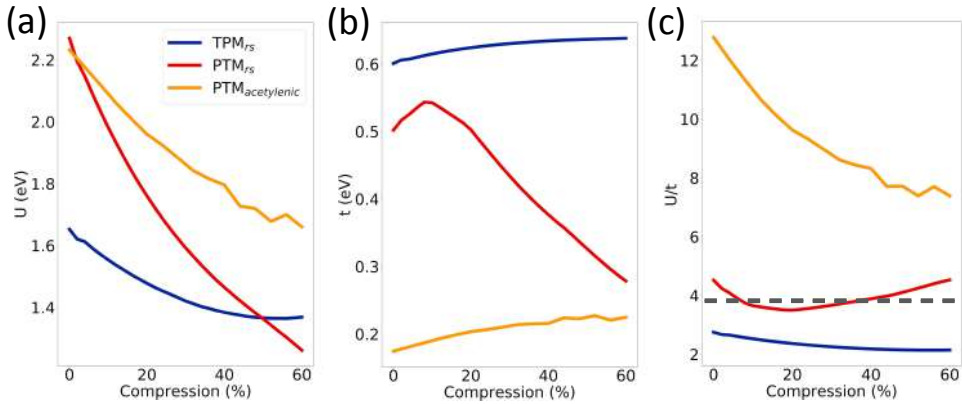


Figure 3.9: Extracted values of (a) U , (b) t , and (c) U/t from our DFT calculations for $PTM_{acetylenic}$ (orange), PTM_{rs} (red), and TPM_{rs} (blue) 2D-CORFs for a 0–60% range of compression. The dashed gray line in (c) indicates the boundary between AFMI and SM phase stability ($U/t = 3.8$) as per the Hubbard model phase diagram for the 2D half-filled hexagonal lattice.^{51,52}

In order to provide further insights, we expanded our analysis to map the DFT/PBE0 band structure results onto a Hubbard-like Hamiltonian, as outlined in the methodology chapter (**Chapter 2**). In this regard, the half-filled Hubbard model,⁵³ with one electron per site, is able to encapsulate the core physics of half-filled systems,^{51,52} thus providing a valuable many-body perspective that can complement the DFT/PBE0 analysis. In short, the Hubbard model specifically addresses the interplay between two key elements: (i) U , which represents an effective localizing, short-range electrostatic repulsion between electrons at individual sites, and (ii) t , a term representing the kinetic energy that character-

izes the ability of electrons to hop among sites. Due to the formal simplicity of such a model, it is particularly suitable for precise many-body methods, such as Quantum Monte Carlo.^{51,52} This method has shown that the ground state phase diagrams of honeycomb lattices, akin to the 2D-CORFs in our study, display SM behavior at lower U/t ratios, and undergo a continuous quantum phase transition to a correlated AFMI state as the U/t ratio increases. Ideally, for exploring transitions between the AFMI and SM phases in half-filled 2D, hexagonally ordered systems, one should be able to tunably access the regimes above and below $U/t \approx 3.8$, respectively.

The procedure described in **Chapter 2** allows for the direct extraction of essential parameters (i.e., t and U) from the DFT/PBE0 band structures at the AFMI state. Essentially, these serve as a reference point to compare the mean-field DFT approach to the many-body treatment. In Figure 3.9, we collect the values of t , U , and U/t for each of the three 2D-CORFs, with some specific values collected in Table 3.1. When **TPM_{rs}** and **PTM_{acetylenic}** undergo compression up to 60%, we observe a respective decrease in U by 0.28 and 0.57 eV. This compression leads to a modest increase in t for both materials (see Table 3.1). As previously discussed, this rise can be attributed to the enhanced π - π overlap between the α C atoms, facilitated by a reduction in the twisting angle, φ . Specifically, for **PTM_{acetylenic}**, the high values of U relative to t result in a pronounced variation in the U/t ratio, which changes from 12.8 to 7.4 across the compression range. Overall, the U/t ratio firmly places **PTM_{acetylenic}** in the AFMI regime according to the Hubbard model phase diagram results,^{51,52} which is in full agreement with the direct inspection of our DFT/PBE0 results discussed above (see Figure 3.8a-d).

However, in the case of $\mathbf{TPM}_{\mathbf{rs}}$, the values of t are notably higher than those for $\mathbf{PTM}_{\mathbf{acetylenic}}$. Moreover, these lead to a U/t ratio that shows minimal change under compression, shifting only from 2.75 to 2.15 eV. Overall, this places $\mathbf{TPM}_{\mathbf{rs}}$ in the SM region of the phase diagram ($U/t < 3.8$, see dashed line in Figure 3.9c), as per the Hubbard-based many-body treatment. Contrarily to the Hubbard picture, our DFT/PBE0 analysis for $\mathbf{TPM}_{\mathbf{rs}}$ reveals that the SM state consistently remains within an energy difference of 48.5 meV per αC above the AFMI state, also narrowing slightly as compression is applied (see Figure 3.8a). Despite the slight mismatch in the outcomes offered by both approaches, they collectively imply that the ground state of $\mathbf{TPM}_{\mathbf{rs}}$ is near the threshold separating the AFMI and SM phases, with compression having only a marginal impact on the electronic properties.

On the other hand, $\mathbf{PTM}_{\mathbf{rs}}$ presents a unique case compared to $\mathbf{PTM}_{\mathbf{acetylenic}}$ and $\mathbf{TPM}_{\mathbf{rs}}$, as it demonstrates a significant non-linear variation in its t values under increasing compression. Initially, for a compression range of 0–10%, t only increases by 0.04 eV, and subsequently decreases by 0.26 eV as compression reaches 60%. Strikingly, this non-linear trend in t mirrors the behavior of the αC – nnC bond distances (as detailed in Figure 3.6b), which is not only in agreement with chemical intuition, but also with the physics governing the quantification of t in the first place (see **Chapter 2**). Regarding U , $\mathbf{PTM}_{\mathbf{rs}}$ registers a higher value than $\mathbf{PTM}_{\mathbf{acetylenic}}$, but it decreases markedly with further compression, even dropping below that of $\mathbf{TPM}_{\mathbf{rs}}$ at 60% compression. As a result, the U/t ratio of the uncompressed $\mathbf{PTM}_{\mathbf{rs}}$ (4.52) dips to its lowest value (3.50) with a modest 20% compression. After that, the U/t ratio incrementally rises to 4.53 at maximum compression. According to the Hubbard model, this suggests that

\mathbf{PTM}_{rs} inherently possesses an AFMI ground state in the uncompressed state, but this can be fine-tuned to approach the boundary between the AFMI and SM phases with compressions around the 20%. Interestingly, achieving such a 20% compression in \mathbf{PTM}_{rs} would only require an estimated out-of-plane pressure of 0.47 GPa.

Although the exact quantitative interpretation is shown to be somewhat contradictory with respect to the approach of choice, both DFT/PBE0 and Hubbard's phase diagram results lead to the same conclusions with respect to the general trends. That is, we demonstrate that TAM-based 2D-CORFs are correlated electronic materials with states that can be tuned by out-of-plane compression. The twistable aryl rings in the 2D-CORFs under study suggest that only modest out-of-plane pressures (<0.5 GPa) can significantly vary their electronic structure, converging towards a graphene-like SM state at modest compression ranges around 20%.

3.2.2. *In-plane strain in TAM-based 2D-CORFs*

From this section on, our attention is shifted to the results concerning the **Publication #2**, which focuses on inducing anisotropic in-plane strain (referred to as ε) to the TAM-based 2D-CORFs considered this far. Our computational methodology is mostly analogous to the one outlined for the out-of-plane compression experiments. However, the tensile stress induced in this study does not require imposing *ad hoc* constraints to certain atoms but, instead, can be readily applied by the progressive elongation of the a cell parameter, as illustrated in Figure 3.10 for $\mathbf{PTM}_{\text{acetylenic}}$. As noted in the figure, we here employed a different unit cell compared to the out-of-plane compression analyses (compare Figure 3.10 with Figure 3.3). In this regard, we opted to apply the tensile stress

such that it is in the same direction as the aryl units covalently coupling the αC centers together. This process could have been induced, otherwise, by the joint elongation of the a and b cell parameters, but would not have allowed us to inspect how the cell parameters change under increasing strain, yielding to similar cell constraints as discussed above for the 2D-CORFs placed between two graphene slabs. Overall, the cell employed herein displays the *correct* atomic orientation with respect to the cell parameters (i.e., the vector connecting the αCs is parallel to the a cell parameter), while keeping the number of atoms per cell as reduced as possible.

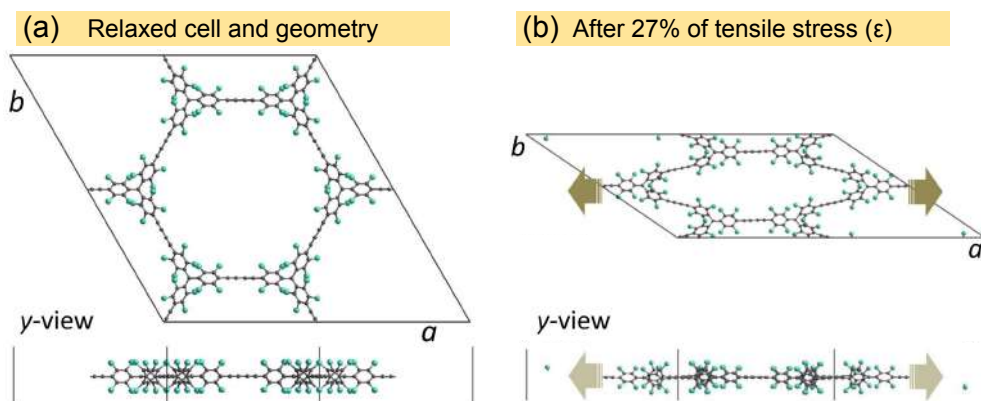


Figure 3.10: (a) Fully relaxed and (b) strained cells for $\text{PTM}_{\text{acetylenic}}$ honeycomb lattice. Top and bottom panels offer different perspectives of the cell. The arrows in the strained structure indicate the direction in which the tensile stress has been applied. Note that it is parallel to the "a" cell parameter.

On a lesser note, it is worth stressing that in the original publication we also considered two extra 2D-CORFs beyond TPM_{rs} , PTM_{rs} and $\text{PTM}_{\text{acetylenic}}$. However, the discussion conducted herein is limited to the latter for consistency. Analogously to the previous sections, the results are split in the description of the structural distortions upon increasing tensile stress (Section 3.2.2.1) and, after that, the associated electronic structure observations are discussed (Section

3.2.2.2). Additionally, it should be acknowledged that the involvement of the PhD candidate in this study began at a later state of the research process. His contributions, while encompassing certain calculations in the following sections (as Young’s modulus calculations, structural analyses and realization of some of the strain calculations), together with the full participation in discussions and manuscript preparation, was initiated after a significant amount of the research had already been completed. That is, most of the optimizations at the initial strains of **TPM_{rs}** and **PTM_{rs}**, the Ab Initio Molecular Dynamics simulations of the former 2D-CORFs, and their corresponding analyses were already preformed.

3.2.2.1. Structural distortions upon progressive strain

Figure 3.11a,b visually summarizes the structural response of **TPM_{rs}** under uniaxial strain. In the absence of strain (i.e. $\varepsilon = 0\%$), all the aryl rings in **TPM_{rs}** are evenly twisted (see $\varphi_1 = \varphi_2 = 33^\circ$ in 3.11a). In contrast, at the maximum strain of $\varepsilon = 28\%$ the aryl rings parallel to the strain direction are flattened (see $\varphi_2 = 4^\circ$ at Figure 3.11b, colored in orange) with respect to the remaining rings, that jointly twist out-of-plane ($\varphi_1 = 75^\circ$, in light blue) due to the strain-induced steric hindrance. Note that the b cell parameter undergoes a severe reduction 12.5%, which is an analogous response as the one discussed previously for compressive strain. Contrarily to compressive stress, where all π - π interactions between α Cs are evenly altered by the deformation, the distortions induced by this kind of anisotropic in-plane strain yield two starkly distinct types of α C π - π interactions. First, there are α C- α C connected by in-plane aryl rings and, secondly, other α C- α C are connected by out-of-plane aryl rings (highlighted in orange and blue respectively). In this regard, these strain-induced conformational changes are likely to favor electron pairing within the flattened

aryl rings (φ_2 in Figure 3.11). In principle, the higher the dihedral angle difference between the two types of aryl rings, $|\varphi_1 - \varphi_2|$, the higher the probability is to induce electron pairing, as discussed on the introduction of this chapter for THIL-like diradicals and further elaborated below.

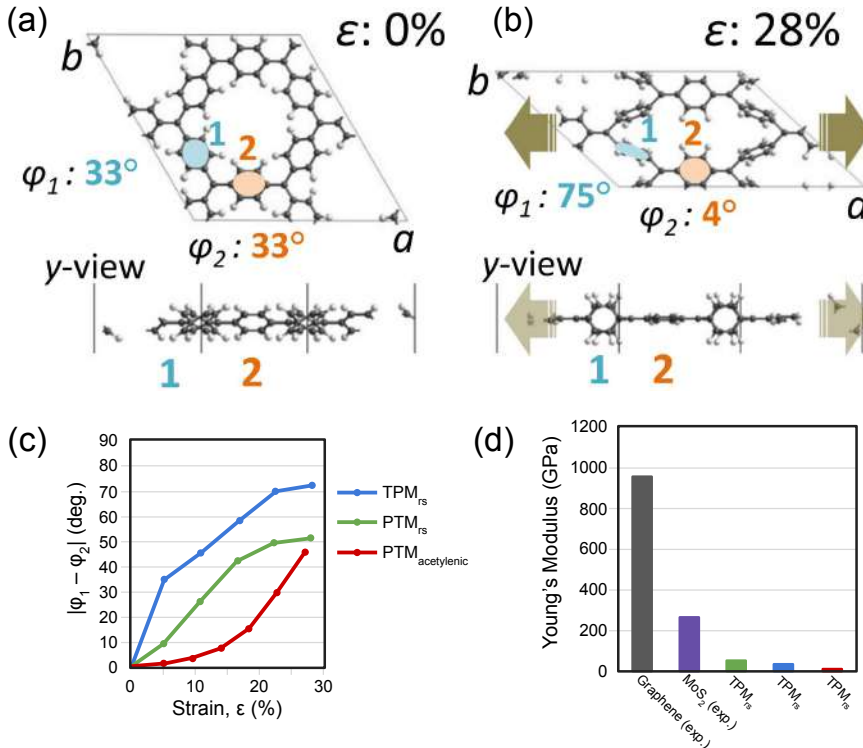


Figure 3.11: (a) Fully relaxed and (b) strained cells for \mathbf{TPM}_{rs} lattice, differentiating between the two types of dihedral twists induced under strain conditions (i.e., φ_1 and φ_2 , colored in light blue and orange, respectively). (c) Dihedral angle difference between the out-of-plane (φ_1) and in-plane (φ_2) twisted aryl rings versus uniaxial strain, ε . (d) Uniaxial Young's modulus (in GPa) calculated for the 2D-CORFs studied in this work and the experimental values for both Graphene and MoS_2 .

Although \mathbf{TPM}_{rs} displays the most significant $|\varphi_1 - \varphi_2|$ variations, \mathbf{PTM}_{rs} and $\mathbf{PTM}_{acetylenic}$ undergo the same conformational changes upon strain, but to a lower extent (see Figure 3.11c). Notably, this observation is in agreement with

the degree of Cl-functionalization in the latter two materials, making their aryl rings more rigidly fixed due to the anticipatedly high steric repulsion between the Cl atoms. Moreover, for **PTM_{acetylenic}**, the acetylenic linkers provide an extra degree of freedom which is not present in the other 2D-CORFs, leading to substantially smaller changes of $|\varphi_1 - \varphi_2|$ with respect to the strain.

Paralleling the analyses conducted for the compression study, we determined the Young’s modulus for **TPM_{rs}**, **PTM_{rs}**, and **PTM_{acetylenic}** 2D-CORFs. Additionally, we compared the obtained values with existing data under similar experimental conditions on graphene and MoS₂. The results demonstrate that the Young’s modulus of all examined 2D-CORFs is considerably lower than that typically observed in inorganic 2D materials, and about 20 times smaller than the one of graphene (approximately 1 TPa for uniaxial strain⁵⁴). This illustrates the relative feasibility with which our 2D-CORFs can be subjected to strain. For instance, single-layer MoS₂ has been uniaxially stretched by approximately 2.5% using various substrate-based techniques.⁵⁵ Given the in-plane Young’s modulus of single-layer MoS₂ (as indicated in purple in 3.11d), these methods can apply in-plane uniaxial tensile stresses of at least 6.5 GPa. When applying similar tensile stress to TPM_{rs}, for example, it would result in a uniaxial strain of around 20%. In more advanced experimental setups, graphene has been stretched uniaxially by nearly 6%, indicating the feasibility of applying higher in-plane tensile stresses of up to approximately 60 GPa.⁵⁶

Overall, it is important to note that Young’s modulus values for **TPM_{rs}**, **PTM_{rs}**, and **PTM_{acetylenic}** are all significantly lower than those for graphene and MoS₂. This suggests that the high stretchability of 2D-CORFs is potentially

due to the twisting of aryl rings, which in the first strain regimes, are the principal changes induced to the structure together with the distortion of the pore.

3.2.2.2. Electronic structure response upon strain

The structural results outlined in the previous section are in line with the current expectations discussed along the introduction of this chapter. That is, increasing the anisotropic in-plane strain also induces the in-plane rotation of certain aryl rings, φ_2 , while some others rotate out-of-plane, φ_1 . From the discussion conducted in the introduction section, we thus expect that such conformational changes alter the relative stability between the AFMI and CSQ electronic structures in favor of the latter. In order to test whether the strain-induced conformational changes depicted in Figure 3.11c lead to a transition from the open-shell AFMI electronic state towards a CSQ diamagnetic state, we have extracted the atom-resolved spin population values on the α C atoms (referred to as $\langle |\mu_{\alpha C}| \rangle$) throughout the full range of strain for all the TAM-based CORFs, by means of employing the Hirshfeld partition scheme.⁵⁷ Although $\langle |\mu_{\alpha C}| \rangle$ should not be taken directly as a face value, in this situation it clearly correlates with the relative stability between the AFMI solution and the competing CSQ electronic structure, as illustrated for the SM state in the previously discussed out-of-plane compression study (see Figure 3.8). The appearance of a significant $\langle |\mu_{\alpha C}| \rangle$ value implies the emergence of low-lying states with non-zero net magnetisation and, thus, it can be interpreted as an indicator of the existence of a spin-polarized state for a particular 2D-CORF and strain condition.

In this regard, Figure 3.12a shows the variation of $\langle |\mu_{\alpha C}| \rangle$ against strain for each studied 2D-CORF. **PTM_{acetylenic}** shows a high and robust $\langle |\mu_{\alpha C}| \rangle$ value throughout stretching, which is a consequence of the strongly localized unpaired

electrons in this material and further affected by the large distance between αC centers and the highly twisted aryl rings (48°). In contrast, for TPM_{rs} and PTM_{rs} , straining the relaxed structure induces a clear transition from the open-shell AFMI solution to the CSQ diamagnetic solution, in which $\langle |\mu_{\alpha\text{C}}| \rangle$ vanishes at moderate amounts of tensile stress. Over the whole set, TPM_{rs} is the most prone to electron pairing and exhibits a full dampening of $\langle |\mu_{\alpha\text{C}}| \rangle$ for strains between 5% to 22% (blue curve in Figure 3.12a).

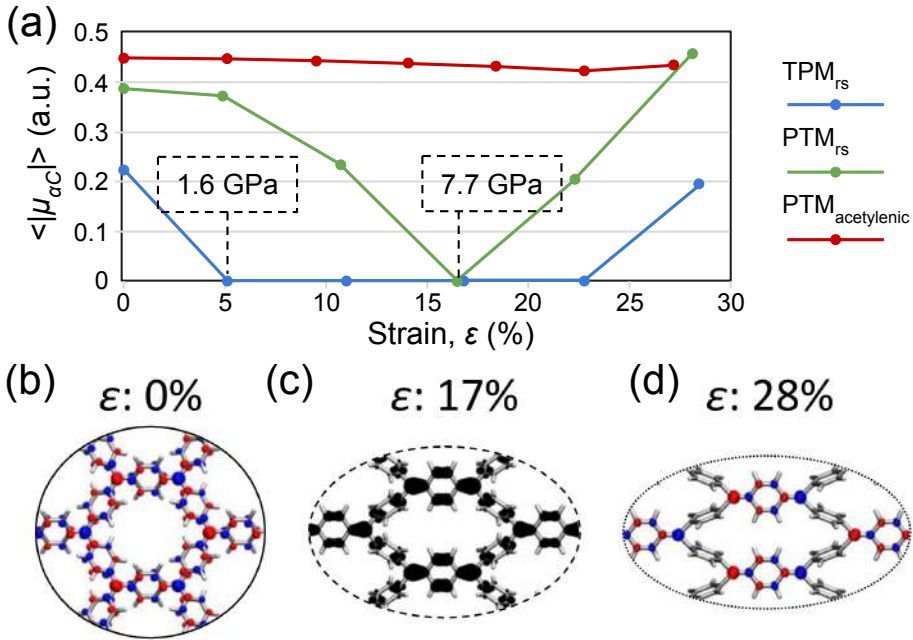


Figure 3.12: (a) Average of the absolute spin population per αC , $\langle |\mu_{\alpha\text{C}}| \rangle$, for each studied 2D-CORF versus uniaxial strain. Estimated in-plane tensile stresses required to induce the AFM-to-quinoidal transition are indicated for TPM_{rs} and PTM_{rs} . (b) Spin density iso-surface (spin up: blue; spin down: red) for TPM_{rs} when relaxed ($\epsilon = 0\%$, AFMI state). (c) Highest occupied crystal orbital density (black) for the semi-strained conformation of TPM_{rs} ($\epsilon = 17\%$, CSQ state). (d) Spin density iso-surface for highly strained TPM_{rs} ($\epsilon = 28.5\%$, AFMI state).

For uniaxial strains above 25%, $\langle |\mu_{\alpha\text{C}}| \rangle$ rises again, which is a consequence of not only the associated αC -nnC bond lengthening (observed in the compression

study, see Figure 3.6), but also due to an apparent twist of the through-bond aryl orientation with respect to the $\alpha\text{C}-\text{nnC}$ bond direction. Likewise, **PTM_{rs}** also shows a complete depletion of $\langle |\mu_{\alpha\text{C}}| \rangle$, which is induced at a higher strain than for **TPM_{rs}** and retained for a smaller range of strains. To further illustrate the general findings, Figure 3.12b-d, shows the transition from the AFMI open-shell solutions in **TPM_{rs}** from $\varepsilon = 0\%$ to $\varepsilon = 28\%$ (see the spin-density isosurface in Figure 3.12b,d), via the CSQ state found at $\varepsilon = 17\%$ (see the highest-occupied crystal orbital colored in black at Figure 3.12c).

The research presented here establishes the feasibility of manipulating the pairing of π -conjugated electrons in TAM-based 2D-CORFs through the application of external uniaxial strain. To substantiate the reliability of these conformational changes, Ab Initio Molecular Dynamics (AIMD) simulations were performed at 300 K for **TPM_{rs}** and **PTM_{rs}** under various strain conditions.

While the attached publication at the end of this chapter also illustrates the findings derived from these simulations, it should be noted that the direct involvement of the PhD candidate in this particular aspect was somewhat limited to the discussion of the results (i.e., the candidate did not carry out the simulations). That said, the results and discussions referring to the AIMD simulations are here summarized briefly to provide a comprehensive overview of the full study.

We found that states previously characterized at 0 K remain robust at 300 K (see Figure 3.13). This persistence is observed despite the presence of thermal vibrations causing fluctuations in the ring torsions. Further insights were gained by examining the $\langle |\mu_{\alpha\text{C}}| \rangle$ values during AIMDs under different strain conditions for both **TPM_{rs}** and **PTM_{rs}** 2D-CORFs. In their relaxed structure ($\varepsilon = 0\%$),

all 2D-CORFs exhibit non-zero $\langle |\mu_{\alpha C}| \rangle$ values (Figure 3.13a). This observation is particularly noteworthy for \mathbf{TPM}_{rs} , where the open-shell AFMI solution is only slightly higher in energy at 0 K compared to the CSQ and SM states.²⁸ Under partial uniaxial strain ($\varepsilon = 17\%$), \mathbf{TPM}_{rs} demonstrates an effective electron pairing, evidenced by the near-complete dampening of $\langle |\mu_{\alpha C}| \rangle$, with minimal impact emerging from thermal fluctuations through time (Figure 3.13b).

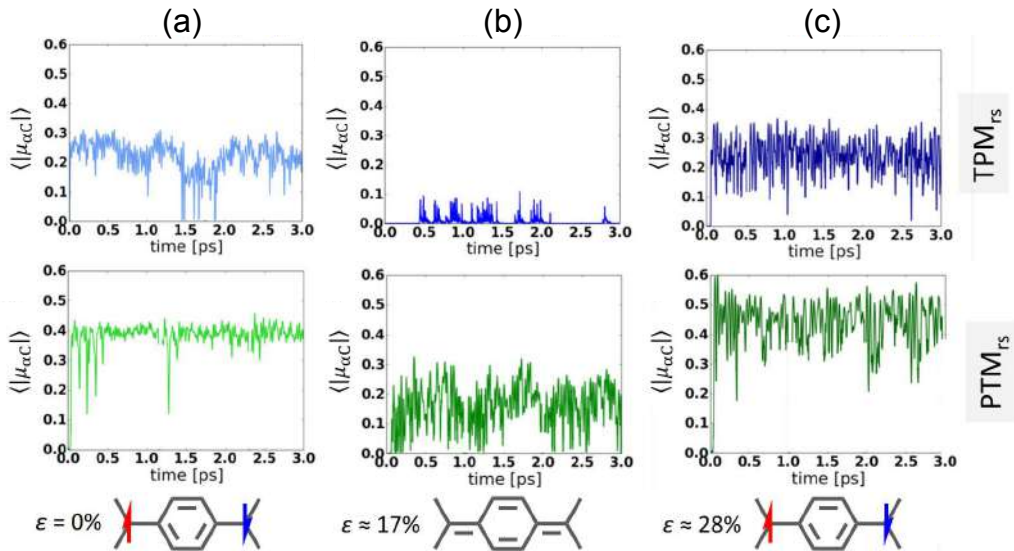


Figure 3.13: Time-resolved evolution of the αC averaged absolute spin population, $\langle |\mu_{\alpha C}| \rangle$, during 3 ps of AIMD simulation at 300 K for (a) 0%, (b) 17% and (c) 28% uniaxial strains for \mathbf{TPM}_{rs} and \mathbf{PTM}_{rs} , colored in blue and green, respectively. The chemical sketches at the bottom of the figure represent the resulting dominant valence bond form for each strain.

Likewise, \mathbf{PTM}_{rs} also shows reduced $\langle |\mu_{\alpha C}| \rangle$ values at $\varepsilon = 17\%$ compared to the unstrained simulation. However, the variance of $\langle |\mu_{\alpha C}| \rangle$ is clearly higher than in \mathbf{TPM}_{rs} under similar conditions (compare the 0.0–0.1 $\langle |\mu_{\alpha C}| \rangle$ range spanned by \mathbf{TPM}_{rs} to the 0.0–0.3 spanned by \mathbf{PTM}_{rs} in Figure 3.13b). We believe this higher variation is associated with an increased noise from thermal vibrations, mostly associated with the lower amount of aryl ring twist due to the bulky

Cl-functionalization. Further stretching of the networks ($\varepsilon = 28\%$) leads to an increase in $\langle |\mu_{\alpha C}| \rangle$ values for both **TPM_{rs}** and **PTM_{rs}** (Figure 3.13c), displaying $\langle |\mu_{\alpha C}| \rangle$ values similar to those resolved at 0 K. Overall, the insights gained from the finite temperature simulations parallel those discussed formerly for the 0 K results, further validating the discussion conducted through the whole chapter.

3.3. Conclusions

This chapter has examined the influence of mechanical stress on the electronic and magnetic properties of three TAM-based 2D Covalent Organic Radical Frameworks: **TPM_{rs}**, **PTM_{rs}**, and **PTM_{acetylenic}**. We initially confirmed their antiferromagnetic insulator (AFMI) ground state and then demonstrated that both isotropic out-of-plane compression and anisotropic in-plane strain can significantly alter their electronic structures. This versatility is highlighted by the transition from an AFMI state to either semimetallic (SM) or closed-shell Quinoidal (CSQ) states under mechanical stress, emphasizing the potential of TAM-based 2D-CORFs as platforms for studying and utilizing correlated electron phenomena in 2-dimensional systems.

Our key finding is that out-of-plane compression induces a transition from an AFMI to a SM state at pressures below 0.5 GPa. This sensitivity to mechanical stress opens up experimental opportunities within practical limits. By employing a hybrid-DFT and Hubbard model approach, we mapped the intricate relationship between the structural adjustments (such as changes in aryl ring twisting angles and C–C bond distances) and the electronic properties under compression. Our analysis suggests that even modest out-of-plane pressures can markedly influence the electronic structure of TAM-based 2D-CORFs. Particularly, applying such pressures to synthesized **PTM_{acetylenic}** is predicted to enhance its AFM cou-

pling strength significantly. Additionally, as \mathbf{TPM}_{rs} molecular fragments have been realized, the synthesis of \mathbf{TPM}_{rs} and \mathbf{PTM}_{rs} 2D-CORFs is potentially feasible. Therefore, these materials are especially promising, since their electronic structure lies at the frontier of the AFMI–SM phase transition (displaying low U/t ratios). Moreover, our results reveal a sophisticated mechanism by which uniaxial strain effectively leads to transitions between AFMI and diamagnetic CSQ states. This transition is primarily driven by uneven twists of the aryl ring within the 2D-CORFs. Our analysis indicates that the strain-induced electronic modulation in these 2D-CORFs is robust even at 300 K. This finding is critical, as it suggests practical viability for technological applications, where consistent performance under varying thermal conditions is essential.

In summary, our analyses underscore the remarkable adaptability of TAM-based 2D-CORFs to mechanical stress, either from out-of-plane compression or in-plane strain. These manipulations lead to significant, reversible transitions in their electronic states, highlighting the potential of these materials in advanced technological applications, where dynamic control over electronic properties is crucial.

References

- (1) X. Hu, W. Wang, D. Wang and Y. Zheng, *Journal of Materials Chemistry C*, 2018, **6**, 11232–11242.
- (2) W. Cheng, H. Chen, C. Liu, C. Ji, G. Ma and M. Yin, *VIEW*, 2020, **1**.
- (3) Y. Morita, S. Nishida, T. Murata, M. Moriguchi, A. Ueda, M. Satoh, K. Arifuku, K. Sato and T. Takui, *Nature Materials*, 2011, **10**, 947–951.

- (4) M. Mas-Torrent, N. Crivillers, V. Mugnaini, I. Ratera, C. Rovira and J. Veciana, *J. Mater. Chem.*, 2009, **19**, 1691–1695.
- (5) V. Lloveras, F. Liko, J. L. Muñoz-Gómez, J. Veciana and J. Vidal-Gancedo, *Chemistry of Materials*, 2019, **31**, 9400–9412.
- (6) R. Frisenda, G. D. Harzmann, J. A. Celis Gil, J. M. Thijssen, M. Mayor and H. S. J. van der Zant, *Nano Letters*, 2016, **16**, 4733–4737.
- (7) C. Simão, M. Mas-Torrent, N. Crivillers, V. Lloveras, J. M. Artés, P. Gorostiza, J. Veciana and C. Rovira, *Nature Chemistry*, 2011, **3**, 359–364.
- (8) Y. Okamoto, M. Tanioka, A. Muranaka, K. Miyamoto, T. Aoyama, X. Ouyang, S. Kamino, D. Sawada and M. Uchiyama, *Journal of the American Chemical Society*, 2018, **140**, 17857–17861.
- (9) L. Wang, X. Shi, S. Feng, W. Liang, H. Fu and J. Yao, *CCS Chemistry*, 2022, **4**, 2748–2756.
- (10) M. B. Smith and J. Michl, *Chemical Reviews*, 2010, **110**, 6891–6936.
- (11) P. M. Zimmerman, Z. Zhang and C. B. Musgrave, *Nature Chemistry*, 2010, **2**, 648–652.
- (12) O. Varnavski, N. Abeyasinghe, J. Aragó, J. J. Serrano-Pérez, E. Ortí, J. T. López Navarrete, K. Takimiya, D. Casanova, J. Casado and T. Goodson, *The Journal of Physical Chemistry Letters*, 2015, **6**, 1375–1384.
- (13) I. Alcón and S. T. Bromley, *RSC Advances*, 2015, **5**, 98593–98599.
- (14) M. Gomberg, *Journal of the American Chemical Society*, 1900, **22**, 757–771.
- (15) G. N. Lewis, D. Lipkin and T. T. Magel, *Journal of the American Chemical Society*, 1944, **66**, 1579–1583.
- (16) Z. Chen, Y. Li and F. Huang, *Chem*, 2021, **7**, 288–332.
- (17) A. H. Maki, R. D. Allendoerfer, J. C. Danner and R. T. Keys, *Journal of the American Chemical Society*, 1968, **90**, 4225–4231.
- (18) O. Armet, J. Veciana, C. Rovira, J. Riera, J. Castaner, E. Molins, J. Rius, C. Miravittles, S. Olivella and J. Brichfeus, *The Journal of Physical Chemistry*, 1987, **91**, 5608–5616.

- (19) J. Thiele and H. Balhorn, *Berichte der deutschen chemischen Gesellschaft*, 1904, **37**, 1463–1470.
- (20) Y. Tobe, in *Physical Organic Chemistry of Quinodimethanes*, Springer International Publishing, 2018, pp. 107–168.
- (21) Z. Zeng, X. Shi, C. Chi, J. T. López Navarrete, J. Casado and J. Wu, *Chemical Society Reviews*, 2015, **44**, 6578–6596.
- (22) J. J. Dressler, M. Teraoka, G. L. Espejo, R. Kishi, S. Takamuku, C. J. Gómez-García, L. N. Zakharov, M. Nakano, J. Casado and M. M. Haley, *Nature Chemistry*, 2018, **10**, 1134–1140.
- (23) J. M. Anglada, J. Poater, I. d. P. R. Moreira and J. M. Bofill, *The Journal of Organic Chemistry*, 2023, **88**, 8553–8562.
- (24) P. Ravat and M. Baumgarten, *Physical Chemistry Chemical Physics*, 2015, **17**, 983–991.
- (25) Y. Su, X. Wang, L. Wang, Z. Zhang, X. Wang, Y. Song and P. P. Power, *Chemical Science*, 2016, **7**, 6514–6518.
- (26) N. M. Shishlov, *Russian Chemical Reviews*, 2006, **75**, 863–884.
- (27) C. Wentrup, M. J. Regimbald-Krnel, D. Müller and P. Comba, *Angewandte Chemie International Edition*, 2016, **55**, 14600–14605.
- (28) I. Alcón, D. Reta, I. d. P. R. Moreira and S. T. Bromley, *Chemical Science*, 2017, **8**, 1027–1039.
- (29) I. Alcón, F. Viñes, I. d. P. R. Moreira and S. T. Bromley, *Nature Communications*, 2017, **8**.
- (30) X.-H. Cao, D. Wu, J. Zeng, N.-N. Luo, W.-X. Zhou, L.-M. Tang and K.-Q. Chen, *Applied Physics Letters*, 2021, **119**.
- (31) S. Wu, M. Li, H. Phan, D. Wang, T. S. Herng, J. Ding, Z. Lu and J. Wu, *Angewandte Chemie International Edition*, 2018, **57**, 8007–8011.
- (32) Y. Yang, C. Liu, X. Xu, Z. Meng, W. Tong, Z. Ma, C. Zhou, Y. Sun and Z. Sheng, *Polymer Chemistry*, 2018, **9**, 5499–5503.

-
- (33) Y. Jiang, I. Oh, S. H. Joo, O. Buyukcakir, X. Chen, S. H. Lee, M. Huang, W. K. Seong, J. H. Kim, J.-U. Rohde, S. K. Kwak, J.-W. Yoo and R. S. Ruoff, *ACS Nano*, 2019, **13**, 5251–5258.
- (34) K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva and A. A. Firsov, *Science*, 2004, **306**, 666–669.
- (35) S. Thomas, H. Li and J.-L. Bredas, *Advanced Materials*, 2019, **31**.
- (36) J.-J. Adjizian, P. Briddon, B. Humbert, J.-L. Duvail, P. Wagner, C. Adda and C. Ewels, *Nature Communications*, 2014, **5**.
- (37) G. Yang, L. Li, W. B. Lee and M. C. Ng, *Science and Technology of Advanced Materials*, 2018, **19**, 613–648.
- (38) X. Du, I. Skachko, A. Barker and E. Y. Andrei, *Nature Nanotechnology*, 2008, **3**, 491–495.
- (39) Z. Li, T. Y. Gopalakrishna, Y. Han, Y. Gu, L. Yuan, W. Zeng, D. Casanova and J. Wu, *Journal of the American Chemical Society*, 2019, **141**, 16266–16270.
- (40) C.-H. Liu, Y. Sakai-Otsuka, P. Richardson, M. R. Niazi, E. Hamzehpoor, T. Jadhav, A. Michels-Gualteri, Y. Fang, M. Murugesu and D. F. Perepichka, *Cell Reports Physical Science*, 2022, **3**, 100858.
- (41) I. Y. Zhang, X. Ren, P. Rinke, V. Blum and M. Scheffler, *New Journal of Physics*, 2013, **15**, 123033.
- (42) V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler, *Computer Physics Communications*, 2009, **180**, 2175–2196.
- (43) A. S. Rodin, A. Carvalho and A. H. Castro Neto, *Physical Review Letters*, 2014, **112**.
- (44) M. Yankowitz, S. Chen, H. Polshyn, Y. Zhang, K. Watanabe, T. Taniguchi, D. Graf, A. F. Young and C. R. Dean, *Science*, 2019, **363**, 1059–1064.
- (45) M. Yankowitz, J. Jung, E. Laksono, N. Leconte, B. L. Chittari, K. Watanabe, T. Taniguchi, S. Adam, D. Graf and C. R. Dean, *Nature*, 2018, **557**, 404–408.
- (46) J.-U. Lee, D. Yoon and H. Cheong, *Nano Letters*, 2012, **12**, 4444–4448.
- (47) R. R. Sumathi, *ECS Journal of Solid State Science and Technology*, 2021, **10**, 035001.

- (48) F. Ronci, S. Colonna, R. Flammini, M. De Crescenzi, M. Scarselli, M. Salvato, I. Berbezier, H. Vach and P. Castrucci, *Applied Surface Science*, 2023, **611**, 155763.
- (49) Y. Gao, S. Kim, S. Zhou, H.-C. Chiu, D. Nélías, C. Berger, W. de Heer, L. Polloni, R. Sordan, A. Bongiorno and E. Riedo, *Nature Materials*, 2015, **14**, 714–720.
- (50) M. H. Garner, R. Hoffmann, S. Rettrup and G. C. Solomon, *ACS Central Science*, 2018, **4**, 688–700.
- (51) Y. Otsuka, S. Yunoki and S. Sorella, *Physical Review X*, 2016, **6**.
- (52) F. F. Assaad and I. F. Herbut, *Physical Review X*, 2013, **3**.
- (53) J. Hubbard, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 1964, **277**, 237–259.
- (54) C. Lee, X. Wei, J. W. Kysar and J. Hone, *Science*, 2008, **321**, 385–388.
- (55) R. Roldán, A. Castellanos-Gomez, E. Cappelluti and F. Guinea, *Journal of Physics: Condensed Matter*, 2015, **27**, 313201.
- (56) K. Cao, S. Feng, Y. Han, L. Gao, T. Hue Ly, Z. Xu and Y. Lu, *Nature Communications*, 2020, **11**.
- (57) F. L. Hirshfeld, *Theoretica Chimica Acta*, 1977, **44**, 129–138.

Publication #1
2D Hexagonal Covalent Organic
Radical Frameworks as tunable
correlated electron systems

FULL PAPER



2D Hexagonal Covalent Organic Radical Frameworks as Tunable Correlated Electron Systems

Raul Santiago, Isaac Alcón, Jordi Ribas-Arino, Mercè Deumal, Ibério de P. R. Moreira, and Stefan T. Bromley*

Quantum materials hold huge technological promise but challenge the fundamental understanding of complex electronic interactions in solids. The Mott metal–insulator transition on half-filled lattices is an archetypal demonstration of how quantum states can be driven by electronic correlation. Twisted bilayers of 2D materials provide an experimentally accessible means to probe such transitions, but these seemingly simple systems belie high complexity due to the myriad of possible interactions. Herein, it is shown that electron correlation can be simply tuned in experimentally viable 2D hexagonally ordered covalent organic radical frameworks (2D hex-CORFs) based on single layers of half-filled stable radical nodes. The presented carefully procured theoretical analysis predicts that 2D hex-CORFs can be varied between a correlated antiferromagnetic Mott insulator state and a semimetallic state by modest out-of-plane compressive pressure. This work establishes 2D hex-CORFs as a class of versatile single-layer quantum materials to advance the understanding of low dimensional correlated electronic systems.

1. Introduction

Hexagonally ordered single layers (SLs) of threefold connected sp^2 carbon sites (α C centers) possess one singly occupied nonhybridized $2p$ orbital per site. For weak intersite interactions these $2p$ electrons tend to form a gapless semimetallic state with Dirac-like linear dispersive bands, as exemplified by SL graphene.^[1] The ability to tune the electron–electron coupling in graphene would permit access to correlated gapped electronic states, which could open the door to further technological applications. Theoretical

work has suggested that application of high (>15%) in-plane biaxial strain to SL graphene could induce a transition from semimetallicity to a correlated antiferromagnetic Mott insulating (AFMI) state.^[2,3] More recent calculations, however, show that an AFMI phase in strained graphene would likely always be less stable than competing diamagnetic insulating valence bond (or quinoidal) states, which start to emerge at <15% strains,^[4] and that are also likely to play a precursor role in structural failure.^[5] Unlike in-plane strain, out-of-plane compression has recently proven to be a viable tool by which to tune the electronic correlation in bilayer (BL) graphene. Specifically, a recently observed correlated insulating state in twisted BL graphene^[6] was shown to be dynamically tunable by applying up to 2.1 GPa out-of-plane pressure^[7] to induce transitions to new states (e.g., superconducting and magnetic). Although the underlying physics is not yet fully understood, the tunability of twisted BL hexagonal systems offers hope that analogous behavior could be found in simpler, potentially easier to analyze, SL systems. Herein, we show how correlated electronic states could be finely tuned in experimentally viable hexagonal SL carbon-based systems via application of relatively modest (<0.5 GPa) out-of-plane compressive pressure.

Recently, we proposed a new class of sp^2 carbon-based SL materials based on linking trivalent triarylmethyl (TAM) α C-containing radical molecules to form 2D covalent organic radical frameworks (CORFs).^[8,9] Since our prediction, some example 2D CORFs have been experimentally prepared.^[10–12] Herein, we focus on 2D CORFs with hexagonal connectivity based on the graphene-like honeycomb lattice (2D hex-CORFs), which we predicted to exhibit spin-polarized antiferromagnetic (AF) ground states and low energy valence bond quinoidal gapped states, linked by an intervening semimetallic transition state.^[9] Compared to the delocalized semimetallic state, the more stable AF and valence bond states have lower space group symmetries and thus can be described as emerging through symmetry breaking.^[9,13] The first experimental examples of a 2D hex-CORF were indeed found to exhibit an AF ground state.^[10,11] 2D hex-CORFs can be seen as expanded graphene lattices where chemical linkers, rather than external strain, increase the distance between the threefold connected carbon sites. In this respect, 2D hex-CORFs are similar to the hypothetical α -graphyne^[14] and α -graphdiyne,^[15] where the relatively

R. Santiago, Dr. J. Ribas-Arino, Prof. M. Deumal, Dr. I. de P. R. Moreira, Prof. S. T. Bromley
Departament de Ciència de Materials i Química Física & Institut de Química Tècnica i Computacional (IQTCUB)
Universitat de Barcelona
c/ Martí i Franquès 1-11, Barcelona 08028, Spain
E-mail: s.bromley@ub.edu

Dr. I. Alcón
Physical and Theoretical Chemistry Department
Freie Universität Berlin
Arnimallee 22, Berlin 14195, Germany

Prof. S. T. Bromley
Institució Catalana de Recerca i Estudis Avançats (ICREA)
Passeig Lluís Companys 23, Barcelona 08010, Spain

The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adfm.202004584>.

DOI: 10.1002/adfm.202004584

increased distance between sp^2 carbon nodes leads to a reduction in the internode electronic coupling. Unlike these hypothetical planar carbon allotropes, 2D hex-CORFs are based on sterically protected sp^2 carbon nodes (e.g., TAMs) leading to more experimentally realizable materials. TAMs are stable and versatile open-shell species, which exhibit the Kondo effect in molecular junctions,^[16] and have been employed as building blocks for molecular spintronics,^[17] nonvolatile memory platforms,^[18] and nanoporous magnets.^[19] The experimentally synthesized TAM building blocks we consider are also not planar, but exhibit twisted aryl rings whose twist angle largely determines their electronic properties.^[20] In turn, SL TAM-based 2D hex-CORFs exhibit out-of-plane atoms and thus have a small finite thickness. In this computational work, we show that the twist angles of aryl ring linkers in 2D hex-CORFs are highly sensitive to out-of-plane compression, which can thus be used to modulate the internode electronic coupling and thereby tune the electrical, magnetic, and optical properties. We show that these SL materials are 2D correlated systems in which applied out-of-plane pressure can shift the AFMI state towards a transition to a semimetallic state.

Our theoretical exploration of 2D hex-CORFs upon out-of-plane compression has been carried out by means of all-electron density functional theory (DFT) based calculations and a mapping of the DFT results onto a Hubbard-like model Hamiltonian. Previously, we have confirmed that the interlinked electronic and structural properties of individual TAMs can be well captured by DFT-based calculations.^[20] Building on previous work,^[8,9] we further confirm that the properties of TAM-based 2D hex-CORFs can be well captured by suitable DFT calculations, and thereby show how their electronic states can be tuned by compression. DFT is a formally exact method^[21] to determine the electron density of quantum systems without the need for an explicit many body treatment. In practical implementations, DFT-based calculations depend on a description of electron correlation through an approximate exchange–correlation potential (V_{xc}). To improve the accuracy of DFT-based treatments of the electronic structure of correlated systems, it is well established that V_{xc} can be significantly improved through incorporation of exact nonlocal Hartree–Fock exchange (HFE). The resulting hybrid DFT-based methods have become very effective approaches to investigate the electronic structure of different types of systems, ranging through strongly correlated transition metal oxides,^[22–24] organic radicals,^[25] and polyradicals,^[26–28] with a remarkable reliability in describing their ground states and magnetic properties. Hybrid functionals provide a computationally effective approach that can be applied in periodic schemes, and which give a good description of magnetic properties for many classes of compounds. Although quantitative results can be sensitive to the exact percentage of HFE employed, most hybrid functionals use between 20% and 50% HFE and tend to provide qualitatively similar results. Our DFT-based calculations employ 25% HFE through use of the PBE0 functional,^[29] to provide a self-consistent description of all electrons in our studied systems.

The use of model Hamiltonians to describe interactions between only those electrons in the system that are thought to be most important is a complementary approach to DFT. The half-filled (i.e., one electron per site) Hubbard model on the

honeycomb lattice^[30] can thus be used to describe the essential electronic structure of SL hexagonal frameworks of sp^2 carbon sites. This model describes the interplay between: i) a localizing effective short-range electrostatic repulsion between electrons on single sites (U), and ii) a kinetic energy term (t), which describes the capacity of electrons to move between sites. Due to the formal simplicity of such models, they are more amenable to being solved through accurate many body methods. Such approaches have revealed that the ground state phase diagram of this model exhibits graphene-like semimetallicity for smaller U/t ratios, and a continuous quantum transition to a correlated AFMI state when U/t is increased.^[31,32] Generally, the Hubbard model is a useful generic model for the essential physics of half-filled 2D systems, but extensions to the model are likely to be necessary for more precise descriptions of specific experimental systems. For example, extended Hubbard-like models with both short- and long-range components of the electron–electron repulsion^[33] can improve the description of different physical realizations of SL half-filled 2D Dirac fermionic systems on the honeycomb lattice (e.g., graphene on different substrates and artificial graphene made from quantum corrals^[34]). However, the presence of heteroatoms or different kinds of hybridized sites (such as in 2D hex-CORFs) leads to highly complex models that are only solvable approximately using mean field or limited correlated schemes.

In this work, we make a link between the DFT-based and model Hamiltonian approaches by extracting approximate U and t values from our PBE0 calculations. In the extreme case of a molecular crystalline lattice of isolated (i.e., nonlinked) TAMs, t is extremely small and U is high leading to a highly robust AFMI state that has been experimentally shown to be resistant to pressures of over 20 GPa.^[35] In ref. [36], relatively high U/t ratios (6.3–16.0) were obtained from DFT calculations of a few unstrained 2D hex-CORF structures. According to the Hubbard model for the 2D honeycomb lattice, these linked structures also have ground states that are firmly within the AFMI part of the phase diagram (i.e., for $U/t > 3.8$).^[31,32] This result is in line with the calculated^[9,36] and experimentally observed^[40] stability of AF phases reported for unstrained 2D hex-CORFs. Due to the approximations in both DFT and model Hamiltonian approaches outlined above, one should be careful not to overinterpret the physical significance of theoretically derived t and U values with respect to any specific material. However, the consistency between the stable states predicted by DFT-based calculations and the Hubbard model phase diagram for high U/t ratios indicate that taking both descriptions into account could be useful for exploring the electronic structure of 2D hex-CORFs with lower U/t ratios. Following the experimentally demonstrated compression-induced tunability of correlated states in twisted BL graphene,^[7] we show that application of out-of-plane compression to 2D hex-CORFs can dramatically reduce their U/t values. We further note that unlike typical magnitudes of out-of-plane compressive pressures used to show observable effects in experiments on other 2D materials (e.g., 1–2.5 GPa for graphene;^[7,37] 0.5–5 GPa for SL MoS_2 ^[38,39]), we predict that relatively small out-of-plane pressures (<0.5 GPa) would be needed for 2D hex-CORFs. Specifically, we show that such modest out-of-plane pressure should enable an AFMI-to-semimetallic transition in suitably fabricated 2D hex-CORFs, thus highlighting

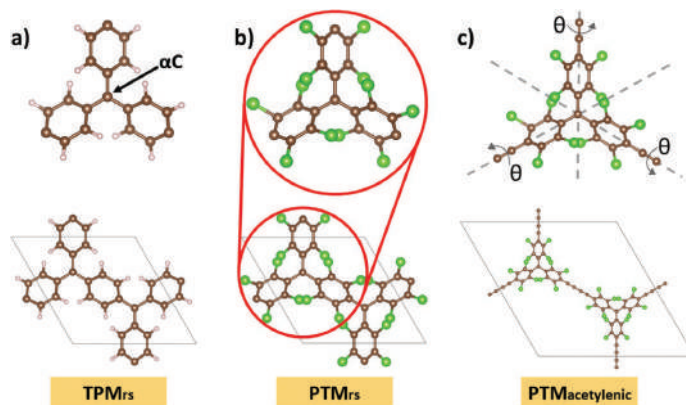


Figure 1. Structures of the three considered 2D hex-CORFs: a) ring-sharing triphenylmethyl (TPM_{rs}) radicals, b) ring-sharing perchlorotriphenylmethyl (PTM_{rs}) radicals, and c) acetylenic coupled PTM radicals ($\text{PTM}_{acylenic}$). Lower: Cuts from each extended 2D hex-CORF structure with the unit cell delimited by a gray line. Upper: Local chemical environment of the corresponding protected triconnected sp^2 carbon sites (αC) for each structure. In (c), we highlight the aryl ring dihedral angles (θ) that are affected by compression in all structures. Atom color key: C: dark brown; H: light brown; Cl: green.

the potential of this class of SL materials for improving our understanding of 2D correlated electronic systems.

2. Results

We consider three examples of TAM-based 2D hex-CORFs based on: a) ring-sharing (rs) triphenylmethyl (TPM) radicals^[9] - TPM_{rs} , b) ring-sharing perchlorotriphenylmethyl (PTM) radicals^[9] - PTM_{rs} , and c) acetylenic linked PTM radicals as found in a recently synthesized and experimentally characterized material ($\text{PTM}_{acylenic}$)^[10,11] (see **Figure 1**). We note that although TPM_{rs} and PTM_{rs} have not yet been synthesized as extended materials, recently, it has been experimentally demonstrated that one can connect TPMs via ring-sharing in molecules containing a ring of six αC centers (i.e., a TPM_{rs} fragment).^[40] This breakthrough opens the door to the possibility of synthesizing extended 2D hex-CORFs based on ring-sharing TAMs. The structures of all three 2D hex-CORFs are shown in **Figure 1**.

We note that, at variance with purely planar SL systems such as graphene, out-of-plane hydrostatic compression significantly affects the dihedral angles and bond distances in 2D hex-CORFs. These changes govern electron localization in the hexagonal lattice of αC centers in these materials, which remains planar throughout. Other materials with nonplanar layers (e.g., black phosphorous) have attracted interest due to their electronic response upon compression.^[41] In black phosphorous, compression affects the coordinates of all atoms in the puckered layers, which affect band dispersion/ordering, which, in turn, can cause a semiconductor-metal transition at 1.2 GPa. Unlike in 2D hex-CORFs, however, this transition is not due to electron correlation effects.

For each of our considered 2D hex-CORFs, we first optimize the cell parameters and all atomic positions of each system to obtain an unstrained reference point. Using an approach

established in other studies modeling compressed SL materials,^[39,42] we apply compressive stress to each 2D hex-CORF by employing selected out-of-plane (i.e., z -direction) structural constraints while allowing for full relaxation of the 2D lattice and unconstrained atoms (see the Experimental Section and the Supporting Information for details). With respect to the original z -direction thickness, we apply compression to each system such that each 2D hex-CORF is compressed by a maximum of 60% (i.e., to 40% of the original thickness). For each material, we have calculated the perpendicular (i.e., in a direction normal to the plane of the layer)^[43] Young's modulus (see the Supporting Information for details) thus enabling us to provide an estimate of the required out-of-plane pressure to induce any percentage of compression. We find that the perpendicular Young's moduli (TPM_{rs} : 5.1 GPa; PTM_{rs} : 13.8 GPa; $\text{PTM}_{acylenic}$: 0.74 GPa) imply that modest pressures (i.e., 0.9, 1.2 and 0.07 GPa respectively) would be needed to achieve the full range of reported compressions (see **Figure 2b**).

In **Figure 2a,b** the main compression-induced structural responses are shown for each 2D hex-CORF. Upon compressing the unstrained materials, the dihedral angles of the aryl rings with respect to the αC centers decrease (i.e., ring flattening) in a linear manner, in line with the applied constraints. The main effect of ring flattening is to increase π - π orbital overlap between αC centers and nearest neighbor carbon (nnC) atoms. The increased π - π overlap also leads to an associated increase in spin delocalization,^[8,20] which we show for PTM_{rs} in a FM state in **Figure 2c**. The in-plane αC -nnC bond distances show more complicated nonlinear dependence on out-of-plane compression, although in all cases tend to increase (i.e., bond stretching) for higher compressions due to increasing steric hindrance between aryl rings. In summary, during out-of-plane compression, there is a competition between increased αC -nnC bond strength due to ring flattening and associated

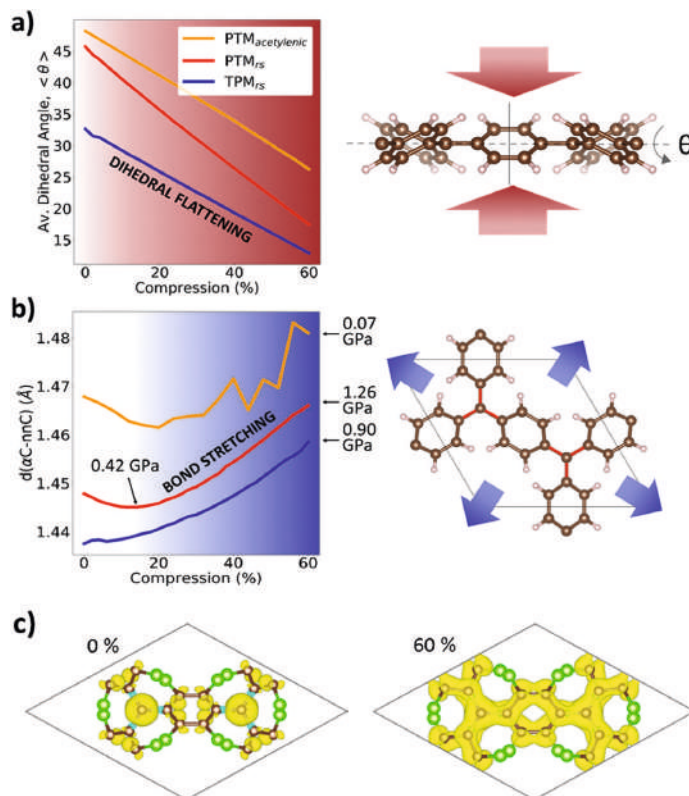


Figure 2. Response of 2D hex-CORFs upon out-of-plane compression. The contributions of ring flattening and in-plane bond stretching in the AF state are indicated by plots of: a) average ring dihedral angle (θ) (see Figure 1), and b) average α C–nnC bond lengths (highlighted in red, right), respectively. The pressures in GPa required for selected compressions in each 2D hex-CORF are indicated in (b). c) Spin density (yellow/cyan) over the bonded C network of PTM_{rs} for 0% (left) and 60% (right) compressions for the FM state. Atom color key: C: dark brown; H: light brown; Cl: green.

increased π – π overlap, and increasing steric hindrance between aryl rings, which tends to expand the in-plane α C–nnC bonds.

In TPM_{rs}, the initial average aryl ring dihedral angle (θ) is the lowest of all considered materials ($\approx 32.6^\circ$) due to the relatively low steric hindrance between the H-functionalized aryl rings.^[20] At the maximum considered compression, ring flattening leads to a (θ) value of only 12.9° . The in-plane α C–nnC bond lengths in TPM_{rs} start at just below 1.44 Å. This value is slightly higher than C–C bond length in graphene (1.42 Å), in line with the conjugated C–C bonding throughout the TPM_{rs} network. With increasing compression, this bond length increases monotonically with a gradually increasing rate up to 1.458 Å at 60% compression.

For PTM_{rs}, the more sterically hindered Cl-functionalized aryl rings induce an initial (θ) value of 45.7° , which reduces to 17.5° at maximum compression. PTM_{rs} has the same topo-

logically connected carbon network but with higher (θ) values that reduce the π – π overlap, leading to longer average α C–nnC bond lengths than in TPM_{rs}. However, unlike TPM_{rs}, for 0–25% compression, the α C–nnC distances first decrease and then increase with increasing aryl ring flattening. This nonmonotonic behavior can be understood due to the relatively reduced π – π overlap in the longer relatively strained α C–nnC bonds in uncompressed PTM_{rs}. Unlike in uncompressed TPM_{rs} where the conjugation is relatively high and the initial α C–nnC bonds are quite short, in PTM_{rs}, ring flattening significantly increases the π – π overlap leading to a strengthening and shortening of these bonds. For PTM_{acetylenic}, the aryl rings are also Cl-functionalized and the initial (θ) value (48.2°) is thus very similar to that of PTM_{rs}. However, the π -conjugation through the longer acetylenic linkers is relatively lower and the initial average α C–nnC bond lengths are considerably longer (1.468 Å). Like

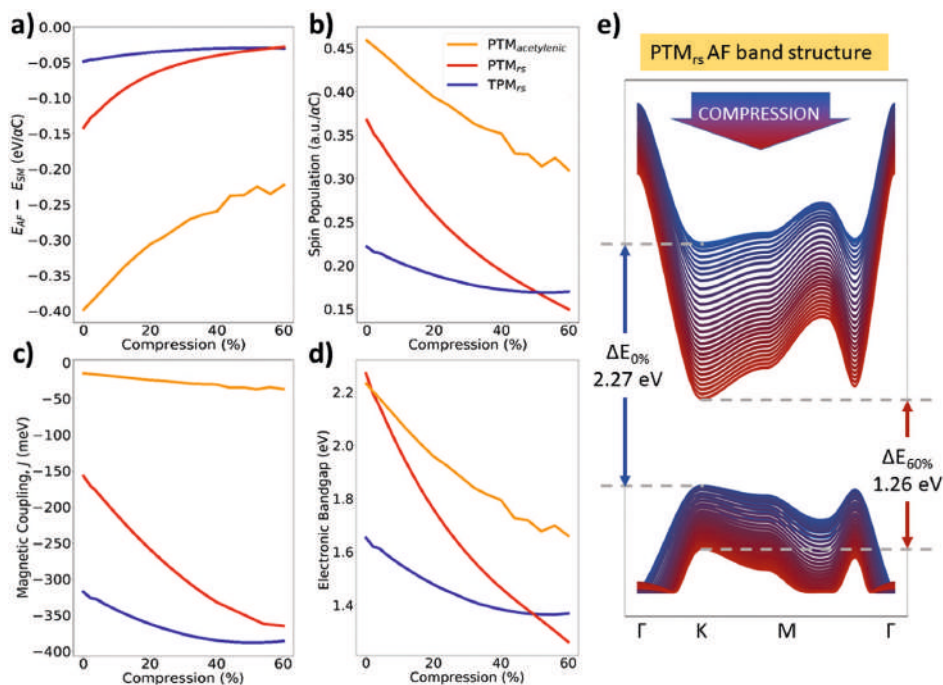


Figure 3. Directly extracted results from DFT-based calculations for out-of-plane compressions from 0% to 60% for each 2D hex-CORF: a) energy difference between the AFMI and semimetallic state (eV per αC), b) average absolute spin moment per αC site, c) AF coupling constant, J (meV), d) AF electronic bandgap (eV), and e) evolution of the AF band structure of PTM_{rs} from 0% (blue) to 60% (red) compression.

PTM_{rs} , compression also leads to an initial increase in π - π overlap and a corresponding dip in the average αC -nnC distances. At higher compression $PTM_{acetylenic}$ also shows significant fluctuations in the average αC -nnC distances. These C-C bond distance fluctuations are observed throughout the conjugated in-plane bonded carbon network of this material (see the Supporting Information) and are thus likely to be linked to variations in π - π overlap. We note that energetically competing spiral conjugated π - π overlap possibilities have been reported in acetylenic chains,^[44] which would sensitively depend on $\langle \theta \rangle$. We also note that these small bond length fluctuations do not appear to indicate that the material is failing structurally, as can be confirmed through the small effect they have on the near-harmonic energy versus compression relation, implying elastic behavior for this compression range (see the Supporting Information).

The structural changes arising from compression induce a significant electronic response in the three 2D hex-CORFs. In all cases and for all considered compressions, our DFT calculations predict a gapped AF ground state. However, with increasing compression for all 2D hex-CORFs the AF state is progressively energetically destabilized with respect to the

semimetallic state (see Figure 3a). This tendency can also be measured by the average absolute spin moment per αC , the AF coupling constant (J), and the electronic bandgap (ΔE), see the Experimental Section and the Supporting Information for details. In all three 2D hex-CORFs, we find that these three indicators decrease with increasing compression (see Figure 3b-d).

With respect to the parameters plotted in Figure 3a-d, TPM_{rs} shows the mildest changes of all three 2D materials with respect to compression. Although TPM_{rs} is the closest energetically to being semimetallic for most of the compression range, with correspondingly smaller spin densities and a relatively small bandgap, the AF magnetic coupling in TPM_{rs} is the largest of all considered materials. In unstrained TPM_{rs} , our PBE0 calculations predict a J value of -318 meV. We note that this value appears to be fairly robust with respect to the percent of HFE used in the functional (see Table S2 in the Supporting Information). At 60% compression, J reaches -386 meV (see also Table 1 for selected J values for each 2D hex-CORF). Recently, a singlet-triplet energy gap of 248 meV was measured for a synthesized molecular fragment of the TPM_{rs} structure containing six spin $1/2$ αC sites in a ring. Using the exact solution for an AF hexagonal Heisenberg system of $S = 1/2$

Table 1. Extracted values of J , U , t , and U/t for $\text{PTM}_{\text{acetylenic}}$, PTM_{rs} , and TPM_{rs} for selected compressions. The U/t values are shaded according to the corresponding ground state on the Hubbard model phase diagram for the 2D half-filled hexagonal lattice: AFMI: orange; semimetallic: blue.

	J [meV]			U [eV]			t [eV]			U/t		
	0%	20%	60%	0%	20%	60%	0%	20%	60%	0%	20%	60%
$\text{PTM}_{\text{acetylenic}}$	-15	-24	-37	2.23	1.96	1.66	0.17	0.20	0.22	12.8	9.63	7.39
PTM_{rs}	-157	-259	-365	2.27	1.76	1.26	0.50	0.50	0.28	4.52	3.50	4.53
TPM_{rs}	-318	-362	-386	1.65	1.48	1.37	0.60	0.62	0.64	2.75	2.37	2.15

spin particles in a 6-membered ring, in which $(E_S - E_T) = 1.369 J$,^[45] we obtain $J = -197$ meV for this experimental system. We note that the average dihedral angle in this molecule is $\approx 7^\circ$ higher than in our extended TPM_{rs} material. In previous work, we showed that the average dihedral angle of aryl rings in TAM-based 2D CORFs determines the energy difference between AF and FM solutions.^[8] In order to compare with experiment, we linearly extrapolate the J versus compression curve for TPM_{rs} (to $\approx 20\%$ compression), to better match the average dihedral angle value of the TPM-based ring molecule. In line with experiment, we find $J \approx -200$ meV for TPM_{rs} under these conditions providing support for the suitability of our hybrid DFT approach for treating these systems.

For $\text{PTM}_{\text{acetylenic}}$, the AF state shows the largest energy difference with respect to the semimetallic state for all considered compressions, but with relatively weak AF coupling. For unstrained $\text{PTM}_{\text{acetylenic}}$, we obtain $J = -15$ meV. We note that our $\text{PTM}_{\text{acetylenic}}$ model is an infinite symmetric SL with an unpaired spin at every αC site. However, experimental $\text{PTM}_{\text{acetylenic}}$ samples consist of a few finite-sized layers, each with a proportion of nonspin carrying sites and each likely to exhibit some structural defects. These differences may explain why the fitting of experimental susceptibility data leads to a relatively larger J value (-46.5 meV) than we obtain.^[9] For this system, the highest considered compression destabilizes the dominant AF state by 0.18 eV per αC . Recently, it was proposed that by replacing the PTM_{rs} nodes in the $\text{PTM}_{\text{acetylenic}}$ structure with oxo-bridged planar TAMs, one could increase the calculated AF coupling strength by approximately four times.^[36] We note that oxo-bridged planar TAMs are relatively unstable to reaction with oxygen,^[46] unlike TAMs, which are kinetically stabilized by sterically hindering access to the central αC (e.g., PTMs and TPMs). As such, 2D hex-CORFs based on oxo-bridged planar TAMs are unlikely to be amenable to experimental synthesis. Herein, we predict that application of small out-of-plane compressive pressures (0 – 0.07 GPa) to the stable as-synthesized $\text{PTM}_{\text{acetylenic}}$ structure could achieve a $2.5\times$ increase in AF magnetic coupling strength ($J = -37$ meV) together with a significant decrease (>0.5 eV) in the magnitude of the electronic bandgap (see Figure 3). For $\text{PTM}_{\text{acetylenic}}$ at high compression, we observe some fluctuations in all the measured electronic parameters (Figure 3a–d), which is also observed in the in-plane C–C bond distances (see Figure 2 and the Supporting Information). As noted above, these variations are likely linked to different θ -dependent π - π overlap possibilities in the acetylenic bridges.

Of the three considered 2D hex-CORFs, PTM_{rs} shows the most dramatic electronic response with respect to compression. For no strain, our calculations predict PTM_{rs} to exhibit an AF state with a slightly larger energy gap than $\text{PTM}_{\text{acetylenic}}$. With

increasing strain, the PTM_{rs} energy gap rapidly decreases to eventually be lower than that of both TPM_{rs} and $\text{PTM}_{\text{acetylenic}}$ for $>50\%$ compression. Moreover, for this high compression, PTM_{rs} has the lowest spin density per atom and energetically is the closest to being semimetallic. As for TPM_{rs} , the AF coupling is significant in this material, with $J = -365$ meV for the highest compression. The evolution of the electronic band structure of PTM_{rs} with increasing compression is shown in Figure 3e, where the decrease in the energy gap and increase in overall band dispersion is clearly seen. For high compression, the difference in energy between the AF and the semimetallic state for both TPM_{rs} and PTM_{rs} is calculated to be only ≈ 30 meV per αC . Such marginal energy differences between distinct electronic states should be treated with caution within a single determinantal picture such as DFT. These results suggest that compression of 2D hex-CORFs can make the semimetallic and AFMI states almost degenerate in energy, and both would contribute to the true ground state.

In addition to directly inspecting the nature of the lowest energy state predicted by DFT calculations, we can also extract U and t parameters from our calculations (see the Experimental Section and the Supporting Information for details), which we can use to compare with predictions from the Hubbard model on the honeycomb lattice. In Figure 4, we plot the extracted values of t , U , and U/t for all three 2D hex-CORFs for all considered degrees of compression (see also Table 1 for selected values). For TPM_{rs} and $\text{PTM}_{\text{acetylenic}}$, compressing the unstrained structures up to 60% leads to a decrease in U of 0.28 and 0.57 eV, respectively. For both these 2D hex-CORFs this compression also leads to a similar increase in t (0.3 – 0.5 eV). For $\text{PTM}_{\text{acetylenic}}$, the extracted U values are relatively high compared to the t values and this entails a corresponding increased sensitivity of the U/t values, which vary from 12.8 to 7.4 . In ref. [24], the smallest reported U/t ratio of 6.3 was obtained in a proposed structure based on oxo-bridged planar-TAMs with vinylenic linkers. In both cases, these U/t ratios firmly place these materials in the AFMI regime according to the Hubbard model phase diagram. For $\text{PTM}_{\text{acetylenic}}$, this picture is fully consistent with a direct inspection of our DFT results (see Figure 3). Ideally, for exploring transitions between the AFMI and semimetallic phases in half-filled 2D-hexagonally ordered systems, one should be able to tunably access the regimes above and below $U/t \approx 3.8$ ^[31,32]

For TPM, the t values are relatively larger than the U values and the corresponding U/t ratio does not change significantly with compression (from 2.75 to 2.15), which places TPM_{rs} in the semimetallic region of the Hubbard phase diagram. Direct inspection of our DFT results for TPM_{rs} indicates that the semimetallic state is always within 48.5 meV per αC of the AF state

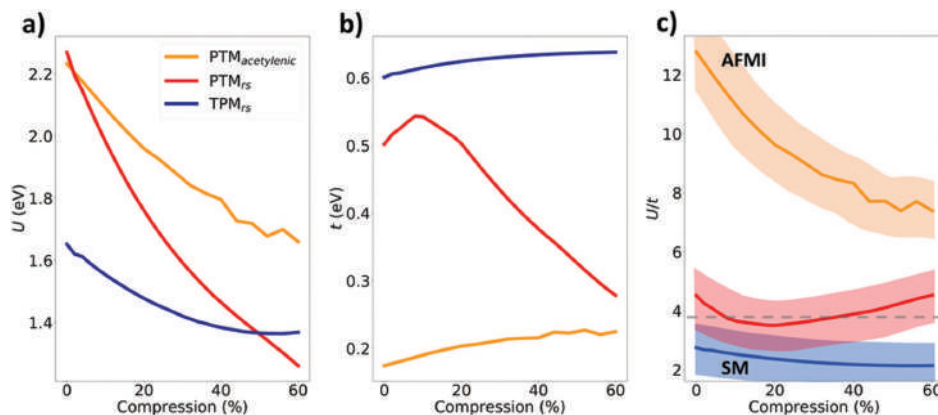


Figure 4. Extracted values of: a) U , b) t , and c) U/t from our DFT calculations for $PTM_{acetylenic}$, PTM_{rs} , and TPM_{rs} 2D hex-CORFs for a 0–60% range of compression. The dashed gray line in (c) indicates the boundary between AFMI and semimetallic (SM) phase stability ($U/t = 3.8$)^[31,32] with respect to the Hubbard model phase diagram for the 2D half-filled hexagonal lattice. The shading around the lines in (c) provides an estimate of the expected variation in U/t values with respect to modest changes in the percent of HFE used in the functional (see Table S2 in the Supporting Information).

and that this energy difference slightly decreases with compression. Although, the quantitative picture provided by these two perspectives differs somewhat, both suggest that TPM_{rs} has a ground state that is quite close to the boundary between AFMI and semimetallic phases and which is only slightly affected by compression.

PTM_{rs} , unlike $PTM_{acetylenic}$ and TPM_{rs} , shows a large and nonmonotonic change in t with respect to increasing compression. The t values for PTM_{rs} lay between those extracted for TPM_{rs} and $PTM_{acetylenic}$ but first increase by 0.04 eV with a 0–10% compression increase and thereafter decrease by 0.26 eV as compression increases to 60%. This nonmonotonic behavior in t is related to the analogous behavior in the αC – nnC bond distances (see Figure 2), which are a structural indicator of the electronic coupling between αC centers. The value of U for uncompressed PTM_{rs} is higher than the corresponding value for $PTM_{acetylenic}$ but rapidly decreases with increasing compression and becomes lower than that of TPM_{rs} for 60% compression. Correspondingly, the U/t ratio for uncompressed PTM_{rs} starts at 4.52, then decreases to a minimum value of 3.50 at 20% compression, and then gradually increases back up to 4.53 at the highest compression. From a Hubbard model perspective, the PTM_{rs} would have an AFMI ground state, which could be tuned to be arbitrarily close to the AFMI–semimetallic phase transition boundary with out-of-plane compression above and below 20%. We note that to achieve 20% compression in PTM_{rs} we estimate that only 0.47 GPa of out-of-plane pressure would be required. A direct inspection of the DFT calculations also suggests that PTM_{rs} exhibits an AF state that can be tuned to be energetically close to the semimetallic phase. However, the latter picture indicates a monotonic progression toward the semimetallic phase with increasing compression. These different tendencies are to be expected due to differences in the methodologies employed. The electronic states in the Hubbard

model are derived from a subset of the actual electrons on a fixed lattice, and effects such as polarization/delocalization of non- αC -based electrons and structural relaxations (e.g., of the linkers), which are described by DFT calculations, are inherently not included. On the other hand, Monte Carlo treatments of the Hubbard model Hamiltonian explicitly describe electronic correlation and attempts to solve it using mean field based approaches (e.g., pure HF) tend to yield lower effective U values than direct many body solutions.^[30] Hybrid DFT functionals attempt to partially bridge the gap between these approaches but results can be sensitive to the percent of HFE employed. In Figure 4c, we include shading to indicate tentative variations in U/t with respect to modest changes in HFE percent.

3. Conclusions

2D hex-CORFs are novel examples of the half-filled 2D honeycomb lattice and thus lie in the same class as graphene and hexagonal quantum corrals.^[33] Theoretical studies using simple but exactly solvable many body model Hamiltonians are still revealing new fundamental physical insights regarding the AFMI–semimetallic transition in such systems.^[31,32] However, even for these structurally uncomplicated systems, such simplified descriptions of their electronic structure can fail to capture essential features found in experiment.^[33] 2D hex-CORFs are relatively chemically complex materials that have a highly nontrivial structural and electronic response to out-of-plane compression. To accurately describe such complexity using a many body Hamiltonian approach would likely require a highly complicated and intractable model. Here, we use an all-electron hybrid DFT approach to describe the full chemical structure and electronic response of our 2D hex-CORFs,

while also extracting parameters relevant to model Hamiltonian descriptions (i.e., U , t , and J). The complementary picture thus obtained shows that compressing suitable 2D hex-CORFs could provide a new means for probing the phase diagram of this important 2D correlated electron system. Specifically, we demonstrate that TAM-based 2D hex-CORFs, like twisted BL graphene, are correlated electronic materials with states that can be tuned by out-of-plane compression. The twistable aryl rings in the SL materials we consider mean that only modest out-of-plane pressures (<0.5 GPa) can significantly vary their electronic structure. This general result can be readily tested as we predict that such pressures applied to the experimentally synthesized PTM_{acetylenic} material should lead to a significant increase in its observed AF magnetic coupling strength. Further, based on synthesized TPM_{is} molecular fragments, we propose that ring-sharing 2D hex-CORFs should be experimentally viable materials and would be particularly promising as they can: i) exhibit ground states laying close to the AFMI–semimetallic transition, and ii) have compression-tunable U/t ratios for low out-of-plane pressures. 2D hex-CORFs are thus promising materials, both from a fundamental point of view to study correlation effects in low dimensional materials and for potential technological applications (e.g., sensors and spintronics/electronics). We hope that our work will inspire further experimental and theoretical work into this fascinating new class of 2D correlated electronic materials.

4. Experimental Section

Unstrained structures of each 2D hex-CORF were obtained via DFT-based optimizations of atomic positions and cell parameters using the PBE0^[29] hybrid functional as implemented in the all-electron FHI-AIMS code.^[47] In all calculations, nonbonded dispersion interactions were included via the TS-vdW approach.^[48] A “light/Tier1” numerical atom-centered orbital (NAO) basis set was used throughout.^[49] This basis set provides results of a similar or higher quality to those obtained with valence triple-zeta plus polarization Gaussian type orbitals^[50] in line with the trends regarding the quality of these NAO basis sets.^[51] The 2D hex-CORF SLs were oriented so that the αC plane was parallel to the x – y plane (with $z = 0$ for the αC plane) and were separated from their periodic images by a vacuum spacing of 40 Å in the z -direction. $6 \times 6 \times 1$ Monkhorst–Pack generated grid meshes of k -points were used for each 2D hex-CORF.

The constrained atom method^[39,42] was used to induce a z -directed out-of-plane compression on the 2D hex-CORFs. In ref. [39], the results from employing this method have been shown to compare very well with experimental results from uniaxial out-of-plane compression on SL MoS₂. Firstly, a set of C atoms was selected both above and below the αC plane for each structure (see the Supporting Information for specific atom sets in each case). To apply compression, the absolute value of the z -coordinates of these constrained atoms was adjusted according to incremental decreases of the dihedral angles of the aryl rings, and then fixed for each increment. The resulting z -constrained structures were then optimized again using the above setup allowing for full relaxation of the unit cell, positions of nonconstrained atoms, and the x and y coordinates of the z -constrained atoms. In the Supporting Information, it is shown that this constrained atom method produces the same structural response as when a 2D hex-CORF is sandwiched within a compressed layered 2D heterostructure. The perpendicular Young's moduli, and thus the external out-of-plane pressure required to induce these compressions, were extracted from the total energy versus unit cell volume curves (see the Supporting Information for details). From

these curves, it could be verified that the 2D hex-CORFs were always in the elastic regime for all considered compressions.

The magnetic coupling constants, J , were evaluated assuming the Heisenberg Hamiltonian whereby the energy difference between AF and FM states is directly related to the Heisenberg spin exchange parameter J_{ij} (see the Supporting Information for further details). The t and U values were extracted from the electronic band structures of each 2D hex-CORF following the procedure detailed in ref. [52] (see also the Supporting Information).

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was supported by the following research grants: MICIUN/FEDER RTI2018-095460-B-I00, PID2019-109518GB-I00 and CTQ2017-87773-P/AEI/FEDER (Spanish government, MINECO), MDM-2017-0767 (“María de Maeztu” program for Spanish Structures of Excellence), and 2017SGR13 and 2017SGR348 (Generalitat de Catalunya, DURSI). R.S. acknowledges a predoctoral FPI grant from MINECO under grant agreement CTQ2017-87773-P/AEI/FEDER. I.A. is grateful for support from the Alexander von Humboldt Foundation. The authors also acknowledge access to supercomputer resources as provided through grants from the Red Española de Supercomputación.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

2D materials, compressive stress, correlated materials, density functional theory calculations, semimetal–insulator transition

Received: May 28, 2020

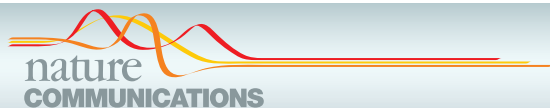
Revised: August 21, 2020

Published online: November 4, 2020

- [1] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, A. A. Firsov, *Science* **2004**, 306, 666.
- [2] S.-H. Lee, S. Kim, K. Kim, *Phys. Rev. B* **2012**, 86, 155436.
- [3] H.-K. Tang, E. Laksono, J. N. B. Rodrigues, P. Sengupta, F. F. Assaad, S. Adam, *Phys. Rev. Lett.* **2015**, 115, 186602.
- [4] S. Sorella, K. Seki, O. O. Brovko, T. Shirakawa, S. Miyakoshi, S. Yunoki, E. Tosatti, *Phys. Rev. Lett.* **2018**, 121, 066402.
- [5] S.-H. Lee, H.-J. Chung, J. Heo, H. Yang, J. Shin, U.-I. Chung, S. Seo, *ACS Nano* **2011**, 5, 2964.
- [6] Y. Cao, V. Fatemi, A. Demir, S. Fang, S. L. Tomarken, J. Y. Luo, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras, R. C. Ashoori, P. Jarillo-Herrero, *Nature* **2018**, 556, 80.
- [7] M. Yankowitz, S. Chen, H. Polshyn, Y. Zhang, K. Watanabe, T. Taniguchi, D. Graf, A. F. Young, C. R. Dean, *Science* **2019**, 363, 1059.
- [8] I. Alcón, D. Reta, I. de P. R. Moreira, S. T. Bromley, *Chem. Sci.* **2017**, 8, 1027.
- [9] I. Alcón, F. Viñes, I. de P. R. Moreira, S. T. Bromley, *Nat. Commun.* **2017**, 8, 1957.

- [10] S. Wu, M. Li, H. Phan, D. Wang, T. S. Herng, J. Ding, Z. Lu, J. Wu, *Angew. Chem., Int. Ed.* **2018**, *57*, 8007.
- [11] Y. Yang, C. Liu, X. Xu, Z. Meng, W. Tong, Z. Ma, C. Zhou, Y. Sun, Z. Sheng, *Polym. Chem.* **2018**, *9*, 5499.
- [12] Y. Jiang, I. Oh, S. Hun, J. Onur, B. Xiong, C. Sun, H. Lee, M. Huang, W. Kyung, S. Jin, H. Kim, J.-U. Rohde, S. Kyu, K. J.-W. Yoo, R. S. Ruoff, *ACS Nano* **2019**, *13*, 5251.
- [13] X. Y. Xu, K. T. Law, P. A. Lee, *Phys. Rev. B* **2018**, *98*, 121406(R).
- [14] B. G. Kim, H. J. Choi, *Phys. Rev. B* **2012**, *86*, 115435.
- [15] X. Niu, X. Mao, D. Yang, Z. Zhang, M. Si, D. Xue, *Nanoscale Res. Lett.* **2013**, *8*, 469.
- [16] R. Frisenda, R. Gaudenzi, C. Franco, M. Mas-Torrent, C. Rovira, J. Veciana, I. Alcon, S. T. Bromley, E. Burzurí, H. S. J. van der Zant, *Nano Lett.* **2015**, *15*, 3109.
- [17] M. Mas-Torrent, N. Crivillers, V. Mugnaini, I. Ratera, C. Rovira, J. Veciana, *J. Mater. Chem.* **2009**, *19*, 1691.
- [18] C. Simão, M. Mas-Torrent, N. Crivillers, V. Lloveras, J. M. Artés, P. Gorostiza, J. Veciana, C. Rovira, *Nat. Chem.* **2011**, *3*, 359.
- [19] D. Maspoch, D. Ruiz-Molina, K. Wurst, N. Domingo, M. Cavallini, F. Biscarini, J. Tejada, C. Rovira, J. Veciana, *Nat. Mater.* **2003**, *2*, 190.
- [20] I. Alcon, S. T. Bromley, *RSC Adv.* **2015**, *5*, 98593.
- [21] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864.
- [22] I. de P. R. Moreira, F. Illas, R. L. Martin, *Phys. Rev. B* **2002**, *65*, 155102.
- [23] F. Tran, P. Blaha, K. Schwarz, P. Novak, *Phys. Rev. B* **2006**, *74*, 155108.
- [24] P. Rivero, I. de P. R. Moreira, F. Illas, *Phys. Rev. B* **2010**, *81*, 205123.
- [25] V. Barone, P. Cimino, *J. Chem. Theory Comput.* **2009**, *5*, 192.
- [26] D. Reta-Mañeru, A. K. Pal, I. D. P. R. Moreira, S. N. Datta, F. Illas, *J. Chem. Theory Comput.* **2014**, *10*, 335.
- [27] D. Cho, K. C. Ko, J. Y. Lee, *Int. J. Quantum Chem.* **2016**, *116*, 578.
- [28] D. Reta, I. de P. R. Moreira, F. Illas, *J. Chem. Theory Comput.* **2016**, *12*, 3228.
- [29] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158.
- [30] S. Sorella, E. Tosatti, *Europhys. Lett.* **1992**, *19*, 699.
- [31] F. F. Assaad, I. F. Herbut, *Phys. Rev. X* **2013**, *3*, 031010.
- [32] Y. Otsuka, S. Yunoki, S. Sorella, *Phys. Rev. X* **2016**, *6*, 011029.
- [33] H.-K. Tang, J. N. Leaw, J. N. B. Rodrigues, I. F. Herbut, P. Sengupta, F. F. Assaad, S. Adam, *Science* **2018**, *361*, 570.
- [34] K. K. Gomes, W. Mar, W. Ko, F. Guinea, H. C. Manoharan, *Nature* **2012**, *483*, 306.
- [35] M. Souto, H. Cui, M. Peña-Alvarez, V. G. Baonza, H. O. Jeschke, M. Tomic, R. Valenti, D. Blasi, I. Ratera, C. Rovira, J. Veciana, *J. Am. Chem. Soc.* **2016**, *138*, 11517.
- [36] S. Thomas, H. Li, J.-L. Bredas, *Adv. Mater.* **2019**, *31*, 1900355.
- [37] M. Yankowitz, J. Jung, E. Laksono, N. Leconte, B. L. Chittari, K. Watanabe, T. Taniguchi, S. Adam, D. Graf, C. R. Dean, *Nature* **2018**, *404*, 557.
- [38] X. Dou, K. Ding, D. Jiang, B. Sun, *ACS Nano* **2014**, *8*, 7458.
- [39] M. Peña-Alvarez, E. del Corro, Á. Morales-García, L. Kavan, M. Kalbac, O. Frank, *Nano Lett.* **2015**, *15*, 3139.
- [40] Z. Li, T. Y. Gopalakrishna, Y. Ha, Y. Gu, L. Yuan, W. Zeng, D. Casanova, J. Wu, *J. Am. Chem. Soc.* **2019**, *141*, 16266.
- [41] A. S. Rodin, A. Carvalho, A. H. Castro Neto, *Phys. Rev. Lett.* **2014**, *112*, 176801.
- [42] T. Hu, J. Dong, *Phys. Rev. B* **2015**, *92*, 064114.
- [43] Y. Gao, S. Kim, S. Zhou, H.-C. Chiu, D. Nélías, C. Berger, W. de Heer, L. Polloni, R. Sordan, A. Bongiorno, E. Riedo, *Nat. Mater.* **2015**, *14*, 714.
- [44] M. H. Garner, R. Hoffmann, S. Rettrup, G. C. Solomon, *ACS Cent. Sci.* **2018**, *4*, 688.
- [45] J. Schnack, in *Quantum Magnetism*, (Eds: U. Schollwöck, J. Richter, D. J. J. Farnell, R. F. Bishop), Lecture Notes in Physics, Vol. 645, Springer, Berlin **2004**, pp. 155–194.
- [46] K. Kato, A. Osuka, *Angew. Chem., Int. Ed.* **2019**, *58*, 8978.
- [47] R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, M. Scheffler, *Comput. Phys. Commun.* **2009**, *180*, 2175.
- [48] A. Tkatchenko, M. Scheffler, *Phys. Rev. Lett.* **2009**, *102*, 073005.
- [49] V. Havu, V. Blum, P. Havu, M. Scheffler, *J. Comput. Phys.* **2009**, *228*, 8367.
- [50] O. Lamiel-García, K. C. Ko, J. Y. Lee, S. T. Bromley, F. Illas, *J. Chem. Theory Comput.* **2017**, *13*, 1785.
- [51] I. Y. Zhang, X. Ren, P. Rinke, V. Blum, M. Scheffler, *New J. Phys.* **2013**, *15*, 123033.
- [52] H. Xiang, C. Lee, H.-J. Koo, X. Gong, M.-H. Whangbo, *Dalton Trans.* **2013**, *42*, 823.

Publication #2
Controlling pairing of
 π -conjugated electrons in 2D
covalent organic radical
frameworks via in-plane strain






ARTICLE

<https://doi.org/10.1038/s41467-021-21885-y>

OPEN

Controlling pairing of π -conjugated electrons in 2D covalent organic radical frameworks via in-plane strain

Isaac Alcón¹, Raúl Santiago², Jordi Ribas-Arino², Mercè Deumal^{1,2}, Ibério de P. R. Moreira² & Stefan T. Bromley^{1,2,3}

Controlling the electronic states of molecules is a fundamental challenge for future sub-nanoscale device technologies. π -conjugated bi-radicals are very attractive systems in this respect as they possess two energetically close, but optically and magnetically distinct, electronic states: the open-shell antiferromagnetic/paramagnetic and the closed-shell quinoidal diamagnetic states. While it has been shown that it is possible to statically induce one electronic ground state or the other by chemical design, the external dynamical control of these states in a rapid and reproducible manner still awaits experimental realization. Here, via quantum chemical calculations, we demonstrate that in-plane uniaxial strain of 2D covalently linked arrays of radical units leads to smooth and reversible conformational changes at the molecular scale that, in turn, induce robust transitions between the two kinds of electronic distributions. Our results pave a general route towards the external control, and thus technological exploitation, of molecular-scale electronic states in organic 2D materials.

¹Institut für Chemie und Biochemie, Physikalische und Theoretische Chemie, Freie Universität Berlin, Berlin, Germany. ²Departament de Ciència de Materials i Química Física & Institut de Química Teòrica i Computacional (IQT-CUB), Universitat de Barcelona, Barcelona, Spain. ³Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. [✉]email: ialcon8@gmail.com; s.bromley@ub.edu

Since the discovery of graphene¹, an increasing number of inorganic 2D materials with a range of physical and chemical properties have been synthesised by top-down approaches (e.g. by exfoliation of layered bulk materials)². Due to their unique combination of extreme thinness and extended planarity, these monolayered systems should represent ideal platforms for tailoring electronic properties through in-plane strain³. However, in practice, 2D inorganic materials possess very high in-plane tensile strengths, which only allow for moderate strains (<5%)³ generating modest electronic changes^{4–6}. In parallel, chemists have developed alternative bottom-up synthesis approaches based on linking arrays of molecular building blocks, to produce 2D covalent organic frameworks (2D-COFs)^{7–9}. These materials are significantly more flexible than their inorganic analogues¹⁰ and allow the integration of the vast structural and electronic versatility of organic molecular chemistry^{9,11} within scalable and robust 2D platforms. Building on previous proposals to take advantage of these traits for applications^{8,12,13}, we show that a particular family of 2D-COFs may provide an ideal basis for highly sensitive strain-control of electronic properties.

Kekulé bi-radicals are promising systems for organic molecular electronics^{14–17} and magnetism¹⁸ due to their intrinsically bi-stable electronic structure whereby two very distinct but energetically close resonant forms compete: namely, the open-shell antiferromagnetic (AFM) and the closed-shell quinoidal diamagnetic states (Fig. 1a). The different character of these two states leads to significant observable differences in the respective associated magnetic¹⁹, optical^{20,21} and structural²² properties. It has been shown that molecular chemical design is an effective way to induce one electronic state or the other¹⁶, where the length of the π -conjugated bridge between radical centres (e.g. a number of benzene rings¹⁹) and its conformation (e.g. dihedral angles²¹) are key parameters. Taking advantage of the effect of temperature on structural conformation (e.g. bond vibrations, aryl ring rotations), molecules have been synthesised in which thermal changes can induce a transition between the two kinds of electronic states^{23,24}. Compared to the response speeds of typical electronic components, heat flow is very slow and so the thermal operation of nanoelectronic devices is difficult to envision. Moreover, such developments are currently focused on discrete molecules, thus hindering the scale-up of this bi-stable switching capacity to the device scale.

Bypassing the difficulties in manipulating individual molecules, our proposal starts with the integration of bi-radicals into 2D covalent organic radical frameworks (2D-CORFs), as schematically shown in Fig. 1. This covalent integration permits the direct manipulation of the structural conformations of bi-radicals through the application of external forces to the material while preserving their well-defined and stable positions within the 2D framework²⁵. Moreover, 2D-CORFs can exhibit multi-radical (e.g. FM^{26–28} or AFM²⁹) and diamagnetic quinoidal closed-shell

solutions³⁰, confirming that the electronic bi-stability from the bi-radical monomers persists in the resulting 2D materials. Recently, different examples of experimentally synthesised 2D-CORFs have been reported, exhibiting multi-radical electronic ground states with either AFM coupling^{31,32} or weak coupling (i.e. paramagnetism) between spins³³.

Dihedral angles of aryl rings are effective conformational parameters to control the localisation/delocalisation of unpaired electrons in π -conjugated organic radicals³⁴ (e.g. triarylmethyls, TAMs) and so determine the balance between the AFM and quinoidal states in bi-radicals^{18,21}. As we have recently shown, compression is an effective manner to homogeneously flatten all aryl rings in 2D-CORFs to significantly increase electron delocalisation³⁵. However, to induce local electron pairing one needs to asymmetrically flatten only some aryl rings with respect to others. Previously, we predicted that such a type of conformational manipulation in paramagnetic 2D-CORFs is possible via in-plane uniaxial strain (arrows in Fig. 1b)^{25,36}. Numerous ways to experimentally induce significant strain in a 2D material placed on a substrate have been proposed such as (i) growing the 2D material on epitaxial substrates with a controlled lattice constant mismatch, (ii) thermal-expansion mismatch between the 2D material and its substrate and (iii) transferring the 2D material onto a flexible substrate and directly stretching, compressing or bending the substrate^{3,37}. Herein, we show how modest in-plane stretching of 2D-CORFs could be a very promising route to externally and sensitively control the balance between both electronic configurations, and thus modulate their resulting optical, magnetic and electrical response.

In this work, we study the effect of uniaxial in-plane strain on a series of 2D-CORFs, including the recently synthesised PTM_{acetylenic} material³², by means of density functional theory (DFT) based calculations. Our results demonstrate that uniaxial in-plane tensile strains may be used to efficiently switch between the AFM multi-radical (i.e. spin polarised or open-shell) and diamagnetic quinoidal (i.e. closed-shell) states in a reversible manner. Such control arises from the effective manipulation of aryl ring dihedral angles, although other parameters such as the structural and chemical design of each 2D-CORF also play important roles. In addition, we demonstrate that the strain-induced electronic switching is robust at finite temperatures, which is a prerequisite for its experimental viability, and thus technological applicability. Overall, our work shows that integrating bi-radical moieties into covalent 2D materials is a promising route to achieve macroscopic control over molecular-scale electronic states.

Results

2D-CORF design and characterisation. Following our previous work^{25,30,34,35} and other studies²⁹ on these types of systems, we employ the PBE0 hybrid density functional³⁸ (which incorporates

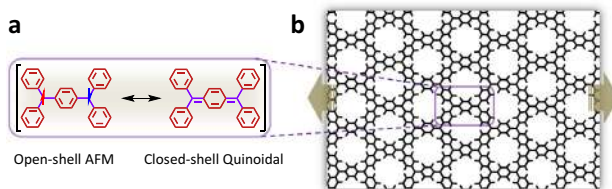


Fig. 1 Bi-radical integration in 2D-CORFs. **a** Schematic representation of the energetically close resonant valence bond forms in the ground state of Thiele's bi-radical: the open-shell AFM (left) and the closed-shell quinoidal (right). **b** 2D-CORFs viewed as extended 2D analogues of bi-radical compounds, where one could potentially switch between the two electronic states via external uniaxial strain (brown arrows).

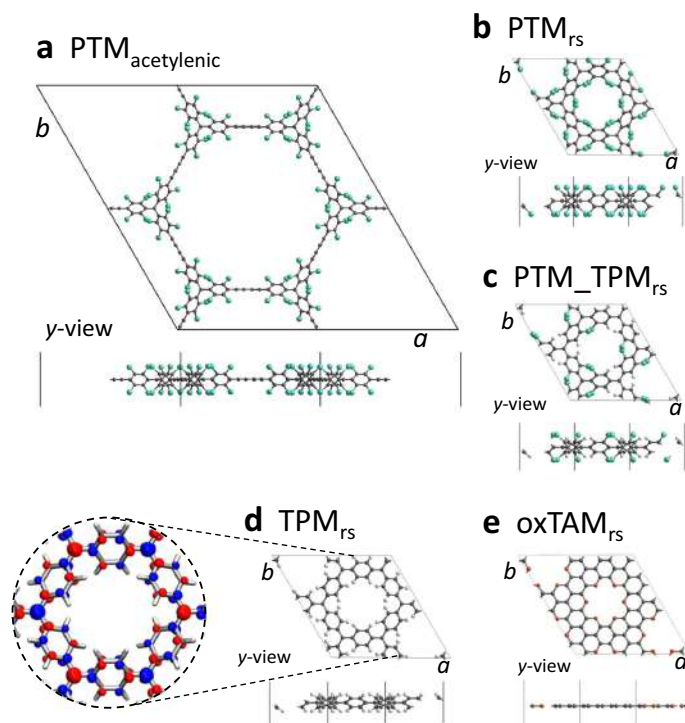


Fig. 2 2D-CORF characterisation. DFT-optimised crystal structures of **a** $\text{PTM}_{\text{acetylenic}}$, **b** PTM_{rs} , **c** $\text{PTM_TPM}_{\text{rs}}$, **d** TPM_{rs} and **e** oxTAM_{rs} 2D-CORFs. Colouring: C—dark grey, H—white, Cl—green, O—red. The spatially resolved spin density for the AFM solution (spin-up: blue, spin-down: red) is shown for TPM_{rs} (**d**).

25% of non-local Hartree Fock exchange—HFE) in order to best capture the electronic structure of 2D-CORF materials. DFT-based calculations using hybrid functionals have proven to have remarkable reliability in describing the ground states and magnetic properties of organic radicals³⁹ and polyradicals^{40–42}. The structures of all considered 2D-CORFs, which have been optimised using DFT/PBE0 calculations (see “Methods” for full details), are shown in Fig. 2. Our considered 2D-CORFs are hexagonal networks of sp^2 radical carbon centres (αC) covalently linked through π -conjugated groups³⁰. The different linkers define the range of different 2D-CORFs that we consider: (i) the experimentally synthesised³² $\text{PTM}_{\text{acetylenic}}$ (Fig. 2a, PTM stands for perchlorotriarylmethyl^{43,44}), (ii) the chlorinated ring-sharing PTM_{rs} (Fig. 2b), (iii) the ring-sharing mixed $\text{PTM_TPM}_{\text{rs}}$ (Fig. 2c, TPM stands for triphenylmethyl⁴⁵), (iv) the ring-sharing TPM_{rs} (Fig. 2d) and (v) the ring-sharing oxTAM_{rs} (Fig. 2e, oxTAM stands for oxo-triarylmethyl⁴⁶). $\text{PTM}_{\text{acetylenic}}$ (**2a**) has been experimentally characterised as an AFM semiconductor³². In PTM_{rs} (**2b**), $\text{PTM_TPM}_{\text{rs}}$ (**2c**), TPM_{rs} (**2d**) and oxTAM_{rs} (**2e**), the αC centres are linked through a single aryl ring, thus significantly enhancing their electronic coupling as compared to $\text{PTM}_{\text{acetylenic}}$ (**2a**). Although such ring-sharing 2D-CORFs have not yet been experimentally reported, an analogue six-membered ring-sharing αC oligomer has been recently reported⁴⁷ and shown to possess a similar electronic structure to bi-radical compounds¹⁸. In the SI

(Section 3) we show that the experimentally derived nearest neighbour magnetic coupling between αC centres in this TPM_{rs} -like oligomer is very similar to that we predict for the TPM_{rs} 2D-CORF, thus confirming the adequacy of our computational approach.

The different chemical functionalisation of the aryl rings in our considered 2D-CORFs determines their dihedral angles through steric hindrance³⁴. Thus, through the series shown in Fig. 2 from **2a** to **2e**, we observe increasingly planarized aryl rings (see the mean dihedral angle in Table 1): from the highly twisted conformation in $\text{PTM}_{\text{acetylenic}}$ to the fully planar structure of oxTAM_{rs} , where oxygen atoms bonded between adjacent aryl rings fix them to be in-plane (see y-views in Fig. 2).

All 2D-CORFs exhibit a multi-radical open-shell ground state solution which is lower in energy than the closed-shell quinoidal minimum, except for TPM_{rs} , where both states are nearly degenerate (see $E_{\text{AFM}} - E_{\text{QUI}}$ in Table 1). The open-shell solution exhibits an alternating spin-up/spin-down ordering between αC centres, as shown with the spatially resolved spin density for TPM_{rs} in Fig. 2d. Formally, this is a partial representation based on a broken symmetry (BS) solution resulting from the limitation of the single determinantal character of DFT-based calculations. Although the spin-polarisation ordering should not be taken directly at face value, such solutions are formally linked to open-shell states of the real system such as the FM or the AFM

Table 1 Mean dihedral angles (degrees), the energy difference between AFM and closed-shell quinoial (QUI) electronic solutions per α C centre (meV), the average absolute value of spin populations on α C centres in the AFM solution ($\langle |\mu_{\alpha C}| \rangle$) and the associated electronic bandgaps for all considered 2D-CORFs.

	PTM _{acetylenic}	PTM _{rs}	PTM TPM _{rs}	TPM _{rs}	oxTAM _{rs}
Mean dihedral angle (degrees)	48.0	46.8	40.1	32.4	0.0
$E_{AFM} - E_{QUI}$ (meV)	-337.3	-136.9	-19.6	3.3	-15.6
$\langle \mu_{\alpha C} \rangle$	0.45	0.39	0.28	0.22	0.12
Bandgap (eV)	2.17	2.36	1.61	1.64	1.12

Note that, for oxTAM_{rs}, the $E_{AFM} - E_{QUI}$ value is taken as the difference between E_{AFM} and the energy of the semimetallic solution (see SI Fig. S2).

states^{48–51}. The degree of BS spin polarisation reflects the extent of the mixture between the open-shell and closed-shell valence bond forms. The appearance of such solutions also implies the emergence of low-lying states with non-zero net magnetisation and thus is a predictor of the existence of paramagnetism/antiferromagnetism for a particular system/condition. In order to quantify the magnitude of the AFM coupling in a periodic calculation one can assume that the magnetic system is described by a defined topology of localised spins and use the energy difference between the BS ground state solution and that of the FM state solution to map these states on to a Heisenberg model as described in the SI (see Table S2 and Section 3 in the SI)⁵². For simplicity, in the remainder, we will refer to BS open-shell solutions as AFM solutions, where we take the degree of spin polarisation in such solutions as an indicator of the strength of the open-shell character of the real state.

The increasing aryl ring twist angles and distance between radical centres, going from left to right through the series of 2D-CORFs in Table 1, also correlates with increasingly localised spin distributions (see values of the average of the absolute α C spin population, $\langle |\mu_{\alpha C}| \rangle$). We also see that increased electronic localisation (e.g. due to more twisted aryl rings) results in a higher energetic cost associated with electron pairing (see $E_{AFM} - E_{QUI}$ in Table 1). For oxTAM_{rs} we could not stabilise the localised quinoial state which spontaneously falls into a fully delocalised semimetallic solution (see SI Fig. S2), in accordance with other fully planar 2D-CORFs (e.g. graphynes)³⁰. The degree of electronic localisation in the AFM solution for each 2D-CORF is also reflected in the tendency for a decrease in the electronic bandgap (see Table 1) and an increase in band dispersion (see Fig. S1 in the SI) through the series of 2D-CORFs. These tendencies are also associated with the strength of the AFM coupling, which increases monotonically through the series of 2D-CORFs (see Table S2 in the SI).

As previously demonstrated for TPM_{rs}³⁰, the quinoial closed-shell solution leads to a local pairing of electrons in specific aryl rings within the framework, which subsequently become more planar³⁰. This structural response is due to the double bonds formed between the involved α C centres and the aryl rings linking them. Herein, we explore the opposite phenomenon: i.e. whether by external manipulation of certain aryl rings we may induce electron pairing. To test this idea we consider the in-plane uniaxial strain of our 2D-CORFs^{25,36}. Our starting electronic configuration is the unstrained open-shell AFM distribution shown in Fig. 2d with an associated alternating spin density. We highlight the response of TPM_{rs} as a reference example of the prototypical structural response of our considered 2D-CORFs to uniaxial strain (see Fig. 3a). For the relaxed TPM_{rs} conformation ($\epsilon = 0\%$), all aryl rings are equally twisted ($\varphi_1 = \varphi_2 = 33^\circ$). Upon stretching ($\epsilon = 28\%$), the aryl rings parallel to the strain direction are flattened ($\varphi_2 = 4^\circ$) and the remaining rings become twisted

out-of-plane ($\varphi_1 = 75^\circ$). These conformational changes lead to a 12.5% reduction of the *b* unit cell parameter (Fig. 3a).

These strain-induced conformational changes are likely to favour electron pairing within the flattened aryl rings (φ_2 in Fig. 3a). In principle, the higher the dihedral angle difference between the two types of aryl rings ($|\varphi_1 - \varphi_2|$) the higher the probability is to induce electron pairing. TPM_{rs} displays the most significant conformational changes, followed by PTM_{TPM_{rs}} and PTM_{rs} (Fig. 3f), which is in accordance with the degree of chlorine functionalisation in the latter two materials, making their aryl rings more rigidly fixed. For PTM_{acetylenic}, the acetylenic linkers provide an extra degree of freedom which is not present in the other 2D-CORFs, leading to smaller changes of $|\varphi_1 - \varphi_2|$ with respect to strain. For the fully planar structure of oxTAM_{rs}, $|\varphi_1 - \varphi_2|$ remains invariable, and equal to zero, throughout the range of considered strain (Fig. 3f).

The large strength of each material, characterised by Young's modulus (YM), depends on the different structural degrees of freedom in each case. Figure 3g shows our calculated YM values for all 2D-CORFs in comparison with experimentally reported values for graphene⁵³ and single layer MoS₂⁵⁴. Here, we can see that, except for the highly rigid oxTAM_{rs}, all 2D-CORFs have a YM that is both significantly lower than that of a typical inorganic 2D material and ~20 times smaller than that of graphene (~1000 GPa⁵³) helping to quantify the relative ease by which our 2D-CORFs can be strained. For the case of single-layer MoS₂, different on-substrate based methods have been employed to uniaxially stretch the material by ~2.5%⁵. Considering the magnitude of the in-plane YM of single-layer MoS₂ (see Fig. 3g), this implies that such experiments can readily apply in-plane uniaxial tensile stresses of at least 6.5 GPa. Taking TPM_{rs} as an example, such tensile stress would induce a uniaxial strain of ~20%. In the case of more specialised experimental set-ups, graphene has been uniaxially stretched by almost 6%, showing that larger in-plane tensile stresses of up to ~60 GPa are also achievable⁴. We also note that the YM values of TPM_{rs}, PTM_{TPM_{rs}}, PTM_{rs} and PTM_{acetylenic} are all significantly smaller than that of oxTAM_{rs} which strongly indicates that aryl ring twisting is the key factor leading to highly stretchable 2D-CORFs, rather than simply their nanoporous structure.

In order to test whether the strain-induced conformational changes depicted in Fig. 3 lead to a transition from the open-shell AFM electronic state towards the closed-shell quinoial diamagnetic state, we have extracted $\langle |\mu_{\alpha C}| \rangle$ values throughout the full range of considered strains for the corresponding electronic solutions. Figure 4a shows the variation of $\langle |\mu_{\alpha C}| \rangle$ against uniaxial strain for each studied 2D-CORF. PTM_{acetylenic} shows a high and robust $\langle |\mu_{\alpha C}| \rangle$ value throughout stretching, which is a consequence of the strongly localised unpaired electrons in this material, which in turn, is due to the large distance between α C centres and the highly twisted aryl rings (48°). Conversely, oxTAM_{rs} shows the

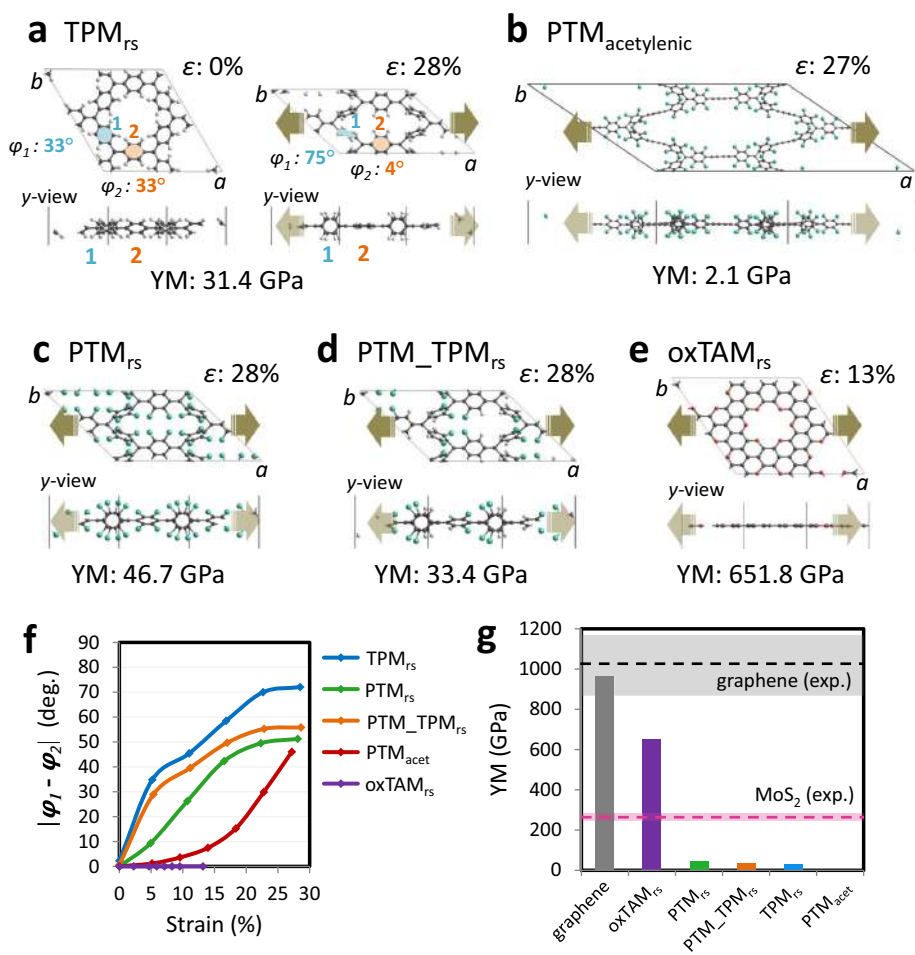


Fig. 3 Structural response to strain. **a–e** Crystal structures of the most stretched conformations for each of the studied 2D-CORFs indicating the associated strain (ϵ) and in-plane Young's modulus (YM). Atom colour key: C—dark grey, H—light grey, Cl—green, O—red. **f** Dihedral angle difference between the out-of-plane (φ_1) and in-plane (φ_2) twisted aryl rings versus uniaxial strain. **g** Calculated in-plane YM values for all 2D-CORFs and graphene (vertical bars) compared to experimental YM values of graphene⁵³ and single layer MoS_2 ⁵⁴ (horizontal dashed lines), where the associated range within which experimental values have been reported are indicated with coloured panels (see Table S1 in the SI for details).

lowest $\langle \mu_{\text{ac}} \rangle$ value in the relaxed conformation. As previously mentioned, the aryl rings in oxTAM_{rs} are fixed in-plane, and so uniaxial strain mainly induces stretching of π - π bonds along the strain direction. As a consequence, unpaired electrons in oxTAM_{rs} become more localised, and thus $\langle \mu_{\text{ac}} \rangle$ monotonically increases with increasing strain (Fig. 4a). In contrast, for TPM_{rs} , $\text{PTM_TPM}_{\text{rs}}$ and PTM_{rs} , uniaxially straining the relaxed structure induces a clear transition from the open-shell AFM solution to the closed-shell diamagnetic quinooidal solution, in which $\langle \mu_{\text{ac}} \rangle$ vanishes. TPM_{rs} , the 2D-CORF with the highest aryl ring twisting

flexibility, is most prone to electron pairing and exhibits a full depletion of $\langle \mu_{\text{ac}} \rangle$ for strains between 5% to 22% (blue curve in Fig. 4a). For uniaxial strains above 25%, $\langle \mu_{\text{ac}} \rangle$ rises again, which is a consequence of the associated α -C-aryl ring bond distances reaching typical values of single carbon-carbon bonds (1.54 Å; see SI Fig. S3). In Fig. 4b–d, we show the three most representative situations, where one may clearly see the transition from the AFM open-shell solutions (see spin-density iso-surfaces in Fig. 4b, d) to the intermediate electron-paired quinooidal distribution (see highest occupied crystal orbital density in Fig. 4c). $\text{PTM_TPM}_{\text{rs}}$ and PTM_{rs}

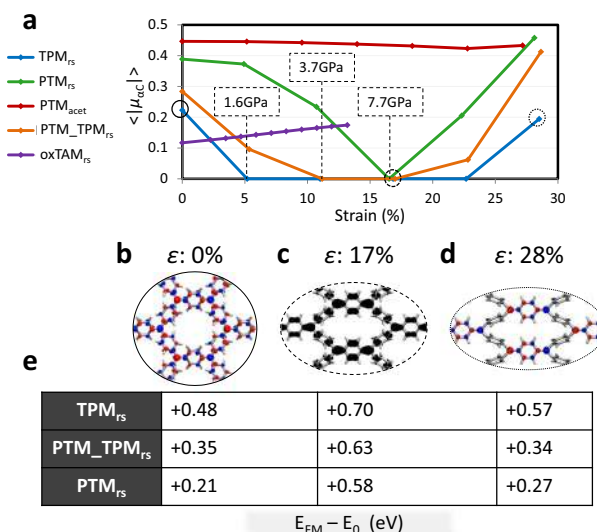


Fig. 4 Electronic response to strain. **a** Average of the absolute spin population per αC for each studied 2D-CORF versus uniaxial strain. Estimated in-plane tensile stresses required to induce the AFM-to-quinoidal transition are indicated for TPM_{rs}, PTM_{TPM_{rs}} and PTM_{rs}. **b** Spin density iso-surface (spin up: blue; spin down: red) for TPM_{rs} when relaxed ($\epsilon = 0\%$). **c** Highest occupied crystal orbital density (black) for the semi-strained conformation of TPM_{rs} ($\epsilon = 17\%$). **d** Spin density iso-surface for highly strained TPM_{rs} ($\epsilon = 28.5\%$). **e** Relative energy per αC (in eV) of the ferromagnetic electronic solution (E_{FM}) with respect to the electronic ground state (E_0) at each conformation (**b–d**) for TPM_{rs}, PTM_{TPM_{rs}} and PTM_{rs}. For the cases when the AFM solution is the ground state these energy differences are related to the degree of magnetic coupling (see also Table S2 in the SI).

also show a complete depletion of $\langle |\mu_{\alpha C}| \rangle$, which, in each case, is induced at a higher strain than for TPM_{rs} and is maintained for a comparatively smaller range of strains (Fig. 4a). This behaviour is due to the more localised unpaired electrons in these two materials (see Table 1), and the lower aryl ring rotational flexibility arising from chlorine functionalisation (see Fig. 3f and SI Fig. S4). We note that the overall picture we obtain from the variation of $\langle |\mu_{\alpha C}| \rangle$ as a function of uniaxial strain (Fig. 4) is in agreement with that from bond length alternation (BLA) analysis (see SI Fig. S5), which is a structural indicator often used to characterise the balance between the AFM and quinoidal states in bi-radical compounds^{22,55}.

Although calculations on graphene have suggested that $>15\%$ in-plane biaxial strain could also induce AFM or quinoidal dimerised states⁵, other theoretical studies point to the spontaneous rupture of the material at such strains⁶. Thus far, due to its ultrahigh in-plane strength, only tensile strains up to $\sim 6\%$ have been experimentally reached in graphene before failure⁴. In contrast, the relatively low tensile strengths of our 2D-CORFs permit the AFM to quinoidal interconversion via experimentally achievable relatively large elastic strains.

Finally, in Fig. 4e we provide the energetic cost associated with FM polarisation of the αC unpaired electrons (e.g. via external magnetic fields) with respect to the ground state at each structural conformation for the TPM_{rs}, PTM_{TPM_{rs}} and PTM_{rs}. This cost can be associated with the accessibility of open-shell states for a given system in a particular conformation. Here, we can see that the FM solution becomes significantly destabilised upon electron pairing, as induced in the three 2D-CORFs with uniaxial strains of ca. 17% (Fig. 4c). This extra energetic cost for closed-shell quinoidal diamagnetic ground states may be associated with the

process of breaking π - π double bonds in order to generate a spin-polarised FM electronic distribution. This effect is particularly significant for PTM_{rs} where the energetic cost to reach the FM solution is nearly three times larger for the “paired” configuration at $\epsilon = 17\%$ as compared to the “unpaired” configuration at $\epsilon = 0\%$. These results highlight the potential of ring-sharing 2D-CORFs as platforms with externally controllable spintronic characteristics.

Overall, the results of Fig. 4 demonstrate that at 0 K it is possible to induce electron pairing of π -conjugated electrons in 2D-CORFs by means of external uniaxial strain. Because of the conformational character of the mechanism (i.e. aryl ring twisting) and the effect of bond vibrations on electron delocalisation³⁴, it is expected that finite temperatures could affect the degree of external control over electron pairing. To test the robustness of our results at finite temperatures, we ran ab initio molecular dynamics simulations (AIMDS; see “Methods” for details) at 300 K for our three most promising 2D-CORFs, namely TPM_{rs}, PTM_{TPM_{rs}} and PTM_{rs}. As shown in SI Fig. S6, the strain-induced manipulation of aryl ring twisting previously characterised at 0 K (Fig. 3) holds at 300 K, despite fluctuations in dihedral angles induced by thermal vibrations. We note that there are some conformational differences between networks (see more details in Section 2 of the SI) but the overall structural response to strain is qualitatively the same for the three tested 2D-CORFs.

To assess whether the aryl ring twist manipulation also leads to a robust control over electron pairing, we have extracted the $\langle |\mu_{\alpha C}| \rangle$ values during the AIMDS at 300 K for the three 2D-CORFs at different strains. As one can see in Fig. 5, all 2D-CORFs show non-zero $\langle |\mu_{\alpha C}| \rangle$ values in the relaxed conformation ($\epsilon = 0\%$; Fig. 5a, d, g). This is particularly interesting for TPM_{rs} where the open-shell

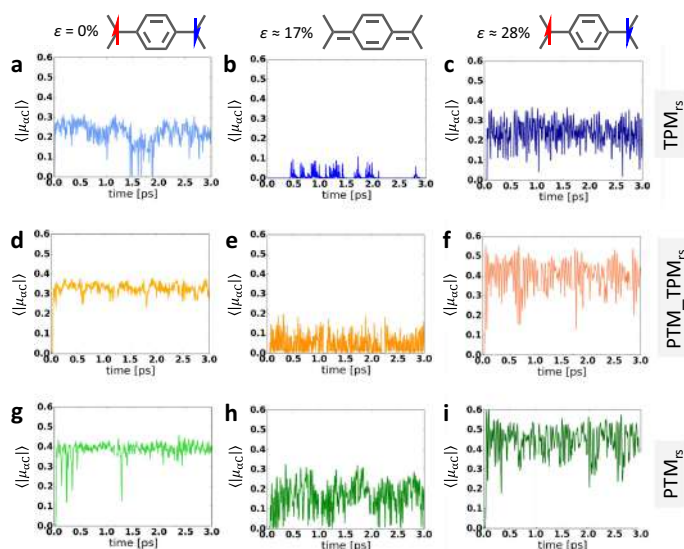


Fig. 5 Electronic response to strain at 300 K. Time-resolved evolution of the αC averaged absolute spin population $\langle |\mu_{ac}| \rangle$ during 3 ps of AIMDS at 300 K for 0%, 17% and 28% uniaxial strains for (a–c) TPM_{rs}, (d–f) PTM_TPM_{rs} and (g–i) PTM_{rs}. The chemical sketches at the top of the figure represent the resulting dominant valence bond form for each strain.

AFM solution is slightly higher in energy at 0 K with respect to the closed-shell quinoidal one (see Table 1). It thus follows that finite temperatures, through bond vibrations inducing electron localisation³⁴, promote the open-shell AFM character (Fig. 5a). This is in good agreement with experimental variable-temperature measurements on the analogous six-membered TPM_{rs} oligomer⁴⁷. $\langle |\mu_{ac}| \rangle$ is larger and more stable for PTM_TPM_{rs} and PTM_{rs} in the relaxed conformations (Fig. 5d, g) due to the more perpendicular chlorinated aryl rings in these networks³⁴. For TPM_{rs}, partial uniaxial strain ($\epsilon = 16\%$) leads to full electron pairing, as demonstrated with the nearly complete depletion of $\langle |\mu_{ac}| \rangle$, showing only a small degree of noise due to thermal fluctuations (Fig. 5b). Such effective pairing may be understood by the almost complete flattening of the aryl rings in TPM_{rs} parallel to the strain direction (see SI Fig. S6b). PTM_TPM_{rs} and PTM_{rs} (Fig. 5e, h) also display vanishing $\langle |\mu_{ac}| \rangle$ values at $\epsilon = 16\%$ (especially notable for PTM_TPM_{rs}, Fig. 5e). However, in these cases, there is an increase of $\langle |\mu_{ac}| \rangle$ noise due to thermal fluctuations, which is significantly detrimental in PTM_{rs}. Following the behaviour at 0 K, further stretching the networks again increases the $\langle |\mu_{ac}| \rangle$ values for all materials (Fig. 5c, f, i), caused by the elongation of carbon–carbon bonds.

Overall, these results confirm the robustness of our proposed mechanical control of electron pairing at finite temperatures (300 K) via strain-induced manipulation of aryl ring twisting in 2D-CORFs. The chemical functionalisation of aryl rings is found to play a key role in determining aryl ring rotational flexibility under uniaxial strain (see Fig. S6 in the SI), in line with the response of insulating 2D-CORFs³⁶, and thus the effectiveness of electron pairing (Fig. 5). Thereby, PTM_TPM_{rs} appears to exhibit the most robust (and so potentially measurable) transition

between the open-shell AFM (Fig. 5d) and closed-shell quinoidal (Fig. 5e) electronic solutions. The behaviour of PTM_TPM_{rs} arises from a balance between electron localisation in the relaxed structure combined with significant conformational flexibility due to the mixed chlorine/hydrogen functionalisation of its aryl rings and is a promising candidate to be experimentally pursued.

Discussion

In this work, we propose a strategy to externally control the transition between open-shell AFM and closed-shell quinoidal electronic states, which have long been studied in the field of molecular electronics^{14–16}. Such states coexist in Kekulé organic bi-radicals¹⁸ and have recently been reported in atomically precise graphene nanoribbons⁵⁶. Although organic chemists have proven that chemical design is a powerful tool to induce one electronic state or another¹⁶, a feasible procedure to achieve dynamic external control over electron pairing was lacking.

Here, we propose 2D-CORFs as ideal platforms to gain such control. We demonstrate that the application of uniaxial strain in 2D-CORFs allows one to effectively pair π -conjugated electrons within such otherwise open-shell multi-radical materials. The key factor for such strain-control comes through the mechanical manipulation of dihedral angles of aryl rings in these materials. Partial strain leads to a flattening of some aryl rings within the networks which, subsequently, leads to an effective electron pairing (quinoidalisation) within them. Further strain stretches the previously generated double bonds, which unpairs the electrons restoring the AFM spin distribution. Such an externally mediated electronic control mechanism is effective not only at 0 K but also at finite temperatures, as shown via AIMDS at 300 K. Our results also highlight the important role of structural and

ARTICLE

NATURE COMMUNICATIONS | <https://doi.org/10.1038/s41467-021-21885-y>

chemical design of 2D-CORFs for enhancing strain-induced electronic control. In this respect, we demonstrate that radical centres need to be close to each other to achieve electron pairing and that the chemical functionalisation of aryl rings, determining the ease with which their dihedral angles may be manipulated, is an important factor.

Ongoing improvements in the bottom-up synthesis of single/few-layer COFs with increasingly large crystalline domains will also assist the experimental realisation of our proposal⁵⁷. The relatively low tensile strengths of our considered 2D-CORF materials also permit large elastic strains with experimentally accessible applied in-plane stresses, which is essential to ensure electronic control in real-world applications. Such high strains are not accessible in stiffer 2D materials, such as graphene, where similar strain-induced electronic transitions have been theoretically predicted⁵ but which are experimentally intractable^{4,6}.

Overall, our proposal merges the fundamental switchability of bi-radicals^{19–21} with the structural robustness and elasticity of organic 2D materials¹⁰. Unlike the effect of strain on most inorganic 2D materials, where minimal bond stretching induces small electronic changes^{4–6}, the remarkable efficiency of aryl ring twisting to induce electron pairing highlights the technological potential of rationally designed 2D organic materials.

Methods

All 2D-CORFs studied in this work were fully optimised using periodic boundary conditions via an efficient cascade methodology. Classical force-field based geometry optimisations using the General Utility Lattice Programme⁵⁸ and the universal force field⁵⁹ were employed to generate pre-optimised structures (both atomic positions and cell-parameters). Then a second full geometry optimisation of both atomic structure and cell parameters was performed using DFT-based calculations and the PBE⁶⁰ functional with a Tier 1 light numerical atom-centred-orbital (NAO) basis set⁶¹. This was followed by a final full optimisation using the PBE0⁶² hybrid functional (incorporating 25% HFE) with the same NAO basis set. In the SI we confirm that DFT calculations using the PBE0 functional are able to accurately to capture local magnetic coupling between α C centres in 2D-CORFs through comparison with results from experiments of a TPM_n oligomer. All DFT calculations employed a 6-6-1 gamma-centred Monkhorst Pack k -point sampling, except for the $\text{PTM}_{\text{acetylenic}}$ 2D-CORF which, due to its relatively large unit cell, was optimised using gamma-point sampling. Convergence criteria for these calculations were 1×10^{-5} eV for total energy and 1×10^{-2} eV/Å for the maximum residual force component per atom. We note that a Tier 1 light NAO basis set provides results of similar or higher quality to those obtained using a triple-zeta plus polarisation Gaussian type orbital basis set⁶². Details of how these NAO basis sets were developed are provided in ref. ⁶³. Single point PBE0 calculations were performed on these fully optimised structures, to generate band structures and isosurface maps of spin and crystal orbital electronic densities. Atomically partitioned spin densities on α C centres ($\mu_{\alpha\text{C}}$) were obtained via the Hirshfeld method⁶⁴. The PBE0 geometry optimisations and single point calculations were performed separately for the open-shell AFM and closed-shell quinoidal solutions for each 2D-CORF. The open-shell AFM solutions were obtained with spin unrestricted DFT calculations and by setting an alternating spin polarisation initial guess over neighbouring α C sites. The closed-shell quinoidal solutions were obtained using closed-shell restricted DFT calculations. We provide an example of an input file (including the utilised basis set specification) used for our calculations in the SI, together with the cell parameters and atomic coordinates of fully optimised geometries for all studied 2D-CORFs in their relaxed conformation ($\epsilon = 0$) for the open-shell AFM electronic solution.

To mimic the externally induced uniaxial strain on each 2D-CORF, we performed a series of restricted optimisations systematically increasing one of the cell parameters while allowing the other in-plane cell parameters and atomic positions to relax. Specifically, the $a(x)$ cell parameter was modified in steps of 0.2 Å for αTAM_{10} , 1 Å for PTM_{10} , $\text{PTM}_{\text{TPM}_{10}}$ and TPM_{10} , and 2 Å for $\text{PTM}_{\text{acetylenic}}$.

The response of 2D-CORFs at finite temperatures was modelled using AIMDS. These were run for 3 ps at 300 K with fixed lattice parameters, using the Bussi–Donadio–Parrinello⁶⁵ thermostat, the hybrid PBE0 functional and a light (Tier-1) NAO basis set (see above). Further information on the post-analysis of AIMDS results is provided in Section 2 of the SI. All DFT-based calculations were performed using the FHI-AIMS code^{63,66}.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request. Optimised 2D-CORF geometries and FHI-AIMS input files used to generate much of the data used in this study can be found in the SI.

Received: 24 November 2020; Accepted: 17 February 2021;

Published online: 17 March 2021

References

- Novoselov, K. S. et al. Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
- Nicolosi, V., Chhowalla, M., Kanatzidis, M. G., Strano, M. S. & Coleman, J. N. Liquid exfoliation of layered. *Mater. Sci.* **340**, 1226419–1226419 (2013).
- Roldán, R., Castellanos-Gomez, A., Cappelluti, E. & Guinea, F. Strain engineering in semiconducting two-dimensional crystals. *J. Phys. Condens. Matter* **27**, 313201 (2015).
- Cao, K. et al. Elastic straining of free-standing monolayer graphene. *Nat. Commun.* **11**, 1–7 (2020).
- Sorella, S. et al. Correlation-driven dimerization and topological gap opening in isotropically strained graphene. *Phys. Rev. Lett.* **121**, 066402 (2018).
- Lee, S. H. et al. Band gap opening by two-dimensional manifestation of Peierls instability in graphene. *ACS Nano* **5**, 2964–2969 (2011).
- Grill, L. et al. Nano-architectures by covalent assembly of molecular building blocks. *Nat. Nanotechnol.* **2**, 687–691 (2007).
- Liu, X. H., Guan, C. Z., Wang, D. & Wan, L. J. Graphene-like single-layered covalent organic frameworks: synthesis strategies and application prospects. *Adv. Mater.* **26**, 6912–6920 (2014).
- Geng, K. et al. Covalent organic frameworks: design, synthesis, and functions. *Chem. Rev.* **120**, 8814–8933 (2020).
- Zhuang, X., Mai, Y., Wu, D., Zhang, F. & Feng, X. Two-dimensional soft nanomaterials: a fascinating world of materials. *Adv. Mater.* **27**, 403–427 (2015).
- DeBlase, C. R. & Dichtel, W. R. Moving beyond boron: the emergence of new linkage chemistries in covalent organic frameworks. *Macromolecules* **49**, 5297–5305 (2016).
- Allendorf, M. D. et al. Electronic devices using open framework materials. *Chem. Rev.* **120**, 8581–8640 (2020).
- Moreno, C. et al. Bottom-up synthesis of multifunctional nanoporous graphene. *Science* **360**, 199–203 (2018).
- Hu, X., Wang, W., Wang, D. & Zheng, Y. The electronic applications of stable diradicaloids: present and future. *J. Mater. Chem. C* **6**, 11232–11242 (2018).
- Tobe, Y. Quinoidmethanes incorporated in non-benzenoid aromatic or antiaromatic frameworks. *Top. Curr. Chem.* **376**, 12 (2018).
- Zeng, Z. et al. Pro-aromatic and anti-aromatic π -conjugated molecules: an irresistible wish to be diradicals. *Chem. Soc. Rev.* **44**, 6578–6596 (2015).
- Dressler, J. J. et al. Thiophene and its sulfur inhibit indenodibenzo[thiophene] diradicals from low-energy lying thermal triplets. *Nat. Chem.* **10**, 1134–1140 (2018).
- Shishlov, N. M. From the Gomberg radical to organic magnets. *Russ. Chem. Rev.* **75**, 863–884 (2006).
- Trinquier, G. & Malrieu, J.-P. Kekulé versus Lewis: when aromaticity prevents electron pairing and imposes polyradical character. *Chemistry* **21**, 814–828 (2015).
- Zeng, Z. et al. Stable Tetrabenzo-Chichibabin's hydrocarbons: tunable ground state and unusual transition between their closed-shell and open-shell resonance forms. *J. Am. Chem. Soc.* **134**, 14513–14525 (2012).
- Ravat, P. & Baumgarten, M. 'Tschtitschibabin type biradicals': benzenoid or quinoid? *Phys. Chem. Chem. Phys.* **17**, 983–991 (2015).
- Mayorga Burrezo, P., Zafra, J. L., López Navarrete, J. T. & Casado, J. Quinoidal/aromatic transformation in π -conjugated oligomers: vibrational Raman studies on the limit of rupture for π -bonds. *Angew. Chem.* **56**, 2250–2259 (2017).
- Su, Y. et al. Thermally controlling the singlet-triplet energy gap of a diradical in the solid state. *Chem. Sci.* **7**, 6514–6518 (2016).
- Wentrup, C., Regimbald-Krnel, M. J., Müller, D. & Comba, P. A thermally populated, perpendicularly twisted alkene triplet diradical. *Angew. Chem. Int. Ed.* **55**, 14600–14605 (2016).
- Alcón, I., Reta, D., Moreira, I., de, P. R. & Bromley, S. T. Design of multifunctional 2D open-shell organic networks with mechanically controllable properties. *Chem. Sci.* **8**, 1027–1039 (2017).
- Ovchinnikov, A. A. Multiplicity of the ground state of large alternant organic molecules with conjugated bonds—(do organic ferromagnetics exist?). *Theor. Chim. Acta* **47**, 297–304 (1978).
- Mataga, N. Possible 'ferromagnetic states' of some hypothetical hydrocarbons. *Theor. Chim. Acta* **10**, 372–376 (1968).
- Rajca, A., Lu, K. & Rajca, S. High-spin poly(arylmethyl) polyradical: fragment of a macrocyclic 2-strand based upon calix[4]arene rings. *J. Am. Chem. Soc.* **119**, 10335–10345 (1997).
- Thomas, S., Li, H. & Bredas, J.-L. Emergence of an antiferromagnetic mott insulating phase in hexagonal π -conjugated covalent organic frameworks. *Adv. Mater.* **31**, 1900355 (2019).

30. Alcón, I., Viñes, F., Moreira, I., de, P. R. & Bromley, S. T. Existence of multi-radical and closed-shell semiconducting states in post-graphene organic Dirac materials. *Nat. Commun.* **8**, 1957 (2017).
31. Yang, Y. et al. Antiferromagnetism in two-dimensional polyradical nanosheets. *Polym. Chem.* **9**, 5499–5503 (2018).
32. Wu, J. et al. Toward π -conjugated 2D covalent organic radical frameworks. *Angew. Chem. Int. Ed.* **57**, 8007–8011 (2018).
33. Jiang, Y. et al. Organic radical-linked covalent triazine framework with paramagnetic behavior. *ACS Nano* **13**, 5251–5258 (2019).
34. Alcón, I. & Bromley, S. T. Structural control over spin localization in triarylmethyls. *RSC Adv.* **5**, 98593–98599 (2015).
35. Santiago, R., Alcón, I., Ribas-Arino, J., Deumal, M., de P. R. Moreira, I. & Bromley, S. T. 2D hexagonal covalent organic radical frameworks as tunable correlated electron systems. *Adv. Funct. Mater.* **31**, 2004584 (2021).
36. Alcón, I. & Bromley, S. T. Triarylmethyl-based 2D covalent networks: virtual screening of chemical functionalisation for optimising strain-induced property control. *Phys. Chem. Chem. Phys.* **20**, 5028–5035 (2018).
37. Dai, Z., Liu, L. & Zhang, Z. Strain engineering of 2D materials: issues and opportunities at the interface. *Adv. Mater.* **31**, 1805417 (2019).
38. Adamo, C. & Barone, V. Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.* **110**, 6158 (1999).
39. Barone, V. & Cimino, P. Validation of the B3LYP/N07D and PBE0/N07D computational models for the calculation of electronic g-tensors. *J. Chem. Theory Comput.* **5**, 192–199 (2009).
40. Reta Mañeru, D., Pal, A. K., Moreira, I. D. P. R., Datta, S. N. & Illas, F. The triplet-singlet gap in the m-Xylylene radical: a not so simple One. *J. Chem. Theory Comput.* **10**, 335–345 (2014).
41. Cho, D., Ko, K. C. & Lee, J. Y. Quantum chemical approaches for controlling and evaluating intramolecular magnetic interactions in organic diradicals. *Int. J. Quantum Chem.* **116**, 578–597 (2016).
42. Reta, D., Moreira, I. D. P. R. & Illas, F. Magnetic coupling constants in three electrons three centers problems from effective hamiltonian theory and validation of broken symmetry-based approaches. *J. Chem. Theory Comput.* **12**, 3228–3235 (2016).
43. Veciana, J. & Ratera, I. Polychlorotriphenylmethyl radicals: towards multifunctional molecular materials. In *Stable Radicals: Fundamentals and Applied Aspects of Odd-Electron Compounds* (ed Hicks, R. G.) 33–80 (John Wiley & Sons, Inc., 2010).
44. Maspoch, D. et al. A nanoporous molecular magnet with reversible solvent-induced mechanical and magnetic properties. *Nat. Mater.* **2**, 190–195 (2003).
45. Gomberg, M. An instance of trivalent carbon: triphenylmethyl. *J. Am. Chem. Soc.* **22**, 757–771 (1900).
46. Sabacky, M. J., Johnson, C. S., Smith, R. G., Gutowsky, H. S. & Martin, J. C. Triarylmethyl radicals, synthesis and electron spin resonance studies of sesquioxanthrydyl dimer and related compounds. *J. Am. Chem. Soc.* **89**, 2054–2058 (1967).
47. Li, Z. et al. Cyclo-para-phenylmethine: an analog of benzene showing global aromaticity and open-shell diradical character. *J. Am. Chem. Soc.* **141**, 16266–16270 (2019).
48. Nishino, M., Yamanaka, S., Yoshioka, Y. & Yamaguchi, K. Theoretical approaches to direct exchange couplings between divalent chromium ions in naked dimers, tetramers, and clusters. *J. Phys. Chem. A* **101**, 705–712 (1997).
49. Noodleman, L. & Davidson, E. R. Ligand spin polarization and antiferromagnetic coupling in transition metal dimers. *Chem. Phys.* **109**, 131–143 (1986).
50. Noodleman, L. Valence bond description of antiferromagnetic coupling in transition metal dimers. *J. Chem. Phys.* **74**, 5737–5743 (1981).
51. Caballol, R., Castell, O., Illas, F., Moreira, I. D. P. R. & Malrieu, J. P. Remarks on the proper use of the broken symmetry approach to magnetic coupling. *J. Phys. Chem. A* **101**, 7860–7866 (1997).
52. Moreira, I. D. P. R. & Illas, F. A unified view of the theoretical description of magnetic coupling in molecular chemistry and solid state physics. *Phys. Chem. Chem. Phys.* **8**, 1645–1659 (2006).
53. Lee, C., Wei, X., Kysar, J. W. & Hone, J. Measurement of the elastic properties and intrinsic strength of monolayer graphene. *Science* **321**, 385–388 (2008).
54. Li, Y. et al. Mapping the elastic properties of two-dimensional MoS₂ via bimodal atomic force microscopy and finite element simulation. *npj Comput. Mater.* **4**, 49 (2018).
55. Li, T. et al. Magnetic bistability in a discrete organic radical. *J. Am. Chem. Soc.* **138**, 10092–10095 (2016).
56. Sun, Q. et al. Coupled spin states in armchair graphene nanoribbons with asymmetric zigzag edge extensions. *Nano Lett.* **20**, 6429–6436 (2020).
57. Sahabudeen, H. et al. Highly crystalline and semiconducting imine-based two-dimensional polymers enabled by interfacial synthesis. *Angew. Chem. Int. Ed.* **59**, 6028–6036 (2020).
58. Gale, J. D. & Rohl, A. L. The general utility lattice program (GULP). *Mol. Simul.* **29**, 291–341 (2003).
59. Rappe, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A. III & Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024–10035 (1992).
60. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
61. Zhang, I. Y., Ren, X., Rinke, P., Blum, V. & Scheffler, M. Numeric atom-centered-orbital basis sets with valence-correlation consistency from H to Ar. *N. J. Phys.* **15**, 123033 (2013).
62. Jensen, S. R. et al. The elephant in the room of density functional theory calculations. *J. Phys. Chem. Lett.* **8**, 1449–1457 (2017).
63. Blum, V. et al. Ab initio molecular simulations with numeric atom-centered orbitals. *Comp. Phys. Comm.* **180**, 2175–2196 (2009).
64. Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chem. Acc.* **44**, 129–138 (1977).
65. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
66. Havu, V., Blum, V., Havu, P. & Scheffler, M. Efficient O(N) integration for all-electron electronic structure calculation using numeric basis functions. *J. Comput. Phys.* **228**, 8367–8379 (2009).

Acknowledgements

This work was supported by MICIUN/FEDER RTI2018-095460-B-I00, PID2019-109518GB-I00 and CTQ2017-87773-P/AEI/FEDER (Spanish government, MINECO), MDM-2017-0767 (“María de Maeztu” programme for Spanish Structures of Excellence), 2017SGR13 and 2017SGR348 (Generalitat de Catalunya, DURSI). I.A. is grateful for support from the Alexander von Humboldt Foundation. R.S. acknowledges funding from MINECO under grant agreement FPI PRE2018-084053. The present work was performed using supercomputer resources as provided through grants from the Red Española de Supercomputación.

Author contributions

I.A. and S.T.B. came up with the original concept and prepared the first version of the paper. I.A. carried out DFT (strain/AIMD) calculations on most of the studied 2D-CORFs and structural/electronic analysis for all materials. R.S. carried out DFT (strain) calculations and analysis of the mixed 2D-CORF and calculated the Young’s modulus for all materials. J.R.A., M.D. and I.P.R.M. provided the magnetic analysis. I.P.R.M. performed the magnetic coupling calculations. All authors discussed the results and contributed to the paper preparation. S.T.B. coordinated the project.

Competing interests

The authors declare no competing interests.

Additional information


Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21885-y>.

Correspondence and requests for materials should be addressed to I.A. or S.T.B.

Peer review information *Nature Communications* thanks Liam Wilbraham and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Chapter 4

Design of antiaromatic organic
diradicals with robust triplet
ground state

4.1. Introduction

The limited presence of reported metal-free organic diradical units with a ferromagnetic (FM) response remains a salient point in the current research landscape.¹ As highlighted at the general introduction of this thesis (**Chapter 1**), once these diradicals can be effectively developed and incorporated into materials, they can exhibit a spin-polarized electronic structure, which positions them as sustainable, lightweight alternatives for spintronic devices^{2,3} such as spin filters,⁴ spin valves,⁵ spin transistors,⁶ and spin wires.⁷ One standout advantage of organic units (when, for example, contrasted with metal complexes) is that their composition primarily stems from abundant elements like carbon, hydrogen, and nitrogen. This suggests that potential devices made from these units could be cost-effective, widely accessible, and potentially reduce the environmental footprint associated with producing and distributing such organic spintronic devices.

However, their widespread usage in practical application depends on having a high Curie temperature, T_C , which marks the transition from a macroscopic ferromagnetic behavior to a paramagnetic state.⁸ Central to achieving this high T_C is the presence of a significant energy gap, ΔE_{ST} , between the singlet (S) and triplet (T) states that must propagate through the whole solid to have bulk ferromagnets. Accordingly, the wider this gap, the greater the thermal energy needed to transition to the singlet state, thereby counteracting paramagnetic effects. Hence, this ensures the retention of the FM behavior of diradicals even at high temperatures. However, the challenge remains, as recent research that successfully characterizes diradicals with such a robust ΔE_{ST} gap above room temperature⁹⁻²¹ or moderately below room temperature²²⁻²⁹ is quite scarce.

In this chapter, we have focused on designing organic diradicals with the aforementioned properties. Overall, our efforts have resulted in the design of a new set of compounds based on pentalene, possessing a robust triplet ground state, where $\Delta E_{ST} \gg 0$. To guide the reader, the structure and content of this introduction are summarized in the following order:

- Overview of the required conditions to achieve a ferromagnetic ground-state from an electronic structure point of view (**Section 4.1.1**)
- Presentation of some qualitative models that, based on the conditions stemming from quantum mechanics, can aid the quick design of organic diradicals with FM interactions (**Section 4.1.2**).
- Presentation of the set of compounds studied in this chapter, which result to be a subset of antiaromatic molecules based on pentalene (**Section 4.1.3**).

After that, the results are presented and discussed in **Section 4.2**, followed by the Conclusions (**Section 4.3**) and, finally, the inclusion of the draft of the publication (**Section 4.4**), which is still in preparation.

4.1.1. Insights from the electronic structure point of view

This section seeks to identify the fundamental components that assist in both qualitative and quantitative predictions of the ΔE_{ST} gap. At its core, it is an electronic problem and, thus, the following electronic Hamiltonian (\hat{H}_{el}) must be considered:

$$\hat{H}_{el} = \hat{T}_e + \hat{V}_{eN} + \hat{V}_{ee} \quad (4.1)$$

Here, \hat{T}_e and \hat{V}_{eN} represent the standard electron kinetic energy and electron-nucleus electrostatic potential operators. In addition, the term \hat{V}_{ee} corresponds to the electron-electron electrostatic repulsion.

For an introductory understanding of the specific conditions that a system must meet to have a high-spin ground state, one can start from the most simple case: two electrons in two atomic orbitals (ψ_1 and ψ_2). The goal is gaining preliminary insight about the preferential spin state of the system. Overall, there are six different arrangements of two electrons in two spin orbitals maintaining the anti-symmetry of the wave function. Drawing from the notation introduced in the Methodology Chapter (**Chapter 2**), these configurations are:

$$\begin{aligned}
 |\Psi_{S_1}\rangle &= |\psi_1\bar{\psi}_1| \\
 |\Psi_{S_2}\rangle &= |\psi_2\bar{\psi}_2| \\
 |\Psi_{T_1}\rangle &= |\psi_1\psi_2| \\
 |\Psi_{T_2}\rangle &= |\bar{\psi}_1\bar{\psi}_2| \\
 |\Psi_{S_3}\rangle &= 1/\sqrt{2} (|\psi_1\bar{\psi}_2| + |\bar{\psi}_1\psi_2|) \\
 |\Psi_{T_3}\rangle &= 1/\sqrt{2} (|\psi_1\bar{\psi}_2| - |\bar{\psi}_1\psi_2|)
 \end{aligned} \tag{4.2}$$

where Ψ_{S_1} and Ψ_{S_2} correspond to the charge transfer Slater determinants in which ψ_1 or ψ_2 is doubly occupied, respectively. On the other hand, Ψ_{T_1} and Ψ_{T_2} are the triplet states characterized by $S_z = \pm 1$ and, lastly, Ψ_{S_3} and Ψ_{T_3} correspond to the antisymmetrized singlet and triplet states with $S_z = 0$. In order to calculate the singlet-triplet gap in its most basic approximation, the difference between the expectation energy values of Ψ_{S_3} and Ψ_{T_3} , which can be determined as

$$\begin{aligned}\langle \Psi_{S_3} | \hat{H} | \Psi_{S_3} \rangle &= h_1 + h_2 + J_{12} + K_{12} \\ \langle \Psi_{T_3} | \hat{H} | \Psi_{T_3} \rangle &= h_1 + h_2 + J_{12} - K_{12}\end{aligned}\tag{4.3}$$

where h_1 and h_2 correspond to the one-electron contributions stemming from the sum of \hat{T}_e and \hat{V}_{eN} terms of the Hamiltonian. Likewise, the terms J_{12} and K_{12} represent the Coulomb repulsion integral and exchange integral, respectively (please do not confuse the exchange magnetic coupling, J_{AB} with neither the Coulomb repulsion integral, J_{12} , nor the exchange integral K_{12}). These integrals can be further expressed as:

$$J_{12} = \iint_{\mathbb{R}^3} \frac{|\psi_1(\mathbf{r}_1)|^2 |\psi_2(\mathbf{r}_2)|^2}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2\tag{4.4}$$

and

$$K_{12} = \iint_{\mathbb{R}^3} \frac{\rho(\mathbf{r}_1)\rho^*(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2\tag{4.5}$$

where $\rho(\mathbf{r})$ corresponds to the overlap density defined as $\rho(\mathbf{r}) = \psi_1(\mathbf{r})\psi_2(\mathbf{r})$ and $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$ is the distance between electrons indexed as 1 and 2, respectively. Now, given the results of equation 4.3, the singlet-triplet gap, ΔE_{ST} , can be succinctly expressed as:

$$\Delta E_{ST} = \langle \Psi_{S_3} | \hat{H} | \Psi_{S_3} \rangle - \langle \Psi_{T_3} | \hat{H} | \Psi_{T_3} \rangle = 2K_{12}\tag{4.6}$$

Notably, as K_{12} is strictly greater than or equal to zero, it can be deduced that the relative energy between the singlet and triplet states is always positive or zero. In other words, the triplet state is invariably expected to be either the ground state or degenerate compared to the singlet state. While the former derivations intentionally omits four out of the six possible states, such a basic

example serves as an illustrative foundation: K_{12} (i.e., the exchange integral) is the principal driving force behind significant positive ΔE_{ST} gaps.

When aiming for a more accurate representation of the ΔE_{ST} gap, or at the very least, a model conceiving antiferromagnetic (AFM) contributions, the picture becomes inherently more complex. Over the years, various models have been formulated in an attempt to address this complexity. Specifically, the models proposed by Hay, Thibault and Hoffmann³⁰ (HTH), Goldberg³¹ and Dougherty³² (GD) and Girerd, Journaux and Kahn³³ (GJK) pioneered this field. These set of models stand as invaluable interpretative tools for the results obtained from *Ab Initio* SCF computations.

Despite the unique nuances of these models, all of them are sustained under similar conceptual frameworks: the different orbitals interact to achieve a more generalized valence bond description, further proving that this feature stabilizes the singlet state. This refined description then emphasizes specific subsets of the Slater determinants, as the ones presented at the beginning of this section (equation 4.2). While a meticulous breakdown of the procedure for each model can be found in the cited literature, the discussion here is focused primarily on the insights extracted from them, particularly since these provide the essential groundwork for designing the molecular structures central to this chapter.

Remarkably, it emerges a consensus on the formulation of the ΔE_{ST} gap across these different models. For example, the HTH treatment leads to

$$\Delta E_{ST}^{\text{HTH}} = 2K_{ab} - \frac{(\epsilon_1 - \epsilon_2)^2}{J_{aa} - J_{ab}} \quad (4.7)$$

where ϵ_1 and ϵ_2 represent the energy of the original canonical singly-occupied molecular orbitals (SOMOs) and the sub-indices a and b distinguish the original canonical orbitals (indexed by 1 and 2) from those derived from an orthogonal linear combination of those orbitals (indexed alphabetically). In parallel to the HTH model, the GJK treatment employs Natural Magnetic Orbitals instead of the canonical orbitals, which leads to the following formulation for the singlet-triplet gap:

$$\Delta E_{ST}^{\text{GJK}} = 2K_{ab} - 4\beta S_{ab} \quad (4.8)$$

Note that the first term aligns seamlessly with what it is derived from the HTH approach. However, the antiferromagnetic (AFM) contribution manifests as $4\beta S_{ab}$, where S_{ab} is the orbital overlap and β is the resonance integral.

Despite the apparent disparities in the AFM term between the HTH and GJK models, both converge towards analogous qualitative interpretations. For a more tangible grasp on this, one can reference to the Generalized Wolfsberg-Helmholtz formula,³⁴ that establishes a linear relationship between S_{ab} and β . Accordingly, as β represents the difference in energy (orbital splitting) between the two orbitals, the AFM term becomes proportional to $(\epsilon_1 - \epsilon_2)^2$, mirroring the trend observed in the HTH model at equation 4.7.

In summary, the intricacies of the singlet-triplet gap, as derived from both the HTH and GJK models, coalesce into a foundational insight: the gap is shaped by the interplay of two pivotal terms. The first, a ferromagnetic (FM) term, is dominated by the exchange integral, K_{12} . The second is an antiferromagnetic (AFM) contribution, which is intimately linked to the square of the energy difference between the singly-occupied molecular orbitals, $(\epsilon_1 - \epsilon_2)^2$. This duality highlights

the delicate interplay and competition between FM and AFM interactions that ultimately determine the electronic nature of the gap.

Hence, achieving a ferromagnetic ground state reduces to the following problem: (i) how to arrange within an organic molecule SOMOs that lay degenerate; and (ii) how to concurrently optimize the conditions to bolster a substantial K_{12} value.

4.1.2. Qualitative models for ground-state properties of diradicals

The most straightforward answer to question (i) formulated above would be to consider molecular symmetry. Group theory demonstrates that certain symmetries can give rise to doubly degenerate MOs. When these symmetries yield degenerate frontier MOs, their energy difference, $(\epsilon_1 - \epsilon_2)$, becomes zero, effectively leading to the vanishing of the AFM contribution according to the electronic structure considerations described so far. This type of MO degeneracy is characteristic of molecules with a three-fold (or higher) rotational axis or those with D_{2d} symmetry.³⁵ For instance, benzene has such a doubly degenerate molecular orbitals, as it belongs to the D_{6h} point group. Yet, benzene typically possesses enough electrons to also fill the degenerate frontier orbital, preventing the formation of open-shell electronic structures in practice. This highlights that, while MO symmetry is crucial, it is not a sufficient condition to guarantee an open-shell ground state.

Expanding upon the principles of molecular symmetry, the Longuet-Higgins³⁶ analysis offered an innovative insight into the design of high-spin hydrocarbons. This theory made significant strides in identifying compounds where the number of electrons harmoniously matches the number of degenerate MOs, anchoring in

the principles of Huckel MO theory³⁷ and Hund's rule. Central to this approach is that the energy levels of a number N of $2p_z$ atomic orbitals combining into N π -molecular orbitals, align predictably as established by the pairing theorem.³⁸ A key observation of Longuet-Higgins was that the maximal number of double bonds in any given resonance structure of a hydrocarbon (M) aligns with the number of degenerate non-bonding molecular orbitals (NBMOs), which in turn reflects the number of electrons populating them.³⁵ In short, this relationship can be represented as:

$$N_{\text{NBMOs}} = N_{\text{unpaired}} = N - 2M \quad (4.9)$$

Following the analysis of Longuet-Higgins, a wide set of alternant hydrocarbons displaying two singly-occupied NBMOs have been designed, which are usually referred to as non-Kekulé hydrocarbons (i.e., cannot be assigned to a closed-shell Kekulé structure).

For illustrative purposes, consider two isomers of xylylene: p-xylylene and m-xylylene, depicted in Figure 4.1a and Figure 4.1b, respectively. First, p-xylylene possesses 8 carbon atoms and 4 double bonds can be counted in one of its resonance structures (the closed-shell quinoidal resonance structure). Then, by the application of equation 4.9, zero NBMOs are anticipated and, thus, a closed-shell ground state is predicted at a first order approximation. In contrast, m-xylylene, displaying the same carbon count, has only 3 double bonds in all its resonance structures, which results in two NBMOs populated by two electrons, yielding a triplet ground state by the application of the Hund's rule. The comparison between p- and m-xylylene constitutional isomers illustrated in Figure 4.1a-b

provide valuable insights, such that subtle topological discrepancies can significantly alter the fundamental ground state properties.

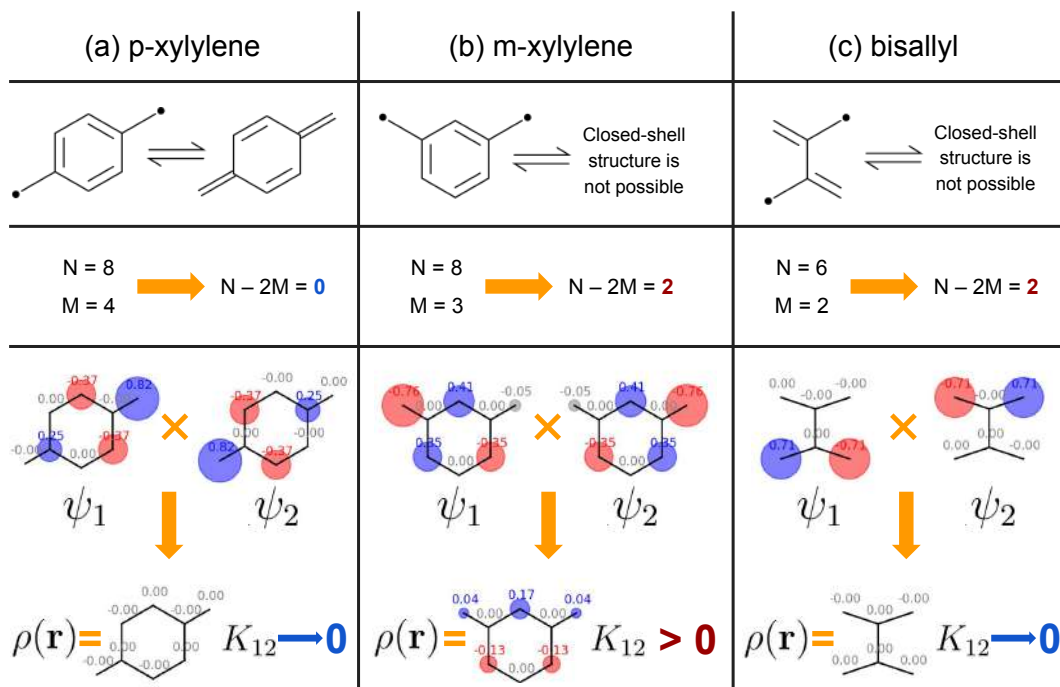


Figure 4.1: Comparison of three compounds structured as a table. In the first row, the names of the compounds are identified (namely, (a) p-xylylene, (b) m-xylylene, and (c) bisallyl). The second row showcases their respective Lewis structures along with the values for N and M . The third row presents the Longuet-Higgins calculation and prediction regarding the number of NBMOs. Lastly, the fourth row visualizes the molecular orbitals, electron distribution, $\rho(\mathbf{r})$, and the Borden-Davidson prediction of K_{12} for each compound.

However, the ground spin states of some alternant hydrocarbons have been found to be contradictory to the Longuet-Higgins prediction. For example, whereas the ground state for tetramethylethane (also referred to as bisallyl, see Figure 4.1c) is predicted to be a triplet according to Longuet-Higgins' analysis. The experimental characterization demonstrates, however, that the singlet and triplet states are nearly degenerate,³⁹ unveiling thus, that the Hund's

rule cannot be applied so straightforwardly in this case. The evidences derived from the bisallyl diradical stress the importance of addressing the second part of the question formulated in the last paragraph of the previous section, i.e., concurrently bolster the FM term dominated by the magnitude of K_{12} .

More recently, models such as the ones pioneered by Ovchinnikov⁴⁰ and Lieb⁴¹ offered further insights on predicting the multiplicity of the ground state of hydrocarbons such as bisallyl. However, the application of these models has been notably limited in this thesis in favor of the model of Borden and Davidson,⁴² (BD), which is also capable of accounting for such complexity. According to the Borden-Davidson model, non-Kekulé hydrocarbons can be classified into two categories with respect to the topology of their singly-occupied NBMOs. In the first one, the NBMOs are found to be *disjoint*, meaning the atomic orbitals contributing to one NBMO do not overlap with those forming the other. This results in the SOMOs being localized in separate regions of the carbon scaffold. Instead, in the second category, the NBMOs comprise a *non-disjoint* set, where some of the atomic orbitals contribute to both NBMOs. As illustrated in Figure 4.1, the NBMOs of p-xylylene and bisallyl clearly indicates that both compounds fall into the disjoint category, as the $2p_z$ basis of the carbon scaffold contributing to the first SOMO, ψ_1 , do not contribute to ψ_2 , and vice versa. Instead, the $2p_z$ basis centered at the carbons of the benzene ring of m-xylylene contribute to both ψ_1 and ψ_2 SOMOs, positioning m-xylylene into the non-disjoint category.

This classification holds significant implications in the FM component of the singlet-triplet gap, particularly concerning the magnitude of the exchange integral K_{12} . In this regard, a close examination of equation 4.5 reveals that the exchange integral attains its maximum values when the electrons are expected

to be close to each other, i.e. $\mathbf{r}_{12} = |\mathbf{r}_1 - \mathbf{r}_2| \rightarrow 0$. Hence, the magnitude of K_{12} is largely associated to the extreme values of the overlap density,⁴³ $\rho(\mathbf{r})$. Accordingly, diradicals comprising a pair of SOMOs that share a significant non-null density region inherently display a non-zero $\rho(\mathbf{r})$ and, by extension, a non-zero K_{12} . Hence, if the SOMOs are spatially separated (forming a disjoint set), the overlap density is expected to be close to zero, leading also to $K_{12} \rightarrow 0$, which, in turn, results in a nearly equivalent energy of the singlet and triplet states according to equation 4.7. Instead, when the NBMOs are non-disjoint, the SOMOs have a substantial overlap with those of the other, sharing common regions within the carbon scaffold (see bottom panels of Figure 4.1 for $\rho(\mathbf{r})$ and K_{12}). As a result, the SOMOs topology leads to non-vanishing regions of $\rho(\mathbf{r})$ when these form a non-disjoint set, leading subsequently to $K_{12} \gg 0$ and a significant ΔE_{ST} gap, as inferred from equation 4.7. Overall, the classification of SOMOs into disjoint and non-disjoint sets provided by the Borden-Davidson model still offers a robust framework for qualitatively predict the ground states of molecules such as p-xylylene and m-xylylene. In addition, this (non-)disjoint classification is able to capture the complexities inherent in predicting the ground state of bisallyl diradical, an aspect where the Longuet-Higgins model falls short.

Overall, the qualitative models discussed earlier have played a critical role in the design of diradicals exhibiting FM interactions between their two electrons. The approach has involved a methodical process: (i) first, Longuet-Higgins' model is applied to determine whether the molecular topology results in two half-filled (i.e. singly-occupied) NBMOs, thus securing non-Kekulé electronic structures and a null AFM contribution to the ΔE_{ST} gap. (ii) After that, the magnitude of K_{12} can be evaluated to determine the (non-)disjoint character of

the SOMOs, leveraging Borden and Davidson’s model for those candidates meeting the first criterion. Hence, molecules that satisfy conditions (i) and (ii) can be selected for further study and computational analysis using DFT and multiterminantal SCF calculations, where the magnetic properties beyond first-order approximations can be explored. This structured approach has enabled the systematic and quick identification of diradicals with desired FM interactions.

4.1.3. Potential candidates: pentalene-based diradicals

Stability is a desired attribute for radical building blocks in many applications. However, organic radicals are prone to undergo dimerization, recombination processes or disproportionation reactions,⁴⁴ leading to a closed-shell structure of the resultant products and, thus, usually being categorized as relatively unstable. Overall, there are three main strategies to enhance the stability of radicals.⁴⁵ (i) The first strategy consists of the steric protection of the spin-bearing units. This involves introducing bulky substituents to shield the radical center, thereby hindering unwanted reactions.^{46–48} While this approach leads to kinetically persistent radicals (as illustrated in Figure 4.2a), it presents an arguable downside: the bulky groups can isolate the radical centers from the rest, resulting in disjointed SOMOs that effectively dampen the desired magnetic properties. (ii) The second strategy does not rely on physical barriers but, instead, on electronic stabilization. By extending the conjugation around the radical center, the unpaired electron of the radical becomes electronically delocalized.^{49–51} This not only lowers its overall electronic energy but also dilutes the spin density, bolstering both its stability (see an example in Figure 4.2b).⁴⁵ (iii) Finally, the usage of lightweight heteroatoms, like oxygen, sulfur and nitrogen, has also been proven beneficial. Compared to full-carbon radicals, these heteroatom-based radicals

exhibit reduced reactivity towards oxidation processes and are less prone to undergo dimerization reactions.^{45,52,53} An example of this strategy can be seen in Figure 4.2c.

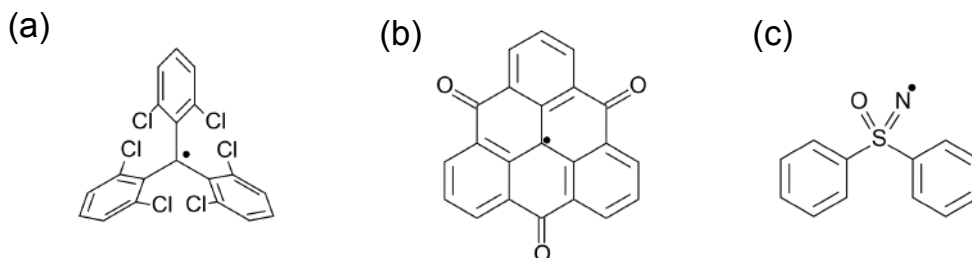


Figure 4.2: *Examples of the three main strategies employed to enhance the stability of radicals. (a) Steric protection of the radical center using bulky substituents. (b) Electronic stabilization through extended conjugation around the radical center. (c) Incorporation of lightweight heteroatoms to reduce radical reactivity.*

The former strategies have been widely employed to produce a large set of diradicals with a triplet ground state based on the aforementioned *m*-xylylene building blocks.^{54–56} However, the work presented here is an attempt to shift away from xylylene-based units in favor of other building blocks, seeking to explore and harness the potential of other promising and innovative units for similar applications. In this regard, a recent study of Winter and co-workers⁵⁷ proposes a starkly different approach in relation to the strategy (ii). While the customary strategy instrumentally utilizes an aromatic π -conjugated unit to dilute the spin density of the spin-bearing unit, $\cdot\mathbf{R}$, the process is at the expense of losing some amount of aromatic character (as depicted in Figure 4.3a). That is, the aromatic moiety, having $(4N+2)$ π -electrons, donates electronic density to stabilize the radical, resulting in a dilution of its aromatic character and the emergence of resonance structures with $(4N+1)$ π -electrons. Conversely, the approach pro-

posed by Winter involves an antiaromatic moiety. As shown in Figure 4.3b, this scenario is mutually beneficial for both the spin-bearing unit and the antiaromatic unit. That is, by diffusing the spin density, the inherent ($4N$) antiaromatic nature of the moiety is attenuated. Here, the radical center donates or withdraws electrons, which reduces the antiaromatic character, bringing about stabilization towards a substantially contributing ($4N-1$) resonance structure. This scenario establishes a cooperative effect where the radical and the substituent stabilize each other, contrasting the traditional unidirectional stabilization seen with the classical aromatic approaches. Note that, while the discussion around Figure 4.3 is centered on the donation of an electron from the antiaromatic coupler to the spin-bearing unit, the opposite process could also take place. That is, it is also possible that the radical center donates the unpaired electron to the aromatic or antiaromatic unit, which would lead to $4N+3$ or $4N+1$, respectively (that are also a $4N+1$ and $4N-1$ with another value of N).

While the previous strategy has been proven experimentally effective for a set of monoradical compounds, their topology does not permit an intuitive way to extend their structure to account for two or more spin-bearing units and, thus, the potential formation of a diradical would be structurally impeded. Moreover, the topology of the carbon scaffold of the presented units does not straightforwardly satisfy the principles of the Borden-Davidson model, arguably hampering the design of diradicals presenting a high-spin ground state. For this reason, we leveraged the principles of their work to select a different set of antiaromatic units that, based on the previous premise, also possess the electronic structure conditions to lead to a diradical with a triplet ground state.

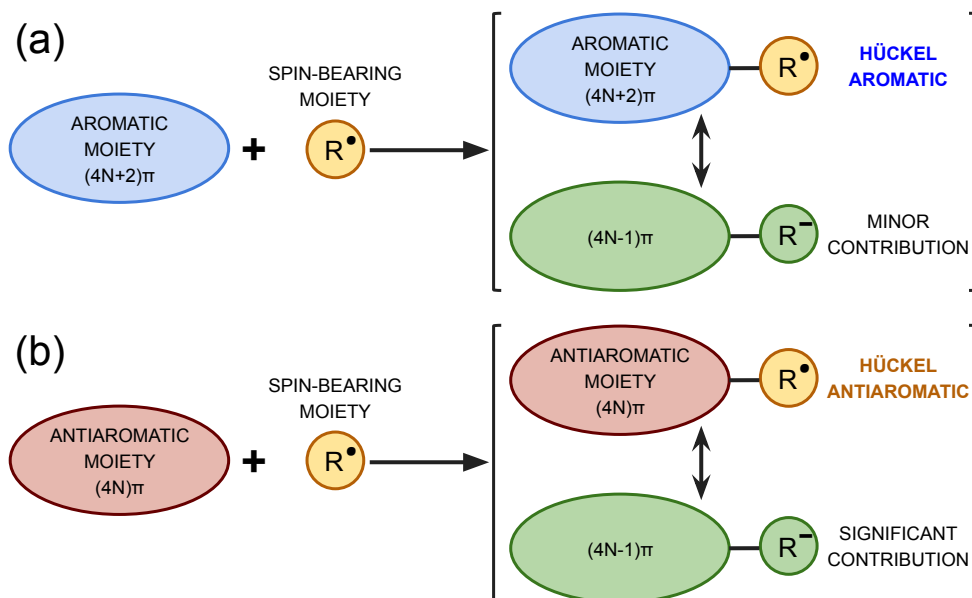


Figure 4.3: Schematic representation of the radical stabilization strategies. (a) The blue moiety represents the traditional aromatic system with $(4N+2)\pi$ electrons interacting with a radical (depicted by the yellow circle), leading to a radical with a minor contribution of a $(4N+1)\pi$ resonance structure and a formally charged residue (both depicted in green). (b) In Winter's approach, the antiaromatic moiety with $(4N)\pi$ electrons (depicted in red), and its interaction with a radical results in a stabilized structure with $(4N-1)\pi$ electrons with a significant contribution. Overall the coloring scheme indicate a descending order of potential stability while progressing from blue \rightarrow green \rightarrow yellow \rightarrow red.

Among the wide amount of antiaromatic units documented in textbooks and scientific publications, the study presented here particularly centers on pentalene derivatives. Pentalene (see Figure 4.4a) stands out as one of the most extensively researched antiaromatic compounds, often employed to probe and test the nature of antiaromaticity itself. Although pentalene has been traditionally viewed as an archetype of instability,^{60–62} recent advancements in synthetic techniques^{63–65} and innovative molecular designs have pivoted the study of certain pentalene derivatives from merely academic curiosities to practical applications. Among

others, structures that have gained attention include those with $[n]$ acene ring insertions^{65–67} (see Figure 4.4b), n -doped hetero-substitutions^{68–70} (Figure 4.4c), and dibenzoannulation^{71–75} (Figure 4.4d), all of which have been successfully isolated and characterized. Beyond their recognized inherent instability, antiaromatic molecules are known for their narrow HOMO-LUMO gap.^{76–78} This feature, coupled with their newfound synthetic viability, has encouraged notable efforts in the synthesis and characterization of structures, culminating in a set of nanohoops^{58,70,75,79–81} and covalent (metal)organic frameworks^{59,82} (as the ones illustrated in Figures 4.4e and 4.4f) with potential applications as batteries for energy storage solutions due to its proved amphoteric redox behavior.⁵⁹

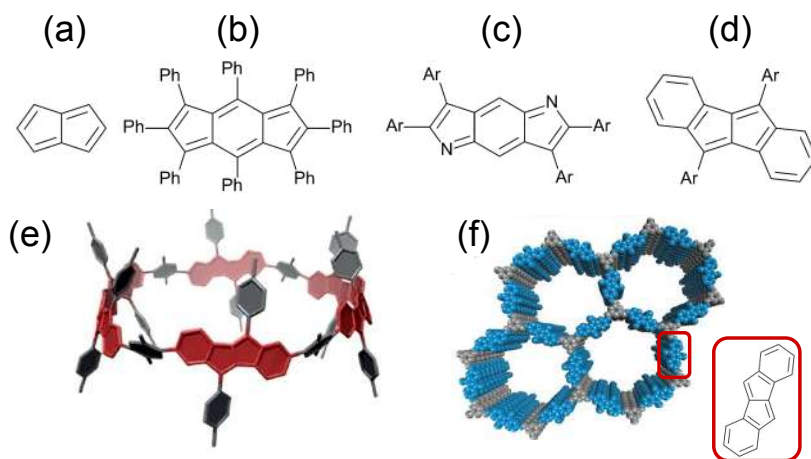


Figure 4.4: (a) The basic structure of pentalene. (b) An example with a fused ring within the pentalene scaffold, indicative of $[n]$ acene ring insertions. (c) A pentalene derivative featuring nitrogen substitutions, representing n -doped hetero-substitutions. (d) An extended structure with additional benzene rings on either side, representing the dibenzoannulation. (e) A three-dimensional rendering of a nanohoop, constituted by six dibenzoannulated pentalenes. (f) A pentalene-based covalent organic framework, with an emphasized region highlighting a repeated unit similar to the zoomed structure on the right, illustrating the fundamental building blocks of the framework. (Images at (e) and (f) have been adapted from the original publications^{58,59})

The structure of the antiaromatic molecules presented in Figure 4.4a-d have inspired us to elaborate various routes to systematically explore the ability of pentalene as a coupler to bolster FM interactions between unpaired electrons. Initially, we explored all feasible connectivities of open-shell cores (**R**) to the pentalene structure (illustrated as route 1 in Figure 4.5), resulting in the whole set of doubly-substituted pentalene constitutional isomers (**R**₂-**PI**). Within this context, we considered the methylenyl (**R=My**) group, marked in gray in Figure 4.5, as a basic model. However, due to its inherent instability and propensity for disproportionation or recombination, it serves better as a theoretical construct rather than a practical option. Shifting our focus to more viable alternatives, we constrained our selection on open-shell units that exhibit persistence under ambient conditions, aligning with increasing the kinetic stability by means of the steric protection strategy previously discussed. In this regard, diphenylmethyl (**R=DPM**) cores emerged as suitable options, as their steric bulk shields the spin-bearing centers and reinforces the kinetic stability of the diradicals. To further enhance the kinetic stability, we included polychloro-diphenylmethylenyl (**R=PDM**) cores, specifically the di-(2,6-dichloro)-diphenylmethylenyl derivative. The introduction of bulky halogen substituents in the ortho positions of the phenyl rings is designed to further bolster the thermal resilience of the diradicals.

Upon selecting the **R**₂-**PI** isomers with the most promising potential for high-spin ground states, we examined the impact of incorporating [*n*]acene into the pentalene framework (route 2 in Figure 4.5). This led to a series of **R**₂-substituted dicyclopenta-[*n*]acenes and *n*-doped diaza-dicyclopenta-[*n*]acenes, denoted to as **R**₂-**DPA**[*n*] and **R**₂-**DADPA**[*n*] for **X=CH** and **X=N**, respectively (see Fig-

ure 4.5). The bare **DPA**[*n*] and **DADPA**[*n*] series are known for their gradually diminishing antiaromaticity as *n* increases.^{66,67} However, according to the stabilization mechanism based on antiaromatic couplers, as the antiaromatic character is diluted in progressing through the series, the stabilization effect of the resulting

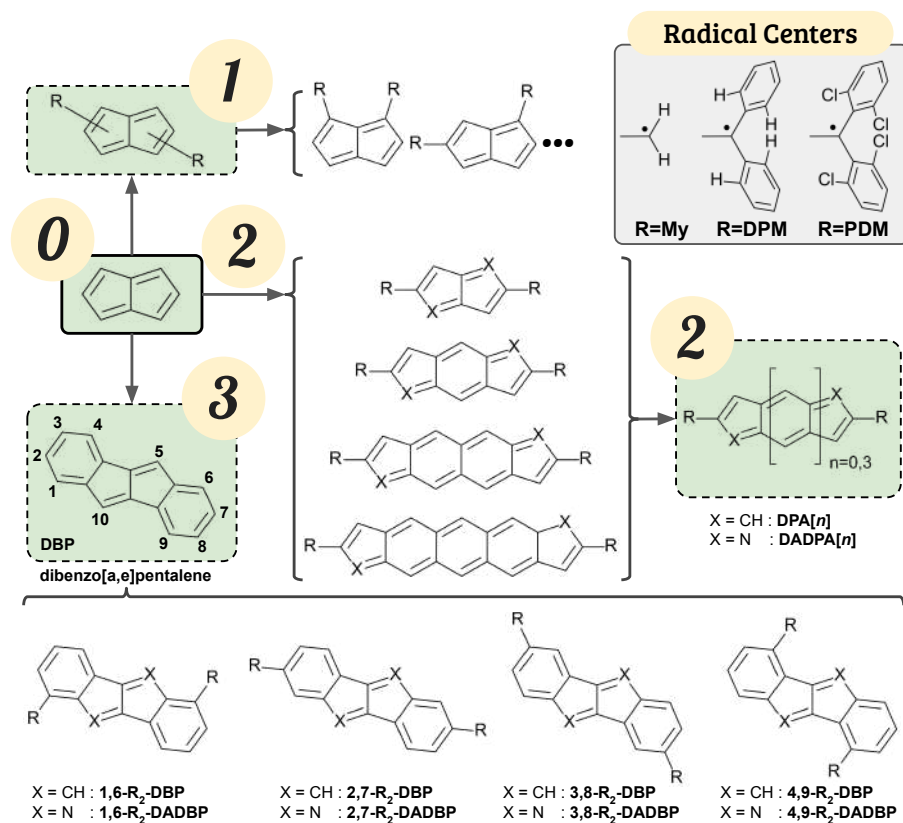


Figure 4.5: Schematic representation of the explored routes for designing antiaromatic molecules based on the pentalene structure. Structure 0 illustrates the basic pentalene unit. Route 1 shows the connection of open-shell cores to the pentalene structure resulting in a set of doubly-substituted pentalene isomers. Route 2 expands on the pentalene framework by incorporating the [*n*]acene units, leading to R_2 -substituted dicyclopenta-*n*]acenes and *n*-doped diaza-dicyclopenta-*n*]acenes. Route 3 illustrates dibenzoannulated derivatives of pentalene, including dibenzopentalene (DBP) and diaza-dibenzopentalene (DADBP) structures. The various molecular cores, including My, DPM, and PDM, and their radical centers are also depicted in grey at the top-right corner of the figure.

R₂-DPA[n] and **R₂-DADPA[n]** diradical units is then expected to decrease. Overall, the purpose of exploring this route is twofold: testing the stabilization mechanism based on antiaromatic couplers and examining the influence on the singlet-triplet gap across the [n]acene series.

Lastly, our exploration extended to dibenzoannulated pentalene derivatives (route 3 in Figure 4.5). This inclusion is motivated by the prevalent success in synthesizing pentalene derivatives based on dibenzopentalene (**DBP**) and diaza-dibenzopentalene (**DADBP**) moieties. As corroborated experimentally, **(DA)DBP** units exhibit significantly reduced antiaromatic characteristics, resulting in one of the most employed choices in the aforementioned nanohoops and covalent organic frameworks applications. As with route 1, we investigated the influence of connectivity on both the antiaromatic character and the singlet-triplet gap.

While the previous objectives delved into assessing the magnetic properties of the pentalene-based diradicals, it is also fundamental to assess their derived stability. According to the mechanism proposed by Winter and co-workers,⁵⁷ cooperative stabilization effects between the bare antiaromatic unit and the spin-holding precursors emerge when covalently linked: the unpaired electron delocalizes towards the antiaromatic coupler, contributing to alleviate the overall antiaromatic character. Although cooperative stabilization effects offer a glimpse into the expected behavior of the pentalene-based diradicals, further proofs must be provided, specially when dealing with such an archetypal unstable precursor as it is pentalene.

However, stability itself remains a concept without a universal reference point, deeply embedded in context-dependent inquiries such as "Stable compared to what?" or "Under what conditions?" From the experimental perspective, stability can be empirically observed in chemical systems under various conditions. However, it is particularly challenging to gauge from the computational standpoint. That is, simulating laboratory conditions, including the medium, temperature, and pressure, within theoretical frameworks, introduces additional layers of complexity to the precise prediction and understanding of this concept.

In the context of the poly-cyclic, fully-conjugated systems discussed in this chapter, the (anti)aromatic character can provide tentative insights into the stability of these systems. Yet, aromaticity is also complex to quantify since it is not a quantum observable nor can be generally measured by any Hamiltonian operator. Although the direct quantification of aromatic stability is elusive, indirect methods can aid the evaluation of aromatic characteristics. Such methods include (i) assessing the response of a molecule upon the effect of an external magnetic field by means of NICS among others (as elaborated in the methodology chapter i.e. **Chapter 2**), which results in induced ring currents within the π -system. On the other hand, (ii) aromaticity insights can be derived from structural analysis. In this regard, examining atomic arrangements, bond lengths, and patterns within the carbon scaffold can reveal the degree of electron delocalization and resonance stabilization inherent to the (anti)aromatic system, giving rise to the well-known set of bond length alternation (BLA) analysis. When these structural insights are combined with magnetic response behavior, they provide a multifaceted picture of aromaticity.

Nonetheless, recent studies challenge the notion of a straightforward correlation between (anti-)aromaticity and (un)stability,^{83,84} casting further doubt on the reliability of aromaticity as a surrogate criterion of stability. Consequently, it becomes critical to (iii) evaluate stability by measuring the energy of the system during carefully orchestrated reactions and processes (further discussion in following sections). It is then the combination of (i) response to specific external stimulus, (ii) structural analysis, and (iii) energy-based assessments that leads to the most accurate picture of the stability.

4.2. Results and discussion

4.2.1. Prediction of the ground-state multiplicity of R_2 -PI isomers

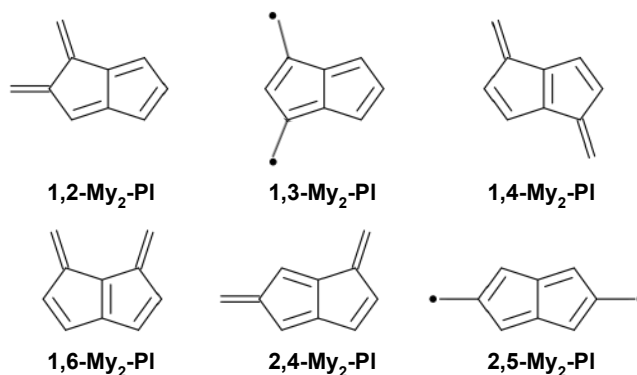


Figure 4.6: Lewis structures of all the possible My_2 -substituted pentalene isomers. Among the whole set of Lewis structures, the figure collects those with the highest number of double bonds.

This section is devoted to inspect the ground state characteristics of pentalene isomers discussed earlier (see route 1 in Figure 4.5). Although most of the analyses carried out in the following sections employ SCF calculations, this specific assessment only requires Longuet-Higgins and Borden and Davidson (non-)disjointness analysis along with Hückel MO calculations to determine the MO topology. In this regard, Figure 4.6 presents some of the Lewis structures of all

possible **My₂-PI** constitutional isomers. While numerous resonance structures can be derived for each isomer, the figure collects those with the highest number of double bonds. Given that all **My₂-PI** isomers have the same carbon count (N=10), the Longuet-Higgins model predicts non-zero NBMOs solely for isomers with 4 double bond pairs, i.e. essentially those without a closed-shell quinoidal Kekulé structure. Now, in accordance with the results collected in Figure 4.6, only **1,3-My₂-PI** and **2,5-My₂-PI** isomers meet this criterion, making them the leading contenders for a triplet ground state as per Longuet-Higgins criterion.

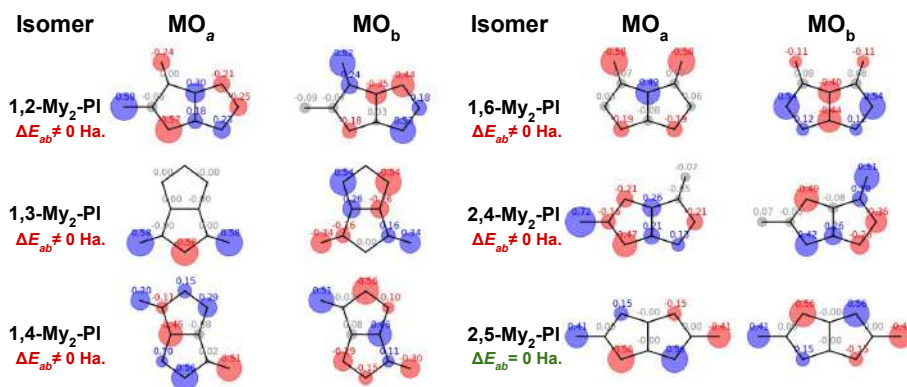


Figure 4.7: *SOMOs topology and ΔE_{ab} (gap between SOMOs indexed as ‘a’ and ‘b’) for all the structural isomers of $My_2\text{-DPA}[0]$. While all the compounds are found to possess a pair of non-disjoint SOMOs, the only isomer displaying a gap of $\Delta E_{ab} = 0$ is $2,5\text{-My}_2\text{-DPA}[0]$.*

To ensure a significant singlet-triplet gap, ΔE_{ST} , we have examined the MO topology of the frontier molecular orbitals across the **My₂-PI** series. Under the assumption of planarity, Hückel MO is valid and, thus, provides a computationally-efficient method to discern the MO topology and determine the (non-)disjoint character of the frontier orbitals. In accordance with the Longuet-Higgins predictions, the Hückel MO calculations (refer to Figure 4.7) reveal an absence of NBMOs for all **My₂-PI** constitutional isomers with a closed-shell

quinoidal resonance structure. Likewise, Hückel MOs topology also confirms that the MOs of **2,5-My₂-P1** are not only degenerate but also form a non-disjoint set. However, the MO topology for **1,3-My₂-P1** displays just one of the two expected NBMOs. Notably, this indicates that the **1,3-My₂-P1** MOs are not truly degenerate, resulting in a potential AFM contribution on the ΔE_{ST} gap and the inappropriate usage of Hund's rule. In summary, the **2,5-My₂-P1** connectivity emerges as the only candidate among the set displaying the required conditions to exhibit a considerable ΔE_{ST} gap and, consequently, it has been the focal point of study in the subsequent subsections.

4.2.2. Qualitative prediction of the ΔE_{ST} gap of R_2 -(DA)DPA

In a similar vein, the prediction of the ground state multiplicity can be carried out in a first approximation for the whole **My₂-DPA[n]** series. As illustrated for **My₂-DPA[0]** (i.e. **My₂-P1**) and **My₂-DPA[3]** in Figure 4.8a-b respectively, the MO topology at a Hückel level of theory is found to be consistent with a triplet ground state as per Longuet-Higgins and Borden-Davidson models. However, recalling the initial discussion in this chapter, the application of ferromagnetically coupled units require a large Curie temperature, T_C , and thus, a large ΔE_{ST} gap.

While the Borden-Davidson model can distinguish between molecular orbital typologies that result in either a triplet or singlet ground state, further insights are required to determine which compounds among the **My₂-DPA[n]** series exhibit a sufficiently large ΔE_{ST} to achieve high values of T_C . It is reasonable to guess by means of general chemical knowledge that all of these molecules display a different ΔE_{ST} . However, Borden-Davidson model, in its current form, is not able to provide quantitative approximation to the ΔE_{ST} gap, but just *guarantee* the triplet ground state in a first approximation.

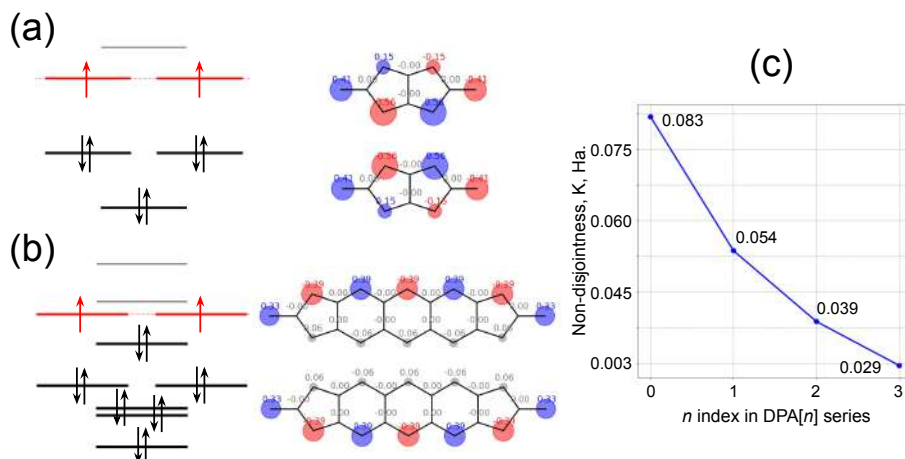


Figure 4.8: Energy levels, occupation and SOMO projected onto the carbon scaffold for (a) My_2 -DPA[0] and (b) My_2 -DPA[3] compounds. For the pair of the two most non-disjoint SOMOs, see inset coefficients of the corresponding linear combination. (c) Non-disjointness parameter (K), computed for the first 4 ($n=0,3$) My_2 -DPA[n] diradicals.

One of the developments carried out during this work has been a further step in this direction. First, the SOMOs obtained by means of Hückel MO theory (ψ_1 and ψ_2) are expanded as the usual linear combination of atomic orbitals:

$$\begin{aligned}\psi_1(\mathbf{r}) &= \sum_i c_{1i} \chi_i(\mathbf{r}) \\ \psi_2(\mathbf{r}) &= \sum_i c_{2i} \chi_i(\mathbf{r})\end{aligned}\tag{4.10}$$

where $\psi_i(\mathbf{r})$ correspond to the i -th SOMO, $\chi_k(\mathbf{r})$ correspond to the $2p_z$ atomic orbital basis and $\{c_{ik}\}$ is the set of expansion coefficients (see the discussion around equation 4.5). Now, as the overlap density is defined as $\rho(\mathbf{r}) = \psi_1(\mathbf{r})\psi_2(\mathbf{r})$, it can be re-expressed as:

$$\rho(\mathbf{r}) = \sum_i c_{1i}c_{2i}[\chi_i(\mathbf{r})]^2 + 2 \sum_i \sum_{j>i} c_{1i}c_{2j}\chi_i(\mathbf{r})\chi_j(\mathbf{r})\tag{4.11}$$

Given the localized nature of the $\chi_k(\mathbf{r})$ basis, the products $\chi_k(\mathbf{r})\chi_l(\mathbf{r}) \rightarrow 0$ as the distance between atoms increases or the basis functions become more localized. Assuming a high-localized limit, all the crossed terms vanish, prevailing only the products of basis located at the same center. Hence, equation 4.11 can be approximated as:

$$\rho(\mathbf{r}) \approx \sum_i c_{1i}c_{2i}[\chi_i(\mathbf{r})]^2 = \sum_i \rho_i(\mathbf{r}) \quad (4.12)$$

where only the first term of equation 4.11 prevails. Note that, while this approximation would not be valid for slowly-decaying and diffuse atomic orbitals with high principal or azimuthal quantum numbers, the present case study is constrained to the $2p_z$ orbitals of carbon atoms, thus, making this approximation reasonable in this context. In a similar manner, incorporating this approximation into the formal definition of K_{12} (equation 4.5) results in:

$$K_{12} = \iint \left[\sum_n \frac{\rho_n(\mathbf{r}_1)\rho_n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} + 2 \sum_n \sum_{m>n} \frac{\rho_n(\mathbf{r}_1)\rho_m(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \right] d\mathbf{r}_1 d\mathbf{r}_2 \quad (4.13)$$

which, again, can be further simplified as done in equation 4.12 for $\rho(\mathbf{r})$ as

$$K_{12} \approx \iint \sum_n \frac{\rho_n(\mathbf{r}_1)\rho_n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 = \sum_n \iint \frac{\rho_n(\mathbf{r}_1)\rho_n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (4.14)$$

Note that the former expression benefits from the commutation relation between integral and discrete summation operators. Now, recovering the explicit expression for $\rho_i(\mathbf{r}) = c_{1i}c_{2i}[\chi_i(\mathbf{r})]^2$ utilized in equation 4.11 it follows that:

$$K_{12} = \sum_n (c_{1i}c_{2i})^2 \iint \frac{[\chi_i(\mathbf{r}_1)]^2[\chi_i(\mathbf{r}_2)]^2}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (4.15)$$

Subsequently, noting that $\chi_i(\mathbf{r}) = \chi_j(\mathbf{r} + \mathbf{r}')$, that is, all the orbitals are equivalent except that they are centered at different positions, all the integrals result identical, i.e.

$$\tilde{k} = \iint \frac{[\chi_i(\mathbf{r}_1)]^2[\chi_i(\mathbf{r}_2)]^2}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 = \iint \frac{[\chi_j(\mathbf{r}_1)]^2[\chi_j(\mathbf{r}_2)]^2}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad \forall i, j \quad (4.16)$$

where \tilde{k} corresponds to the self-repulsion Coulomb integral, i.e. $\langle ii|ii \rangle$. Finally, the expression of K_{12} can be expressed in a compact notation as

$$K_{12} = \tilde{k} \sum_n (c_{1_i} c_{2_i})^2 \propto \langle \mathbf{c}_1 \circ \mathbf{c}_2 | \mathbf{c}_1 \circ \mathbf{c}_2 \rangle^2 = \mathcal{K}_{12} \quad (4.17)$$

where "o" corresponds to the Hadamard (i.e. point-wise) product.

To elaborate on the implications of the formulation collected in equation 4.17, consider how it reflects on the (non-)disjoint nature of molecular orbitals. Taking, for instance, a scenario with disjoint SOMOs, any coefficient that is nonzero in ψ_1 is necessarily zero in ψ_2 and vice versa by the very definition of disjointness. Consequently, the computation of the point-wise product of their vector coefficients, $\mathbf{c}_1 \circ \mathbf{c}_2$, results in a vector of zeros, thus leading to the conclusion that \mathcal{K}_{12} is zero and confirming the disjoint character of the SOMOs. Conversely, when considering non-disjoint SOMOs, some coefficients that are nonzero in ψ_1 are also nonzero in ψ_2 . In this case, the point-wise product $\mathbf{c}_1 \circ \mathbf{c}_2$ must also have non-zero elements in agreement with a significant overlap density. The magnitude of \mathcal{K}_{12} , thus, serves as rough indicator of the degree of non-disjointness between the SOMOs. This methodical approach does not only keep consistent with the Borden-Davidson deductions, but also provides a quantitative perspective on how structural modifications can impact the ΔE_{ST} gap.

For example, in evaluating the diradical character within the **My₂-DPA**[*n*] series, as depicted in Figure 4.8c, a discernible trend is observed with the computed \mathcal{K}_{12} values: i.e. an increment in the number of rings within the pentalene moiety coincides with a marked decrement in \mathcal{K}_{12} . Notably, the **My₂-DPA**[0] unit displays the most substantial ΔE_{ST} gap, suggesting a robust triplet character in comparison to the rest of compounds. As exemplified by these results, leveraging \mathcal{K}_{12} as an early-stage diagnostic tool has proven to be of great utility during the early stages of the work reviewed herein, offering preliminary insights into the molecular interactions that precede the more computationally taxing *Ab Initio* calculations.

4.2.3. Quantitative prediction of the ΔE_{ST} gap of **R₂-(DA)DPA**

As anticipated from the previous discussion, the evaluation of the ΔE_{ST} gap for **My₂-DPA**[0] and **My₂-DADPA**[0] reveals a triplet ground state with a large gap of ca. 5.7 kcal/mol and 8.5 kcal/mol, respectively (see blue symbol in Figures 4.9a and 4.9b). Now, considering a Boltzmann distribution with partition function

$$Z = 3 \cdot e^{\frac{-E_T}{RT}} + 1 \cdot e^{\frac{-E_S}{RT}} \quad (4.18)$$

the triplet state population at room temperature (i.e., $RT \approx 0.6$ kcal/mol) for **My₂-DPA**[0] reduces to

$$P_T = \frac{3 \cdot e^{\frac{-E_T}{RT}}}{3 \cdot e^{\frac{-E_T}{RT}} + 1 \cdot e^{\frac{-E_S}{RT}}} \cdot 100 = \frac{3}{3 + 1 \cdot e^{\frac{-\Delta E_{ST}}{RT}}} = \frac{3}{3 + 1 \cdot e^{-\frac{5.7}{0.6}}} \approx 100\% \quad (4.19)$$

indicating that these kind of pentalene-based diradicals could display excellent magnetic properties once ensembled in a material. Notably, the ΔE_{ST} gap for the protected diradical analogues **DPM₂-(DA)DPA**[0] and **PDM₂-(DA)DPA**[0],

are similar to those observed in their unprotected **My**-substituted analogues (see red and yellow symbols in Figure 4.9). This implies that integrating **DPM** and **PDM** units into the molecular structure does not markedly quench the magnetic interactions between the unpaired electrons and, thus, the sterically protected analogs still hold substantial FM interactions.

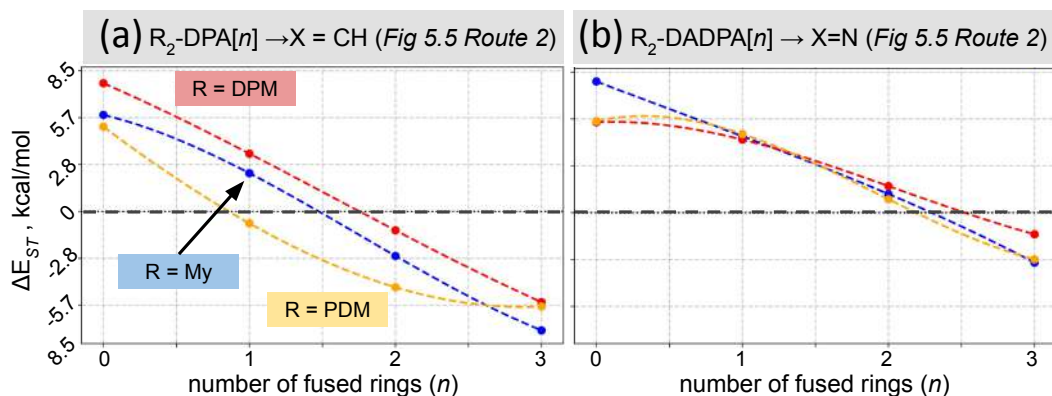


Figure 4.9: Singlet-triplet gap (ΔE_{ST}) as a function of the increasing number of fused 6-member rings (n from 0 to 3) for (a) $R_2\text{-DPA}[n]$ and (b) $R_2\text{-DADPA}[n]$, respectively. Colored data points represent different cores: $R = \text{My}$ (blue), $R = \text{DPM}$ (red), and $R = \text{PDM}$ (yellow). Dashed lines are included as guides to the eye and do not indicate continuity in any case.

In line with the predictions discussed in the previous subsection, the ΔE_{ST} gap drastically drops upon increasing the number of inserted 6-member rings. However, a crossover is observed in both series after $n=2$, reversing the energetic stability of the singlet and the triplet states. Notably, such a crossover is at odds with the prediction of models based solely on the topology of the SOMOs, as a pair of non-disjoint NBMOs is found also at DFT level of theory for all compounds of the $\text{My}_2\text{-DPA}[n]$ and $\text{My}_2\text{-DADPA}[n]$ series. Accordingly, the DFT results highlight the significant role of electron correlation in expanded

(diaz-)pentalene series, leading to apparent second-order effects that limit the applicability of models based solely on the topology of the SOMOs.

The insights extracted from DFT/PBE0 calculations have been validated by Complete Active Space (CAS) and Difference-Dedicated Configuration Interaction (DDCI) methods (outlined in **Chapter 2**). Accordingly, we have been able to determine the performance of the DFT approach and confirm the critical role of the second-order effects on the ΔE_{ST} gap. As a proof of principles, we have restricted our exploration to the $n=0$ and $n=2$ elements of both **My₂-DPA[n]** and **My₂-DADPA[n]** series, i.e., just after the ΔE_{ST} crossover. Interestingly, we found an excellent agreement between DFT/PBE0 and DDCI results (see first two rows of Table 4.1), which demonstrates that PBE0 is a suitable choice of DFT functional to adequately assess ΔE_{ST} gaps for this specific set of compounds.

In order to further analyze the origin of such second-order effects, we systematically expanded the active space from the simplest 2-orbitals and 2-electrons, CAS(2,2), up to 10-orbitals and 10-electrons, CAS(10,10) (see last three rows of Table 4.1). The good agreement between DDCI and CAS(2,2) demonstrates the small impact of electron correlation at $n=0$, as the latter only accounts for the correlation between the SOMOs of the diradical. In contrast, for $n=2$, the predicted ordering of singlet and triplet energy states is reversed when progressing from CAS(2,2) to DDCI. While DDCI predicts an antiferromagnetic interaction of the unpaired electrons (AFM, $\Delta E_{ST} < 0$), CAS(2,2) wrongly renders a ferromagnetic interaction (FM, $\Delta E_{ST} > 0$). Notably, the correct relative energy of the spin states is recovered for CAS(4,4) and CAS(10,10), where a higher degree of electron correlation is allowed by expanding the active space.

Table 4.1: Comparison of singlet-triplet energy gaps (ΔE_{ST} in kcal/mol) for My_2 -DPA[0], My_2 -DPA[2], and their diaza-derivatives (My_2 -DADPA[0] and My_2 -DADPA[2]) calculated using DFT/PBE0, DDCI(2,2) and CASSCF, using active spaces of increasing size (2,2) \rightarrow (10,10). The values for DFT/PBE0 and DDCI calculations are colored in green to emphasize their agreement. Values colored in red correspond to CAS(n,m) cases (being n and m the number of electrons and orbitals, respectively) where the sign of ΔE_{ST} changes with respect to DDCI(2,2).

Method	My_2 -DPA[0]	My_2 -DADPA[0]	My_2 -DPA[2]	My_2 -DADPA[2]
DFT/PBE0	+7.14	+8.01	-2.80	+1.14
DDCI(2,2)	+7.28	+7.52	-2.12	+2.00
CAS(2,2)	+7.56	+6.50	+0.70	+2.35
CAS(4,4)	+17.89	+16.86	-8.18	+5.91
CAS(10,10)	+9.72	+12.13	-4.21	+1.71

4.2.4. Assessment of the (anti)aromatic character of R_2 -(DA)DPA

As discussed earlier in the introduction section, we have assessed the stability of the R_2 -(DA)DPA[n] diradicals by means of three criteria: (i) response of the diradicals and bare antiaromatic couplers upon an external magnetic field by means of NICS calculations, (ii) structural analysis by means of BLA analysis and (iii) energetic criteria. In this regard, the presentation of the results and discussion is first devoted for the most elemental unit of the set, i.e. My_2 -(DA)DPA[0], under (i) and (ii) criteria and readily extended to the rest of the cases. After that, the trends regarding the energetic criteria, (iii), are presented for the whole series at once.

4.2.4.1. NICS and BLA results of My_2 -(DA)DPA[0]

As corroborated by theoretical studies⁸⁵ and NMR analysis of 1,3,5-tri-tert-butylpentalene,^{63,86} the instability of the isolated DPA[0] moiety is structurally expressed through a pseudo-Jahn-Teller distortion from a D_{2h} configuration, with uniformly distributed C–C distances,⁸⁷ to a C_{2h} symmetry. This latter configu-

ration is marked by a striking alternation of short and long C–C bonds, a pattern made noticeable by the BLA analysis depicted in Figure 4.10a. This alternation has been shown to result in a localized π system, which serves to mitigate the antiaromatic character of the compound.⁸⁸ Therefore, the presence or absence of an alternating carbon scaffolds may be indicative of the degree of antiaromaticity in the analogous **My**₂-**DPA[0]** diradical.

Interestingly, the alternating C–C bond pattern observed in **DPA[0]** experiences a significant shift with the incorporation of **My** open-shell units. As depicted in Figure 4.10a, the resultant **My**₂-**DPA[0]** diradical recovered a D_{2h} point group symmetry. Another particularly noteworthy characteristic of **My**₂-**DPA[0]** is the short C–C bond length of 1.364 Å between the **My** and **DPA[0]** units, marked in red in Figure 4.10a, alongside an apparent distortion within the pentalene moiety, as evidenced by the 1.387 Å distances colored in green. Overall, these observations suggest that the incorporation of the **My** unit together with the additional two unpaired electrons significantly impacts the electronic structure throughout the carbon scaffold, potentially altering its aromatic character.

In order to evaluate the(anti) aromatic character of **My**₂-**DPA[0]**, the chemical shielding has been examined by means of NICS, using the bare **DPA[0]** structure as a baseline comparison. Essentially, aromatic compounds are known to generate a diatropic electron current when subjected to an external magnetic field, which in turn causes the molecules to be shielded upon the external field.^{62,89} In contrast, antiaromatic compounds promote a reversed paratropic electron flow, resulting in a deshielding effect. The main insight from NICS is thus that this indirectly measured flow tropicity can be used as a proxy for gaug-

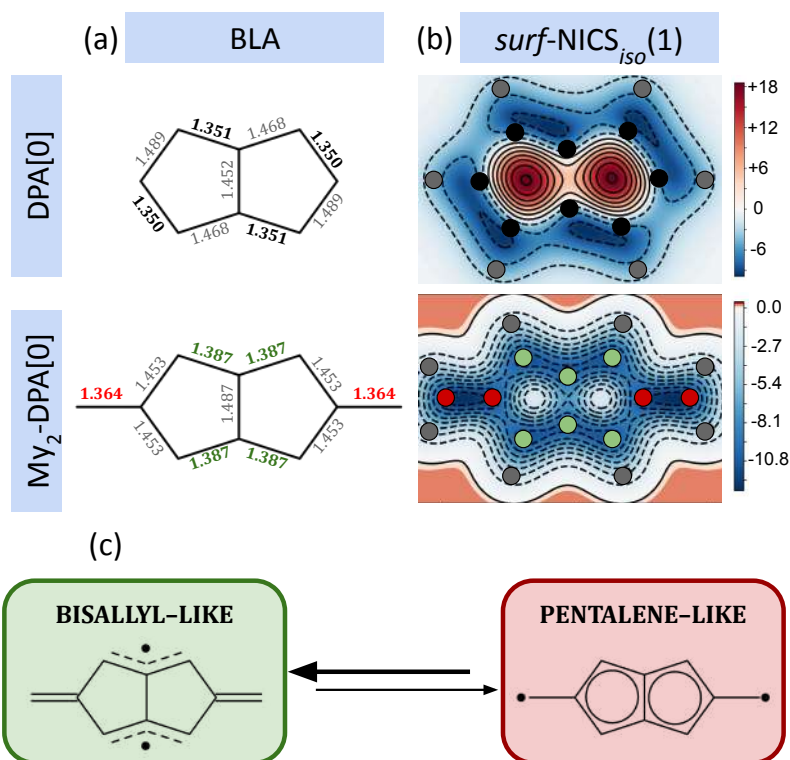


Figure 4.10: (a) BLA analysis for bare DPA[0] and My₂-DPA[0] diradical, with bolded distances indicating shorter values and gray distances representing larger values. (b) *surf*-NICS_{iso}(1), where the color gradient ranges from red to blue, indicating regions from high deshielding to high shielding in ppm, respectively. The position of all carbon atoms is colored in black for DPA[0], while carbon atoms in My₂-DPA[0] are colored according to BLA. The position of hydrogen atoms is colored in gray in both compounds. (c) Average resonance structures for My₂-DPA[0], with bisallyl-like structures (green) and pentalene-like structures (red), emphasizing the former being more likely. Note that the analyses for DPA[0] and My₂-DPA[0] have been carried out for their respective ground electronic states (closed-shell singlet and triplet, respectively).

ing the (anti)aromatic character. While the shielding or deshielding effects are most pronounced at the centers of carbon rings, our analysis was broadened to include the calculation of NICS_{iso}(1) over the entire molecular plane, a method referred here to as *surf*-NICS_{iso}(1). This method, previously utilized effectively

by Lampkin and co-workers,⁹⁰ provides insights not only of the (de)shielding attributes linked to the electron current's tropicity, but also hints at which dominant resonance structure takes place.

Turning the attention first to the (anti)aromaticity insights, the *surf*-NICS_{iso}(1) profile for the bare **DPA[0]** unit reveals a notably high NICS_{iso}(1) value of approximately 18 ppm at the center of both five-membered rings (see Figure 4.10b). This significant deshielding at these specific positions is indicative of the expected antiaromatic nature of the compound. In contrast, the *surf*-NICS_{iso}(1) map for **My₂-DPA[0]** shows a slight negative NICS_{iso}(1) value of 0.05 ppm at the ring centers. This noticeable disparity when compared to the **DPA[0]** suggests a considerable dampening in antiaromaticity for the diradical derivative relative to its antiaromatic precursor.

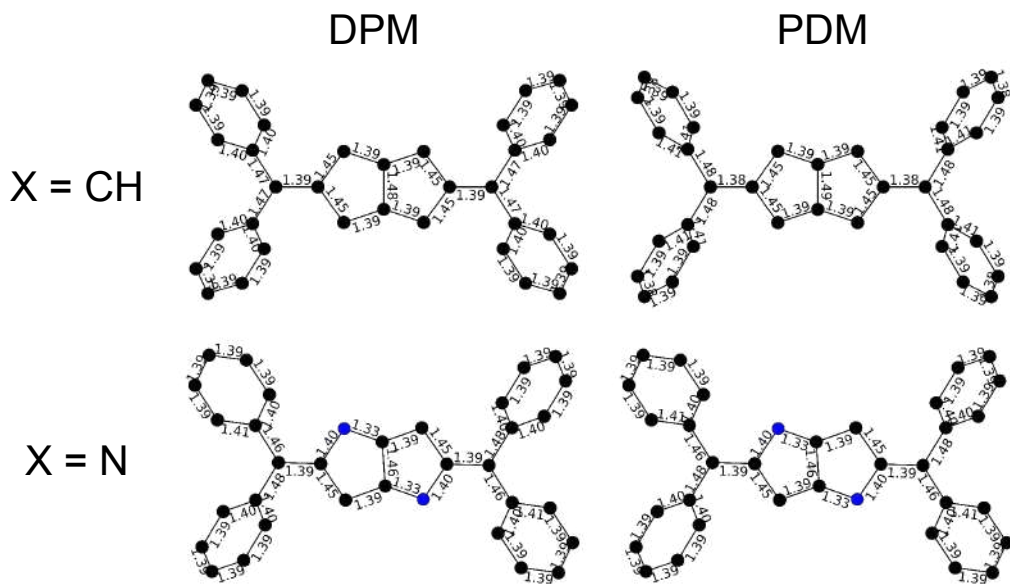


Figure 4.11: Carbon scaffold and distance of **DPA[0]** ($X=CH$) and **DADPA[0]**-based ($X=N$) radicals (in rows), with the **DPM** and **PDM** spin-bearing units (in columns).

Moreover, in line with the BLA results discussed above, *surf*-NICS_{iso}(1) also reveals alternant shielded regions along the carbon scaffold, indicating a set of localized π - π interactions and further evidence of the characteristic C_{2h} pseudo-Jahn-Teller distortion of the bare **DPA[0]**. In contrast, the *surf*-NICS_{iso}(1) map of **My₂-DPA[0]** depicts two local π -conjugated regions (see Figure 4.10b), each consisting of three carbon atoms, that resemble the structure of the bisallyl diradical mentioned in the introduction of this chapter. Moreover, a noticeable shielded region between **My** and **DPA[0]** carbon atoms is also apparent, further proving the strong interaction between the open-shell cores and the coupler suggested above. Thus, our findings unveil that the electronic structure of the **DPA[0]** in **My₂-DPA[0]** resembles more closely to that of a non-aromatic, bisallyl-like diradical rather than an antiaromatic, pentalene-like one (see Figure 4.10c, green and red, respectively). The results discussed thus far for **My**-substituted radicals can be straightforwardly extended to the **DPM** and **PDM**-substituted analogs by means of the BLA analyses. As illustrated in Figure 4.11, **DPM₂-DPA[0]** and **PDM₂-DPA[0]** diradicals also display the bisallyl-like carbon scaffold. In the case of **DPM₂-DADPA[0]** and **PDM₂-DADPA[0]** systems, however, the BLA analyses do not provide so much insight because the N atoms inherently alter the bond length distribution, blurring the conclusions that might be extracted from this type of analyses. For this reason, we will exclusively focus on the energy criterion in a following section.

Collectively, our findings illustrate a substantial attenuation of the antiaromatic character of **R₂-DPA[0]** in comparison to its **DPA[0]** precursor. Importantly, the NICS and BLA analyses reveal that **R₂-DPA[0]** exploits the synergistic stabilization energy introduced by Winter,⁵⁷ thus providing further

evidences that pentalene derivatives are not merely interesting for their magnetic properties but also for their severely dampended antiaromatic character.

4.2.4.2. *Extension of NICS and BLA results on R_2 -(DA)DPA[n] series*

The evaluation of the (anti)aromatic nature within the **My₂-DPA[n]** and **My₂-DADPA[n]** series was performed through NICS measurements at the center of each ring (refer to Figure 4.12). We distinguished between the centers of the five-membered rings (C_5 , marked in blue) and the two types of six-membered rings (central and side, i.e. C_6^M and C_6^S , indicated by orange and green, respectively). For this analysis, we concentrated on the ZZ component of the magnetic tensor, situated 1 Å above the molecular plane, and referred to as $NICS_{ZZ}(1)$. As elaborated in the methodology section (**Chapter 2**), the key contributions to the chemical shielding indicated by $NICS_{ZZ}(1)$ are derived from the π -system. Consequently, this metric is considered a more precise indicator of the (anti)aromatic character of planar molecular structures compared to the $NICS_{iso}(1)$ variant used previously. It is important to note that the *surf*- $NICS_{iso}(1)$ analysis conducted in the preceding section aimed to provide insights into additional aspects of the electronic structure, which $NICS_{ZZ}(1)$ alone may not fully encompass.

As illustrated in Figure 4.12a, the $NICS_{ZZ}(1)$ analysis for the **My₂-DPA[n]** series reveals that the relative NICS value between triplet (T) and open-shell singlet (OSS) states consistently expands upon increasing n in all ring types (compare the values of 2.3 and 0.3 ppm in **My₂-DPA[0]** to 1.2 and -2.3 of **My₂-DPA[1]** and so on). This trend suggests an enhanced aromatic character in the OSS electronic configuration relative to the triplet across the series. Notably, this observation is consistent with the chemical knowledge, as the electronic structure associated with the anti-parallel spin alignment can favor the electron

correlation by, for example, bond formation. Thus, when such bond formation does not result in an antiaromatic electronic structure, the $\text{NICS}_{\text{ZZ}}(1)$ lower in accordance to that, thus revealing more aromatic character in the OSS solution compared to the triplet.

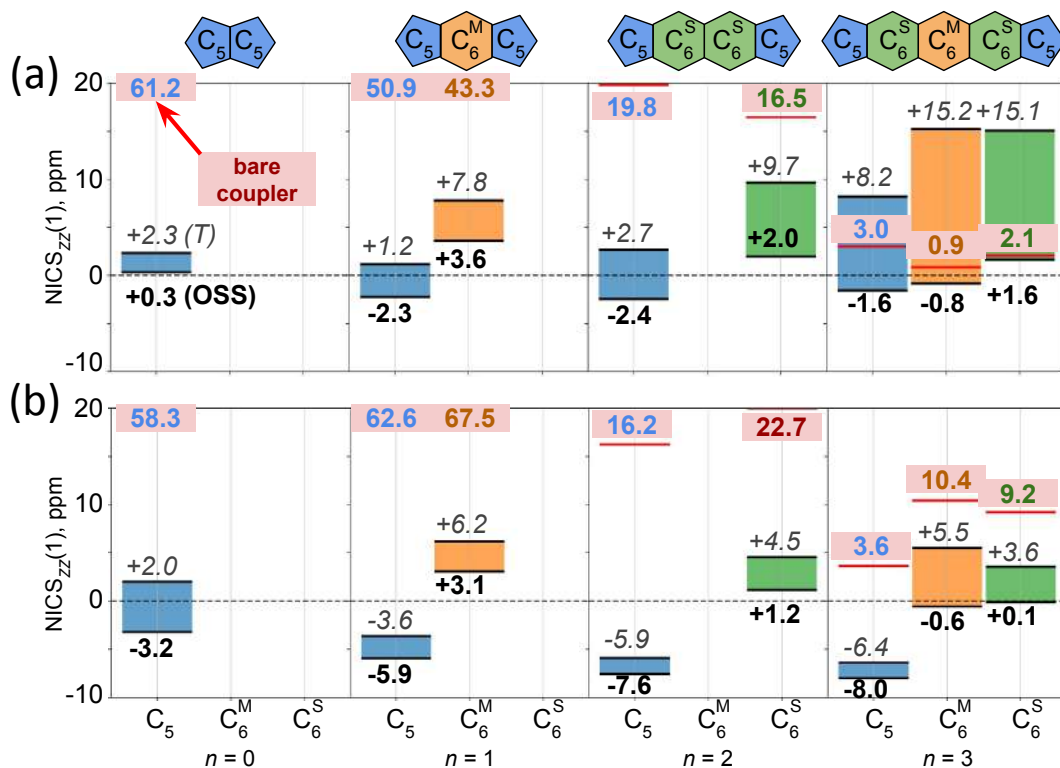


Figure 4.12: $\text{NICS}_{\text{ZZ}}(1)$ results for diradicals ($R=My$) and bare couplers from (a) DPA[n] and (b) DADPA[n] series across the number of fused [n]acenes, n . $\text{NICS}_{\text{ZZ}}(1)$ values for bare couplers are highlighted in red within each ring. For diradicals, $\text{NICS}_{\text{ZZ}}(1)$ values are distinguished between triplet (*italic*) and singlet (**bold**) states. A color-coded $\text{NICS}_{\text{ZZ}}(1)$ gap between these states is presented for clarity: blue, orange, and green correspond to rings C_5 , C_6^{M} and C_6^{S} , respectively.

Additionally, the trends along the series evidence a pronounced decline in $\text{NICS}_{\text{ZZ}}(1)$ values for the bare **DPA[n]** couplers, which is highlighted in dark red background color in Figure 4.12a. This trend suggests a gradual dilution of the

antiaromatic character with an increasing number of fused rings. Interestingly, the difference in $\text{NICS}_{ZZ}(1)$ between the bare coupler and the analogous diradical derivative notably decreases upon advancing through the series. Specifically, the **My₂-DPA[0]** (2.3 ppm) and **My₂-DPA[1]** (1.3 ppm) diradicals display a significant deviation in $\text{NICS}_{ZZ}(1)$ values when compared to their bare **DPA[0]** (61.2 ppm) and **DPA[1]** (50.9 ppm) precursors, respectively. In contrast, the $\text{NICS}_{ZZ}(1)$ for **My₂-DPA[2]** (2.7 ppm) aligns more closely with that of **DPA[2]** (19.8 ppm), and surpasses it in the **My₂-DPA[3]** case (compare 8.2 ppm and 3.0 ppm for the diradical and the precursor, respectively). In essence, the NICS data substantiate a discernible attenuation of the antiaromatic character in the diradicals with the extension of the series.

In the **My₂-DADPA[n]** series (see Figure 4.12b), we have observed a notable trend of decreasing $\text{NICS}_{ZZ}(1)$ values relative to their all-carbon analogs. For example, a shift from 0.3 ppm in the open-shell singlet state of **My₂-DPA[0]** to -3.2 ppm in **My₂-DADPA[0]** is observable in comparing the values reported in Figure 4.12a and Figure 4.12b. This decrease in NICS values is likely attributed to the presence of nitrogen atoms embedded within the ring structure, which potentially allows C=N imine interactions. Moreover, the addition of nitrogen atoms is also thought to enrich the electron density of the π -system, potentially shifting the electronic balance from a $4n$ to a $4n+2$ electron configuration. Furthermore, the difference in $\text{NICS}_{ZZ}(1)$ values between the bare precursors and their diradical counterparts is more substantial within the diaza series compared to the all-carbon series, hinting at the tendency for the **My₂-DADPA[n]** structures in favor of the diradical formation.

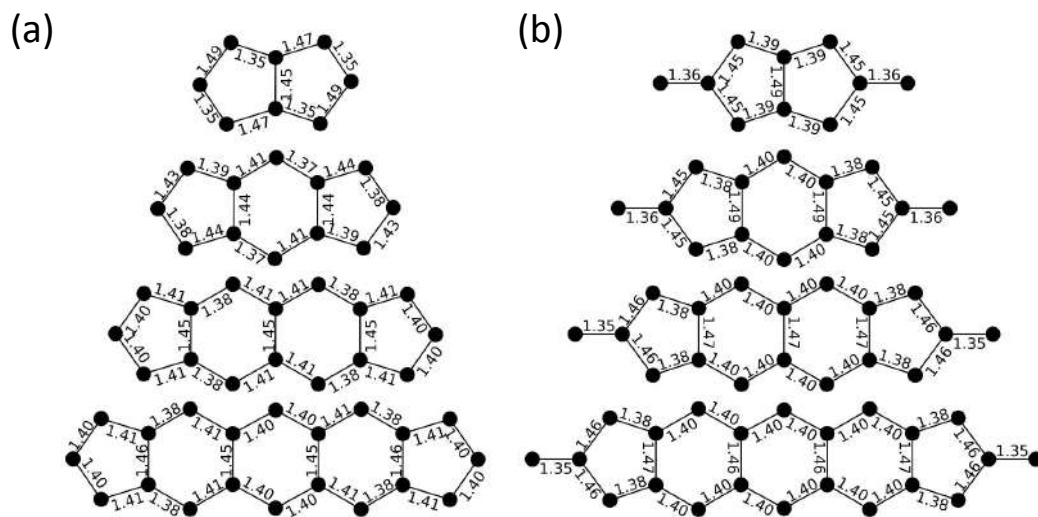


Figure 4.13: Carbon scaffold and distance of (a) $\text{DPA}[n]$ bare precursors and (b) *My*-substituted radicals.

Finally we observed that the $\text{NICS}_{ZZ}(1)$ results are in line with the BLA study carried out for the whole $\text{My}_2\text{-DPA}[n]$ set. We found that the characteristic alternate pseudo-Jahn-Teller distortion from D_{2h} to C_{2h} point group symmetry spotted for the bare $\text{DPA}[0]$ disappears when progressing through the series. Likewise, the bisallyl-like character of the corresponding $\text{My}_2\text{-DPA}[0]$ diradicals becomes less apparent as n increases (see Figure 4.13). In light of the comparison of our $\text{NICS}_{ZZ}(1)$ and BLA trends with the expected results, adding higher-order acenes within the (diaz)pentalene moiety progressively dilutes its antiaromatic character. In addition, we have also assessed that while the bare precursors grow less antiaromatic as n is increased, the affinity for forming diradicals decreases, i.e, there are less severe structural and NICS differences between them. Accordingly, for $\text{DPA}[2]$ and $\text{DPA}[3]$ a smaller driving force to form the diradicals can be expected compared to $\text{DPA}[0]$ and $\text{DPA}[1]$. Overall, $\text{R}_2\text{-DPA}[0]$ and $\text{R}_2\text{-DPA}[1]$ (and their diaza-analogs) stand out as the most

promising candidates of the series for the successful diradical formation based on the synergistic effect between the antiaromatic coupler and the spin-bearing units.

4.2.4.3. *Assessment of the thermodynamic stabilization energy*

Although the aromatic criteria explored in previous sections typically correlate with chemical stability, recent literature^{57,83} suggests the relationship between (anti)aromaticity and (in)stability for radicals may not be as clear-cut as previously thought. For this reason, the relative stability of the compounds of both **My₂-DPA[n]** and **My₂-DADPA[n]** series have been assessed by considering the radical stability energy (RSE). Importantly, RSE assessments are contingent upon a balanced chemical equation, which inherently reflects the intrinsic stability of the reference species used upon the same process experimented by the radical. Consequently, the potential RSE measurements are as varied as the reference species available. Although it is unlikely that a genuine RSE measure exist, the specific isodesmic reaction^{52,91} presented in Figure 4.14 is recognized for minimizing the influence of the reference species, thereby serving as a valuable practical approximation of intrinsic radical stability.^{83,92} Importantly, the reactants and products of the RSE, as defined herein, is such that the diradical appears on the products, which thus imply that an stability of the diradical with respect to its reduced (or hydrogenated) form lead to a negative sign of the RSE.

As illustrated in Figure 4.14, we have evaluated the energy balance of the RSE, E_{RSE} , along both **My₂-DPA[n]** and **My₂-DADPA[n]** series, shown in blue and red, respectively. We found that the diradicals are consistently stabilized with respect to the reference pair of **My** radicals in all the cases. Accordingly, this observation unveils that there is a significant energy gain in the resonance energy

due to the presence of the extra unpaired electrons (note the negative sign in all the cases). Remarkably, the trends observed through the RSE analysis reflect those previously identified via BLA and NICS measurements. That is, E_{RSE} for $\text{My}_2\text{-DPA}[0]$ and $\text{My}_2\text{-DADPA}[0]$ is significantly higher (by approximately 15 kcal/mol) than for the subsequent compounds in the series.

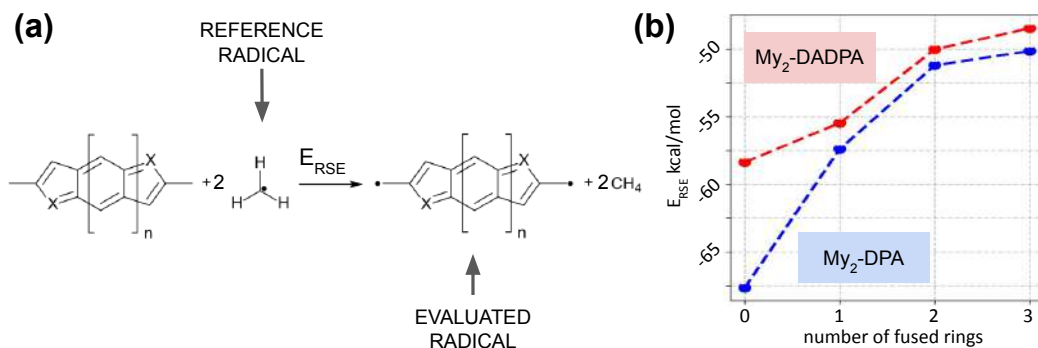


Figure 4.14: (a) Isodesmic reaction used to determine the radical stabilization energy (RSE) for the $\text{My}_2\text{-(DA)DPA}[n]$ series, utilizing the designated My reference radical for comparison. (b) Graphical representation of the calculated RSE values (in kcal/mol) for $\text{My}_2\text{-DPA}[n]$ (blue) and $\text{My}_2\text{-DADPA}[n]$ (red) series as a function of the number of fused rings (n).

In conclusion, the analysis conducted in this study underscores that the initial members of both $\text{My}_2\text{-DPA}[n]$ and $\text{My}_2\text{-DADPA}[n]$ series exhibit not only the most pronounced ferromagnetic interactions but also derive the greatest advantage from the stabilization mechanism based on antiaromatic couplers. The aromatic stabilization and the substantial energy gains reflected in the resonance energies position those as particularly compelling candidates for synthesis.

4.2.5. Properties of dibenzoannulated pentalene diradicals

In this subsection, the efforts are shifted to the **DBP** and **DADBP** derivatives, as detailed in route 3 of Figure 4.5. Although the exploration of **DPA**[n] and **DADPA**[n] couplers has been instrumental in understanding the proper-

ties of pentalene-based diradicals, it is experimentally apparent that **DBP** and **DADBP** derivatives exhibit substantially greater stability. Consequently, hypothetical synthetic approaches that involve the direct condensation of spin-bearing units onto the coupler are likely to be more feasible with **DBP** and **DADBP** derivatives than with **DPA**[*n*] and **DADPA**[*n*].

To investigate the magnetic and stability characteristics associated with **DBP** and **DADBP**, we conducted analyses parallel to those in the prior subsections. Initially, we employed the Longuet-Higgins and Borden-Davidson models to examine the number of non-bonding molecular orbitals (NBMOs) and the topology of the MOs. Despite the existence of 28 unique diradical structural isomers, our study selectively focused on the four depicted in Figure 4.15 to establish a proof of concept. Our findings suggest that the diradical isomers **2,7-R₂-DBP** and **4,9-R₂-DBP** are characterized by a pair of non-disjoint NBMOs, as depicted in Figure 4.15. On the other hand, the SOMOs in **1,6-R₂-DBP** and **3,8-R₂-DBP**, while still presenting a pair of NBMOs, display a disjoint topology. This particular arrangement renders them prone to exhibit negative ΔE_{ST} gaps.

Consistent with the approach in previous sections, our qualitative observations have been corroborated by quantitative PBE0/6-311G* calculations of the ΔE_{ST} gaps. As detailed in Table 4.2, both the **2,7-My₂-DBP** and **4,9-My₂-DBP** isomers demonstrate large ΔE_{ST} gaps of approximately 3.5 and 3.1 kcal/mol, respectively. Mirroring the **R₂-DPA**[*n*] series, these gaps significantly surpass the thermal energy at room temperature (approximated as $RT \sim 0.6$ kcal/mol), indicating their potential as robust FM building blocks. In alignment with these results, the diaza counterparts, namely **2,7-My₂-DADBP** and **4,9-My₂-DADBP**, manifest even greater ΔE_{ST} gaps in comparison to their **DBP**

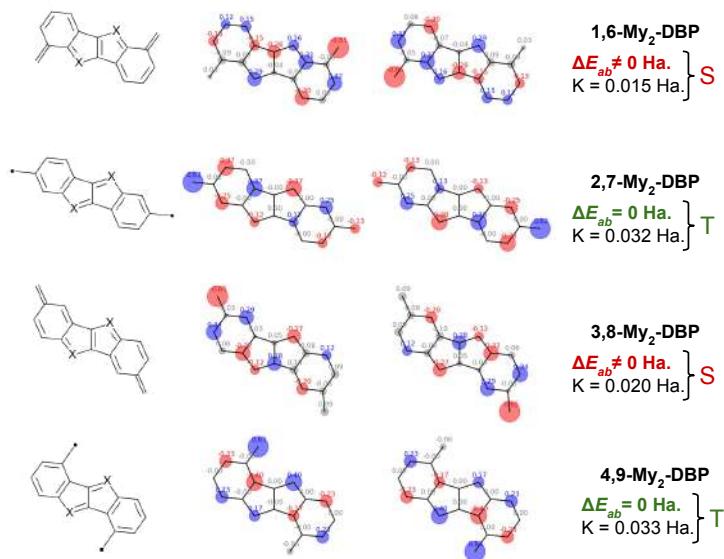


Figure 4.15: Lewis structures, SOMOs topology and ΔE_{ab} (gap between the SOMOs) for all the structural isomers of My_2 -DBP studied herein. While all the compounds are found to possess a pair of non-disjoint SOMOs, the only isomers displaying a gap of $\Delta E_{ab} = 0$ are 2,7- My_2 -DBP and 4,9- My_2 -DBP.

analogs. Additionally, the **1,6- My_2 -DBP** and **3,8- My_2 -DBP** isomers exhibit markedly smaller ΔE_{ST} gaps. Notably, **3,8- My_2 -DBP** presents a pronounced inversion of the singlet-triplet gap, a phenomenon observed in both the **DBP** and **DADBP** series.

Once more, the DFT results for **DBP** and **DADBP** are revealed to be fully in agreement with the qualitative insight derived from the topology-based models for this class of systems. Building on the established analysis, we have proceeded to quantify the ΔE_{ST} gap for isomers featuring a 2,7-connectivity with **DPM** and **PDM** open-shell cores, aiming to boost the kinetic stability by means of steric protection. Although both the examined diradicals have positive ΔE_{ST} values, it is the **PDM** derivatives that consistently exhibit a substantial ΔE_{ST}

gap (as reported in Table 4.2). This finding underscores **2,7-PDM₂-DBP** and **2,7-PDM₂-DADBP** as the most promising candidates of the whole set.

In line with the same analytical approach used for the **DPA[n]** series in preceding sections, both *surf*-NICS_{iso}(1) and BLA evaluations were conducted on the **My₂-DBP**-based derivatives (see Figure 4.16). The **2,7-My₂-DBP** isomer displays similar characteristics to those observed in the **My₂-DPA[n]** set. Specifically, the C–C bond connecting the **My** units to the **DBP** core is notably short, suggesting substantial contribution of the core in the π -conjugation (see Figure 4.16a). However, a difference compared to the **My₂-DPA[n]** series is notable, as BLA analysis shows that the pentalene core in the bare **DBP** is less prone to structural distortions after the introduction of open-shell groups. Complementing this observation, *surf*-NICS_{iso}(1) findings (illustrated in Figure 4.16b) reveal that the five-membered rings in **DBP** exhibit less negative NICS values than those in the isolated **DPA[0]**, which suggests a reduced antiaromatic character in **DBP**.

Table 4.2: ΔE_{ST} gaps (in kcal/mol) for *My₂-DBP* and *My₂-DADBP* structural isomers. Values correspond to structures depicted in Figure 6, including ΔE_{ST} gaps for sterically protected 2,7-DPM₂ and 2,7-PDM₂ analogs in the last two rows. Negative ΔE_{ST} gaps are highlighted in red. Asterisk denote the presence of a closed-shell singlet state (being the ground state as per DFT level of theory), otherwise, the open-shell singlet state is utilized.

Isomer	DBP	DADBP
1,6-My₂	0.09	2.37
2,7-My₂	3.46	5.06
3,8-My₂	-2.23	-14.85*
4,9-My₂	3.07	3.58
2,7-DPM₂	1.60	1.72
2,7-PDM₂	3.03	3.68

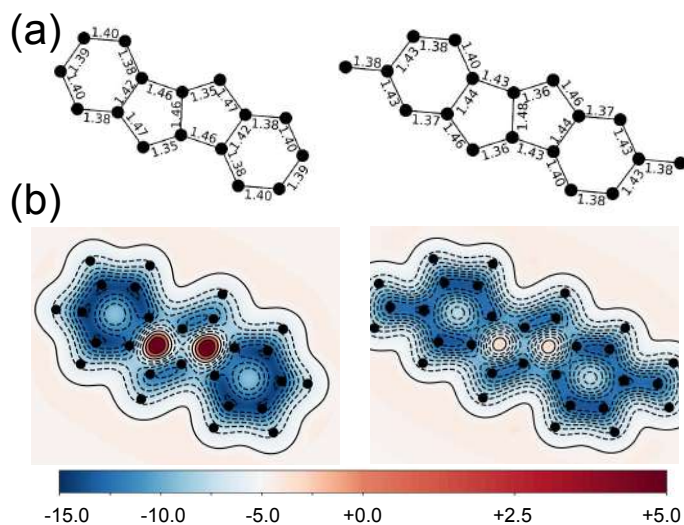


Figure 4.16: (a) Carbon scaffold and bond distances of DBP (left) and 2,7-My₂-DBP (right). (b) Surf-NICS_{iso}(1) contour plots for the bare DBP coupler and the 2,7-My₂-DBP diradical. The color gradient ranges from red to blue, indicating regions from high deshielding to high shielding (in ppm), respectively.

Overall, the set of results for **DBP** suggests there might be a less severe structural rearrangement required to alleviate antiaromatic character with the advent of diradical formation, unlike the bisallyl-like resonance structures that typically emerge in **R₂-DPA[n]**. Yet, the NICS_{iso}(1) values at the center of the 5-membered rings in the diradical are substantially lower than in bare **DBP** (see Figure 4.16b). Accordingly, the **2,7-R₂-DBP** diradical still emerges as a potentially stable and viable candidate.

4.3. Conclusions

In this work, we leverage (i) the fundamental quantum mechanical aspects, (ii) qualitative models that emerge out from them, and (iii) the potential ability of antiaromatic molecules to successfully design and characterize a set of pentalene-based diradicals displaying large and positive singlet-triplet gaps, i.e., ΔE_{ST} .

Overall, our investigation of pentalene-based diradicals have been clearly parceled into two primary categories: dicyclopenta- $[n]$ acene (**DPA** $[n]$ and **DADPA** $[n]$) and dibenzopentalene (**DBP** and **DADBP**) derivatives.

For the **DPA** $[n]$ and **DADPA** $[n]$ series, we have first identified the best connectivity to bolster FM interactions in the first element of the series, **DPA** $[0]$, by means of computationally-light and intuitive Longuet-Higgins and Borden-Davidson models. In this regard, the 2,5-connectivity has been recognized as the only adequate one to bolster FM interactions at the most fundamental level of approximation. Expanding upon the Borden-Davidson model, we have been able to derive an analytical expression to quantify the magnitude of the non-disjointness level. Subsequently, we have proved by means of DFT, CASSCF and DDCI that the **DPA** $[n]$ and **DADPA** $[n]$ compounds with $n < 2$ generally display FM interactions, featuring a proved 100% population of the triplet ground state at room temperature. Likewise, we identified that the role of second order effects in **DPA** $[n]$ and **DADPA** $[n]$ become increasingly relevant, revealing that diradicals of $n \geq 2$ are effectively prone to undergo singlet-triplet gap inversions.

Subsequently, we have utilized BLA, NICS and RSE criteria to prove and assess the mutually-beneficial cooperative effect between the spin-bearing units and antiaromatic moieties. Interestingly, our results uncover that those antiaromatic precursors possessing the most accentuated antiaromatic character are not only those displaying the largest gap, but also taking the most profit from the synergistic effect stemming from antiaromatic character of their precursors.

Our results from diradicals based on **DBP** and **DADBP** couplers further highlight the fundamental role of the connectivity between the open-shell cores

and the coupler in shaping the MO topology and, thus, the ΔE_{ST} gap. Core-coupler connectivity has been found to critically influence the topology of the SOMOs, setting the stage for tailoring the magnetic interactions in these types of systems. Overall, the large ΔE_{ST} gaps together with a dampened antiaromatic character provided a solid foundation for further exploration into thermally stable organic diradicals. The insights gained from this study not only contribute to the body of knowledge on organic electronics and spintronics materials but also suggest a more pragmatic role for pentalene beyond academic settings. Among all the diradicals explored across this work, **2,7-PDM₂-DBP** and **2,7-PDM₂-DADBP** derivatives, as well as diaza-substituted **PDM₂-DPA[1]** and **PDM₂-DADPA[1]** stand out as the most promising candidates for further study and applications.

References

- (1) C. Shu, Z. Yang and A. Rajca, *Chemical Reviews*, 2023, **123**, 11954–12003.
- (2) *Stable Radicals*, ed. R. G. Hicks, Wiley, 2010.
- (3) J. Linder and J. W. A. Robinson, *Nature Physics*, 2015, **11**, 307–315.
- (4) D. Cho and J. Y. Lee, *The Journal of Physical Chemistry C*, 2023, **127**, 8256–8262.
- (5) J. Devkota, R. Geng, R. C. Subedi and T. D. Nguyen, *Advanced Functional Materials*, 2016, **26**, 3881–3898.
- (6) A. Eftekhari, L. Li and Y. Yang, *Journal of Power Sources*, 2017, **347**, 86–107.
- (7) E. A. Weiss, M. J. Ahrens, L. E. Sinks, A. V. Gusev, M. A. Ratner and M. R. Wasielewski, *Journal of the American Chemical Society*, 2004, **126**, 5577–5584.
- (8) J. Ribas-Gispert, *Química de coordinación*, Omega, 2000.
- (9) J. Veciana, C. Rovira, M. I. Crespo, O. Armet, V. M. Domingo and F. Palacio, *Journal of the American Chemical Society*, 1991, **113**, 2552–2561.

- (10) K. Inoue and H. Iwamura, *Angewandte Chemie International Edition in English*, 1995, **34**, 927–928.
- (11) H. Quast, W. Nüdling, G. Klemm, A. Kirschfeld, P. Neuhaus, W. Sander, D. A. Hrovat and W. T. Borden, *The Journal of Organic Chemistry*, 2008, **73**, 4956–4961.
- (12) A. Rajca, K. Shiraishi and S. Rajca, *Chemical Communications*, 2009, 4372.
- (13) P. J. Boratyński, M. Pink, S. Rajca and A. Rajca, *Angewandte Chemie International Edition*, 2010, **49**, 5459–5462.
- (14) S. Suzuki, T. Furui, M. Kuratsu, M. Kozaki, D. Shiomi, K. Sato, T. Takui and K. Okada, *Journal of the American Chemical Society*, 2010, **132**, 15908–15910.
- (15) Y. Li, K.-W. Huang, Z. Sun, R. D. Webster, Z. Zeng, W. Zeng, C. Chi, K. Furukawa and J. Wu, *Chemical Science*, 2014, **5**, 1908.
- (16) K. Kato, K. Furukawa and A. Osuka, *Angewandte Chemie International Edition*, 2018, **57**, 9491–9494.
- (17) N. Gallagher, H. Zhang, T. Junghoefer, E. Giangrisostomi, R. Ovsyannikov, M. Pink, S. Rajca, M. B. Casu and A. Rajca, *Journal of the American Chemical Society*, 2019, **141**, 4764–4774.
- (18) S. Arikawa, A. Shimizu, D. Shiomi, K. Sato and R. Shintani, *Journal of the American Chemical Society*, 2021, **143**, 19599–19605.
- (19) A. Shimizu, T. Morikoshi, K. Sugisaki, D. Shiomi, K. Sato, T. Takui and R. Shintani, *Angewandte Chemie International Edition*, 2022, **61**.
- (20) P. Saha, N. Chrysochos, B. J. Elvers, S. Pätsch, S. I. Uddin, I. Krummenacher, M. Nandeshwar, A. Mishra, K. V. Raman, G. Rajaraman, G. Prabusankar, H. Braunschweig, P. Ravat, C. Schulzke and A. Jana, *Angewandte Chemie International Edition*, 2023, **62**.
- (21) Z. Zhu, D. Zhang, T. Xiao, Y.-H. Fang, X. Xiao, X.-G. Wang, S.-D. Jiang and D. Zhao, *Angewandte Chemie International Edition*, 2023.
- (22) G. Kothe, K.-H. Denkel and W. Sümmermann, *Angewandte Chemie International Edition in English*, 1970, **9**, 906–907.

- (23) D. R. McMasters and J. Wirz, *Journal of the American Chemical Society*, 2000, **123**, 238–246.
- (24) A. Rajca, K. Shiraishi, M. Vale, H. Han and S. Rajca, *Journal of the American Chemical Society*, 2005, **127**, 9014–9020.
- (25) N. M. Gallagher, J. J. Bauer, M. Pink, S. Rajca and A. Rajca, *Journal of the American Chemical Society*, 2016, **138**, 9377–9380.
- (26) W. Wang, C. Chen, C. Shu, S. Rajca, X. Wang and A. Rajca, *Journal of the American Chemical Society*, 2018, **140**, 7820–7826.
- (27) C. Shu, H. Zhang, A. Olankitwanit, S. Rajca and A. Rajca, *Journal of the American Chemical Society*, 2019, **141**, 17287–17294.
- (28) S. Tang, L. Zhang, H. Ruan, Y. Zhao and X. Wang, *Journal of the American Chemical Society*, 2020, **142**, 7340–7344.
- (29) K. Wang, P. Liu, F. Zhang, L. Xu, M. Zhou, A. Nakai, K. Kato, K. Furukawa, T. Tanaka, A. Osuka and J. Song, *Angewandte Chemie International Edition*, 2021, **60**, 7002–7006.
- (30) P. J. Hay, J. C. Thibeault and R. Hoffmann, *Journal of the American Chemical Society*, 1975, **97**, 4884–4899.
- (31) A. H. Goldberg and D. A. Dougherty, *Journal of the American Chemical Society*, 1983, **105**, 284–290.
- (32) D. A. Dougherty, *Accounts of Chemical Research*, 1991, **24**, 88–94.
- (33) J. J. Girerd, Y. Journaux and O. Kahn, *Chemical Physics Letters*, 1981, **82**, 534–538.
- (34) M. Wolfsberg and L. Helmholz, *The Journal of Chemical Physics*, 1952, **20**, 837–843.
- (35) H. Iwamura, in *Advances in Physical Organic Chemistry*, Elsevier, 1990, pp. 179–253.
- (36) H. C. Longuet-Higgins, *The Journal of Chemical Physics*, 1950, **18**, 265–274.
- (37) E. Huckel, *Zeitschrift für Physik*, 1933, **83**, 632–668.
- (38) C. A. Coulson and G. S. Rushbrooke, *Mathematical Proceedings of the Cambridge Philosophical Society*, 1940, **36**, 193–200.
- (39) P. Dowd, W. Chang, C. J. Partian and W. Zhang, *The Journal of Physical Chemistry*, 1993, **97**, 13408–13412.

- (40) A. A. Ovchinnikov, *Theoretica Chimica Acta*, 1978, **47**, 297–304.
- (41) E. Lieb, T. Schultz and D. Mattis, *Annals of Physics*, 1961, **16**, 407–466.
- (42) W. T. Borden and E. R. Davidson, *Journal of the American Chemical Society*, 1977, **99**, 4587–4594.
- (43) O. Kahn, *Molecular magnetism*, Courier Dover Publications, 2021.
- (44) Z. Chen, Y. Li and F. Huang, *Chem*, 2021, **7**, 288–332.
- (45) K. Kato and A. Osuka, *Angewandte Chemie International Edition*, 2019, **58**, 8978–8986.
- (46) T. Kubo, *The Chemical Record*, 2014, **15**, 218–232.
- (47) V. Diez-Cabanes, G. Seber, C. Franco, F. Bejarano, N. Crivillers, M. Mas-Torrent, J. Veciana, C. Rovira and J. Cornil, *ChemPhysChem*, 2018, **19**, 2572–2578.
- (48) D. Ruiz-Molina, J. Veciana, F. Palacio and C. Rovira, *The Journal of Organic Chemistry*, 1997, **62**, 9009–9017.
- (49) S. Rajca, A. Rajca, J. Wongsriratanakul, P. Butler and S.-m. Choi, *Journal of the American Chemical Society*, 2004, **126**, 6972–6986.
- (50) K. Sudoh, T. Hatakeyama, K. Furukawa, H. Nakano and Y. Matano, *Journal of Porphyrins and Phthalocyanines*, 2018, **22**, 542–551.
- (51) T. Kushida, S. Shirai, N. Ando, T. Okamoto, H. Ishii, H. Matsui, M. Yamagishi, T. Uemura, J. Tsurumi, S. Watanabe, J. Takeya and S. Yamaguchi, *Journal of the American Chemical Society*, 2017, **139**, 14336–14339.
- (52) D. Griller and K. U. Ingold, *Accounts of Chemical Research*, 1976, **9**, 13–19.
- (53) Y. Kim and E. Lee, *Chemistry – A European Journal*, 2018, **24**, 19110–19121.
- (54) A. Olankitwanit, S. Rajca and A. Rajca, *The Journal of Organic Chemistry*, 2015, **80**, 5035–5044.
- (55) A. Rajca and S. Utamapanya, *Journal of the American Chemical Society*, 1993, **115**, 10688–10694.
- (56) A. Rajca, K. Lu and S. Rajca, *Journal of the American Chemical Society*, 1997, **119**, 10335–10345.

- (57) R. Zhang, A. Ellern and A. H. Winter, *Angewandte Chemie International Edition*, 2021, **60**, 25082–25088.
- (58) J. S. Wössner, D. Wassy, A. Weber, M. Bovenkerk, M. Hermann, M. Schmidt and B. Esser, *Journal of the American Chemical Society*, 2021, **143**, 12244–12252.
- (59) J. Sprachmann, T. Wachsmuth, M. Bhosale, D. Burmeister, G. J. Smales, M. Schmidt, Z. Kochovski, N. Grabicki, R. Wessling, E. J. W. List-Kratochvil, B. Esser and O. Dumele, *Journal of the American Chemical Society*, 2023, **145**, 2840–2851.
- (60) R. R. Parkhurst and T. M. Swager, in *Topics in Current Chemistry*, Springer International Publishing, 2014, pp. 141–175.
- (61) T. Bally, S. Chai, M. Neuenschwander and Z. Zhu, *Journal of the American Chemical Society*, 1997, **119**, 1869–1875.
- (62) V. I. Minkin, M. N. Glukhovtsev and B. Y. Simkin, *Aromaticity and antiaromaticity*, John Wiley & Sons, Incorporated, 1994.
- (63) H. Hopf, *Angewandte Chemie International Edition*, 2013, **52**, 12224–12226.
- (64) T. Xu, Y. Han, Z. Shen, X. Hou, Q. Jiang, W. Zeng, P. W. Ng and C. Chi, *Journal of the American Chemical Society*, 2021, **143**, 20562–20568.
- (65) S.-J. Jhang, J. Pandidurai, C.-P. Chu, H. Miyoshi, Y. Takahara, M. Miki, H. Sotome, H. Miyasaka, S. Chatterjee, R. Ozawa, Y. Ie, I. Hisaki, C.-L. Tsai, Y.-J. Cheng and Y. Tobe, *Journal of the American Chemical Society*, 2023, **145**, 4716–4729.
- (66) M. Makino and J.-i. Aihara, *Phys. Chem. Chem. Phys.*, 2008, **10**, 591–599.
- (67) S. Motomura, M. Nakano, H. Fukui, K. Yoneda, T. Kubo, R. Carion and B. Champagne, *Physical Chemistry Chemical Physics*, 2011, **13**, 20575.
- (68) S. Tanaka, K. Satake, A. Kiyomine, T. Kumagai and T. Mukai, *Angewandte Chemie International Edition in English*, 1988, **27**, 1061–1062.
- (69) K. Hanida, J. Kim, N. Fukui, Y. Tsutsui, S. Seki, D. Kim and H. Shinokubo, *Angewandte Chemie International Edition*, 2021, **60**, 20765–20770.
- (70) J. Usuba and A. Fukazawa, *Chemistry – A European Journal*, 2021, **27**, 16127–16134.

- (71) M. Baranac-Stojanović and M. Stojanović, *Physical Chemistry Chemical Physics*, 2019, **21**, 3250–3263.
- (72) Z. U. Levi and T. D. Tilley, *Journal of the American Chemical Society*, 2009, **131**, 2796–2797.
- (73) T. Kawase, A. Konishi, Y. Hirao, K. Matsumoto, H. Kurata and T. Kubo, *Chemistry – A European Journal*, 2009, **15**, 2653–2661.
- (74) D. T. Chase, B. D. Rose, S. P. McClintock, L. N. Zakharov and M. M. Haley, *Angewandte Chemie International Edition*, 2010, **50**, 1127–1130.
- (75) A. Konishi, Y. Okada, M. Nakano, K. Sugisaki, K. Sato, T. Takui and M. Yasuda, *Journal of the American Chemical Society*, 2017, **139**, 15284–15287.
- (76) A. Minsky, A. Y. Meyer and M. Rabinovitz, *Tetrahedron*, 1985, **41**, 785–791.
- (77) G. I. Warren, L. J. Zocchi, L. N. Zakharov and M. M. Haley, *Chemistry – A European Journal*, 2023, **29**.
- (78) R. Casares, Á. Martínez-Pinel, S. Rodríguez-González, I. R. Márquez, L. Lezama, M. T. González, E. Leary, V. Blanco, J. G. Fallaque, C. Díaz, F. Martín, J. M. Cuerva and A. Millán, *Journal of Materials Chemistry C*, 2022, **10**, 11775–11782.
- (79) B. Esser, J. S. Wössner and M. Hermann, *Synlett*, 2022, **33**, 737–753.
- (80) D. Wassy, M. Pfeifer and B. Esser, *The Journal of Organic Chemistry*, 2019, **85**, 34–43.
- (81) J. S. Wössner, J. Kohn, D. Wassy, M. Hermann, S. Grimme and B. Esser, *Organic Letters*, 2022, **24**, 983–988.
- (82) X. Li and J. Yang, *Journal of the American Chemical Society*, 2018, **141**, 109–112.
- (83) L. Lin and J. Zhu, *The Journal of Organic Chemistry*, 2021, **86**, 15558–15567.
- (84) C. Foroutan-Nejad, *The Journal of Organic Chemistry*, 2023, **88**, 14831–14835.
- (85) I. G. Cuesta, S. Coriani, P. Lazzeretti and A. M. J. S. de Merás, *ChemPhysChem*, 2006, **7**, 240–244.
- (86) K. Hafner and H. U. Süss, *Angewandte Chemie International Edition in English*, 1973, **12**, 575–577.

- (87) A. Toyota and S. Koseki, *The Journal of Physical Chemistry*, 1996, **100**, 2100–2106.
- (88) S. M. Quintero, M. M. Haley, M. Kertesz and J. Casado, *Angewandte Chemie International Edition*, 2022, **61**.
- (89) G. Merino, T. Heine and G. Seifert, *Chemistry – A European Journal*, 2004, **10**, 4367–4371.
- (90) B. J. Lampkin, P. B. Karadakov and B. VanVeller, *Angewandte Chemie International Edition*, 2020, **59**, 19275–19281.
- (91) T. Ackermann, *Angewandte Chemie*, 1977, **89**, 921–921.
- (92) M. L. Coote and A. B. Dickerson, *Australian Journal of Chemistry*, 2008, **61**, 163.

Publication #3
Rational design of organic
diradicals with robust high-spin
ground state based on
antiaromatic building blocks

Rational Design of Organic Diradicals with Robust High-Spin Ground State Based on Antiaromatic Building Blocks

Raul Santiago,* M. Àngels Carvajal, Jordi Poater, Ibério de P. R. Moreira,
Stefan T. Bromley, Mercè Deumal, Jordi Ribas-Ariño*

* E-mail: raul.sant.1972@gmail.com & j.ribas@ub.edu

Abstract

Fully-organic molecules with high-spin ground states are promising building blocks for new lightweight flexible magnetic materials with potential for emerging applications (e.g. spintronics and quantum sensing). Here, we employ a modular design strategy to tune the topology of the frontier orbitals and explore the potential of combining open-shell cores with antiaromatic couplers in order to engineer a new class of high-spin diradicals. Our selection of couplers is anchored in pentalene-based moieties, with special focus on these derivatives displaying an attenuated antiaromatic character and, thus, a larger thermodynamic stability. Our results reveal that some of these diradicals stand out for their robust triplet ground states with substantial singlet-triplet energy gaps (e.g., $\Delta E_{ST} \sim 5$ kcal/mol). We have also gauged the thermodynamic stability of the resulting diradicals, revealing that these diradicals are tilted to a non-aromatic electronic structure, thus, potentially more stable than their antiaromatic bare precursors.

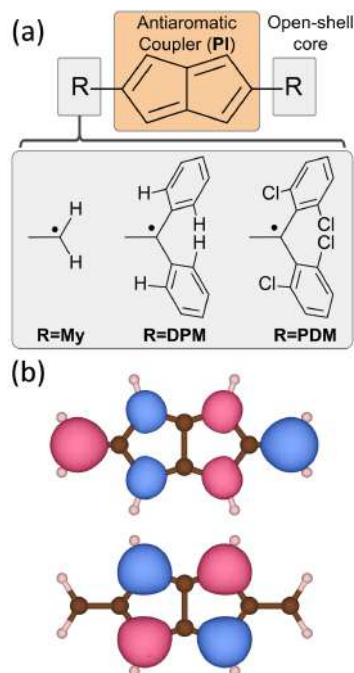
1. Introduction

Antiaromatic molecules are attracting increasing interest for applications in chemistry and materials science.^{Group5} This trend is primarily driven by both advances in synthetic techniques^{Hopf2013, Xhu2021, Jhang2023} and innovative designs devoted to enhancing the thermodynamic and kinetic stability of these molecules.^{REF} Thermodynamic stability can be achieved through the attenuation and dilution of their antiaromatic character (*e.g.*, by dibenzoannulation ^{Group6} or n-doped hetero-substitutions^{Tanaka1988,Hanida2021, Usuba2021}). Likewise, the inclusion of bulky substitutions has also been granted as an alternative strategy to increase the time and thermal resilience (*i.e.*, kinetic stability) of the unstable precursors ^{Hafner1973, Tanaka1988, Karas2023}. Furthermore, the unique electronic properties, specifically a low HOMO-LUMO gap, render antiaromatic molecules as promising candidates for organic electronics ^{Group2} and optoelectronic applications.^{Group3} The confluence of strategic design, accessible synthesis and the inherent electronic attributes of antiaromatic compounds have not only demonstrated their viability but also broadened their practical utilization. Such advances have led to the successful synthesis of a whole class of nanostructures^{Group1} and covalent organic frameworks,^{Sprachmann2023} establishing their potential for real-world applications.

In parallel to the synthetic efforts to prepare new antiaromatic molecules, open-shell organic molecules possessing a robust high-spin ground state is a long-standing research goal with recognised potential for spintronics,^{REF} memory devices,^{REF} spin filters,^{REF} and sensors.^{REF} The practical application of thermally stable high-spin organic magnets requires large energy gaps, ΔE_{ST} , between the singlet (S) excited state and the triplet (T) ground state.^{REF} Many organic compounds, including recently reported antiaromatic molecules,^{Group7} possess an open-shell singlet ground state with an antiferromagnetic (AFM) interaction between spins. Instead, examples of fully-organic diradicals displaying large ferromagnetic (FM) interactions between spins are still intriguingly uncommon, with only a few studies reporting compounds that meet these conditions.^{Group4} Bearing in mind the advancements in the synthesis of antiaromatic compounds, a clear opportunity emerges to leverage the potential of these moieties as building blocks for novel, robust high-spin units.

One of the most simple, yet illustrative, examples of the potential of antiaromatic molecules to bolster ferromagnetism in π -conjugated organic systems is 2,5-dimethylenylpentalene,^{Dias2003} **My₂PI**, which can be considered as a functionalized pentalene *coupler* with two added terminal methylenyl *cores* (**R=My** in Figure ^{F1}a) that serve as unpaired electron sources. **My₂PI** is a

non-alternant and non-Kekulé π -conjugated hydrocarbon that possesses a doubly-degenerate set of non-bonding molecular orbitals (NBMOs), as presented in Figure {F1}b. Under these conditions, Hund's rule promotes the parallel alignment of unpaired electrons, thus favoring a triplet ground state in a first approximation.{Iwamura1990} In addition, the specific carbon scaffold of **My₂PI** results in a non-disjoint set of Singly-Occupied Molecular Orbitals{Dias2003} (SOMOs), *i.e.*, the SOMOs cannot be localized on distinct groups of non-overlapping atoms by any linear combination between them.{Borden1977} Overall, the molecular orbital (MO) topology typified by **My₂PI** significantly impairs the energy stability associated with antiparallel spin alignment (favoring, *e.g.*, bond formation).{Iwamura1990} Consequently, the antiparallel spin alignment primarily results in unfavorable electron-electron Coulomb repulsion, favoring a triplet ground state and making **My₂PI** a notable example of a topological ferromagnetic molecule in a first approximation.



{F1} **Figure 1.** (a) Schematic overview of the modular diradical design featuring an antiaromatic pentalene coupler (PI, framed in orange) and two open-shell cores as substituents $R = \text{My}$, DPM , PDM (framed in gray). Some hydrogen atoms in DPM are explicitly set to highlight the differences with PDM . (b) SOMOs of My_2PI illustrated at a $0.04 \text{ e}/\text{\AA}^3$ isosurface with blue and red regions indicating positive and negative electron density probabilities, respectively.

In this study, we focus on the design of high-spin antiaromatic diradicals featuring non-bonding and non-disjoint singly-occupied molecular orbitals (SOMOs). Our approach is modular, involving the assembly of diradicals from separate building blocks, *i.e.*, two open-shell cores and an antiaromatic coupler (see Figure {F1}a). Specifically, we rely on couplers based on pentalene-like moieties that have been tailored for enhanced thermodynamic stability by means of ring insertion{Makino2008} and annulation.{BaranacStojanovi2019} These techniques have been proven effective to attenuate and dilute the antiaromatic character, thereby reducing the risk of redox and dimerization reactions.{Kawase2010, Hermann2018, Hopf2013} Moreover, we have also considered a set of diazapentalene-based couplers, a specific class of *n*-doped hetero-pentalene.{REF} Several studies have demonstrated the higher thermal resilience of diaza-derivatives compared to their full-carbon counterparts due to imine interactions (C=N).{REF} Regarding the open-shell cores, methylenyl (**R=My** in Figure {F1}a) provides the simplest example, but its anticipated inherent instability upon thermodynamically favored disproportionation or recombination processes makes it more suitable as a theoretical proof of concept rather than a realizable choice. Hence, for real-world applications, we grounded our selection inspired by persistent open-shell units at ambient conditions. In this regard, diphenylmethyl (**R=DPM** in Figure {F1}a) cores offer steric protection to the spin-bearing centers and enhance the kinetic stability of diradicals.{REF} In order to further enhance the kinetic stability, we have also included polychloro-diphenylmethylenyl (**R=PDM** in Figure {F1}a) cores in this study, specifically di-(2,6-dichloro)-diphenylmethylenyl. The introduction of bulky halogen substituents in the ortho positions of the phenyl moieties in **PDM** can aid the thermal resilience of the diradicals.{REF}

Methods like Extended Huckel{Hoffmann1963} (EH) and those rooted in spectral graph theory{Mallion1978} (SGT, further elaborated in **SI Section S{1}**) provide preliminary insights into the MO topology for these planar π -conjugated hydrocarbons. Here, we use these methods to design new high-spin antiaromatic diradicals with a specific MO topology akin to **My₂PI** at marginal computational cost. However, EH and SGT simple models lack the capability to fully capture higher-order effects like electron correlation of non-bonding SOMOs with the rest of the π -MOs.{Borden1994} Therefore, we have employed advanced *ab initio* methods for a rigorous evaluation of ΔE_{ST} gaps and magnetic properties. This is particularly crucial, as the performance of existing methods, like Density Functional Theory (DFT) and Correlated Wave-Function (CWF), is yet to be assessed in the context of this set of antiaromatic-based diradicals.

In the strategic design of these diradicals, we have also considered the stability of the compounds compared to their bare antiaromatic precursors. However, stability is a multifaceted

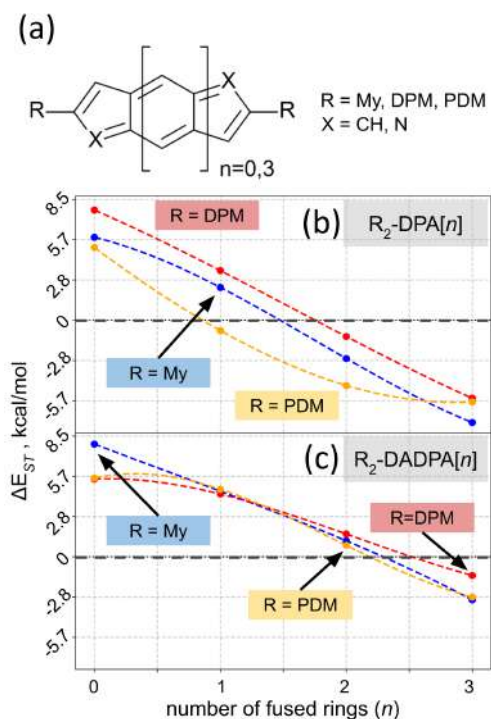
concept that lacks a universally accepted definition or measurement methodology. When dealing with fully-conjugated polycyclic hydrocarbons, the (anti)aromatic character can provide a comparative indicator of stability across different chemical systems. Accordingly, we opted for Bond Length Alternation (BLA), Nuclear Independent Chemical Shifts{Schleyer1996} (NICS), and Anisotropy of the Induced Current Density{Geuenich2005} (ACID) as indirect measures of aromaticity. These measures have provided a robust framework to gauge the thermodynamic stability of the studied pentalene-based diradicals and couplers from their (anti)aromatic character.

Overall, our findings reveal a new class of promising pentalene-based diradicals with robust triplet ground states and large ΔE_{ST} gaps, together with a few examples exhibiting singlet-triplet gap inversion. Although pentalene has been historically studied in academic settings to depict the properties stemming from antiaromatic units, its potential for practical applications has been argumentatively overlooked. While this trend is justified, our work underscores that the inherent thermodynamic instability of pentalene, and some of its derivatives, is markedly mitigated when their diradical derivatives are considered. This insight could contribute to pivoting the role of pentalene-like systems beyond mere academic curiosities to feasible candidates for emerging applications.

2. Results and Discussion

Our results and discussion is organized into five subsections. In the first subsection, we focus on the evaluation of ΔE_{ST} gaps for diradicals derived from pentalene and diaza-pentalene derivatives. We have also assessed the effect on ΔE_{ST} of the insertion of poly-acenes within the pentalene core (*i.e.*, expanded pentalenes) and the replacement of the methylenyl radical units by other protected open-shell cores. As mentioned in the introduction, we employ **My** as our theoretical benchmark, and **DPM** or **PDM** as potentially realizable units. In the second subsection, we gauge the thermodynamic stability of these diradicals in comparison to their bare antiaromatic precursors by means of BLA and aromaticity criteria (NICS and ACID). In the third subsection, we use standard functional groups to further kinetically stabilize the most promising couplers introduced this far. In the fourth subsection, we discuss similar ΔE_{ST} , BLA and NICS results for diradicals based on dibenzopentalene couplers, which is used as an example of an alternative thermodynamic stabilization strategy for pentalene and diazapentalene couplers via external dibenzoannulation. To conclude, we finally report the ΔE_{ST} gaps and viability of all newly suggested diradicals in the fifth subsection.

ΔE_{ST} gap in core-expanded and substituted pentalene diradicals. The set of antiaromatic diradicals explored in this section consists of two open-shell **R** cores (**R=My, DPM, PDM**) linked by either a dicyclopenta- $[n]$ acene coupler, **DPA** $[n]$, or a diaza-dicyclopenta- $[n]$ acene coupler, **DADPA** $[n]$ ($X=CH$ and $X=N$ in Figure {F2}a, respectively). The bare **DPA** $[n]$ and **DADPA** $[n]$ series result from the insertion of an $[n]$ acene moiety within a pentalene or diazapentalene unit, which is a recognized strategy to enhance the thermodynamic stability of such antiaromatic compounds by gradually attenuating their antiaromaticity as n increases.{Makino2008, Motomura2011} We have corroborated these trends for both the bare precursors and the diradical analogues by means of formation and combustion energy calculations (see details in **Section S{X} of the SI**).



{F2} Figure 2. (a) Chemical structures of diradicals based on dicyclopenta- $[n]$ acenes (R_2 -DPA $[n]$, $X=CH$) and diaza-dicyclopenta- $[n]$ acenes (R_2 -DADPA $[n]$, $X=N$). Plots (b) and (c) show the singlet-triplet gap (ΔE_{ST}) as a function of the increasing number of fused 6-member rings (n from 0 to 3) for R_2 -DPA $[n]$ and R_2 -DADPA $[n]$, respectively. Colored data points represent different cores: R=My (blue), R=DPM (red), and R=PDM (yellow). Dashed lines are included as guides to the eye and do not indicate continuity in any case.

The evaluation of ΔE_{ST} gap for **My₂-DPA[n]** diradicals at DFT level (PBE0/6-311G* exchange-correlation functional) reveals a triplet ground state with a large gap of *ca.* 5.7 kcal/mol for the first compound of the series ($n=0$, **My₂-DPA[0]** colored in blue in Figure {F2}b). Considering that thermal energy at room temperature (RT) is roughly 0.6 kcal/mol, the ΔE_{ST} gap effectively results in a 100% population of the triplet state at room temperature.{NOTE1} The nitrogen-doped counterpart, **My₂-DADPA[0]**, exhibits an even greater ΔE_{ST} gap of *ca.* 8.5 kcal/mol, as depicted in Figure {F2}c (colored in blue). Notably, the ΔE_{ST} gaps of **DPM₂-DPA[0]** and **PDM₂-DPA[0]** analogues, which are characterized by enhanced kinetic stability, also show comparable values, suggesting that the incorporation of **DPM** and **PDM** cores does not significantly quench the magnetic interactions of these diradicals. A potential explanation for the variations of ΔE_{ST} gaps for **DPM** and **PDM** derivatives can be ascribed to the dihedral twist of the phenyl rings with respect to the coupler plane, as explained in **Sections S{2} and S{3} of the SI**. It should be mentioned that **DPA[0]** has also been predicted to mediate strong magnetic couplings between metal centers in 2D Cr-based metal organic frameworks.{Li2018}

Interestingly, our DFT results show a reduction of the ΔE_{ST} gap upon increasing the number of inserted 6-member rings (see Figure {F2}b-c). Remarkably, a crossover is observed in both series after $n=2$, reversing the energetic stability of the singlet and the triplet states. Such a crossover is at odds with the predictions of models based solely on the electronic structure topology because the SOMOs of all diradicals of **My₂-DPA[n]** and **My₂-DADPA[n]** series are predicted to be non-disjoint NBMOs, thus, anticipating for all of them a triplet ground state at a first approximation (see **Section S{2} of the SI** for a visual representation of the SOMOs derived from SGT). Accordingly, our results highlight the significant role of electron correlation in expanded (diaz-)pentalene series and the limitations of the rational design based solely on the topology of the SOMOs.

The observed trends can be rationalized in two different ways. First, a close examination of the spin distribution of **R₂-DPA[0]** and **R₂-DADPA[0]**, provided by Mulliken's atomically-partitioned spin populations, reveals stark contrasts between the singlet and triplet states configurations (see **Section S{3} of the SI**). In the singlet state, the spin distribution resembles that of the two allylic subunits of the well-known tetramethylenethane diradical.{Borden1977,Pozun2013} In contrast, in the triplet state, the spin distribution is delocalized towards the terminal open-shell core units, arguably enhancing the triplet stability with respect to the singlet by reducing the electron Coulomb repulsion. This effect is also observed when progressing along the **R₂-DPA[n]** and **R₂-DADPA[n]** series. However, the gain from delocalizing the spin density becomes progressively smaller relative to the total number of atoms. As a result, the impact of spin density delocalization on the ΔE_{ST} gap is

expected to decrease with increasing molecular size. Alternatively, the trends can be understood on the basis of a qualitative analysis of the SOMOs of $\mathbf{R}_2\text{-DPA}[0]$ and $\mathbf{R}_2\text{-DADPA}[0]$. The shape of the SOMOs (see Figure {F1}a) suggests a high degree of overlap density, as the orbitals are confined to the same limited spatial region. This density overlap is generally associated with substantial positive ΔE_{ST} gaps and implies significant exchange integrals.{REF-Kahn} To further substantiate these qualitative observations, we calculated the exchange integrals across each series (see Section S{3} of the SI). We found that the exchange integral markedly decreases with increasing n in both series, which aligns with our finding that the SOMOs become increasingly diffuse with each increment in the series and, thus, the overlap density drops.

Method/Radical	$\text{My}_2\text{-DPA}[0]$	$\text{My}_2\text{-DADPA}[0]$	$\text{My}_2\text{-DPA}[2]$	$\text{My}_2\text{-DADPA}[2]$
DFT/PBE0	+7.14	+8.01	-2.80	+1.14
DDCI(2,2)	+7.28	+7.52	-2.12	+2.00
CAS(2,2)	+7.56	+6.50	+0.70	+2.35
CAS(4,4)	+17.89	+16.86	-8.18	+5.91
CAS(10,10)	+9.72	+12.13	-4.21	+1.71

{T1} **Table 1.** Comparison of singlet-triplet energy gaps (ΔE_{ST} in kcal/mol) for $\text{My}_2\text{-DPA}[0]$, $\text{My}_2\text{-DPA}[2]$, and their diaza-derivatives ($\text{My}_2\text{-DADPA}[0]$ and $\text{My}_2\text{-DADPA}[2]$) calculated using DFT/PBE0, DDCI(2,2) and CASSCF, using active spaces of increasing size (2,2) \rightarrow (10,10). The values for DFT/PBE0 and DDCI calculations are colored in green to emphasize their agreement. Values colored in red correspond to CAS(N,N) cases where the ΔE_{ST} reverses with respect to DDCI(2,2).

The DFT/PBE0 results in Figure {F2}b-c have been validated by other hybrid DFT functionals (reported in Section S{4} of the SI) and also employing Complete Active Space (CAS) and Difference-Dedicated Configuration Interaction (DDCI) methods. The good agreement found between DFT/PBE0 and DDCI results (see first two rows of Table {T1}) demonstrates that PBE0 is a suitable choice of DFT functional to adequately assess ΔE_{ST} gaps for these kinds of systems. Regarding the CASSCF results, we systematically expanded the active space from the simplest 2-orbitals and 2-electrons, CAS(2,2), up to 10-orbitals and 10-electrons, CAS(10,10), to examine the impact of electron correlation along the series. Accordingly, we calculated the ΔE_{ST} gap in both $\mathbf{R}_2\text{-DPA}[n]$ and $\mathbf{R}_2\text{-DADPA}[n]$ at $n=0$ and $n=2$, i.e., just after the ΔE_{ST} crossover of $\mathbf{R}_2\text{-DPA}[n]$. We carried out the calculations only for methylenyl ($\mathbf{R}=\text{My}$) core as a proof of principle. As observed in Table {T1}, DDCI and CAS(2,2) results predict a similar ΔE_{ST} value for both $\text{My}_2\text{-DPA}[0]$ and $\text{My}_2\text{-DADPA}[0]$. The good agreement between DDCI and CAS(2,2) demonstrates the small impact of electron correlation at $n=0$, as the latter only account for the correlation between the SOMOs of the diradical. In contrast, for $n=2$, the predicted ordering of singlet and triplet energy states is reversed

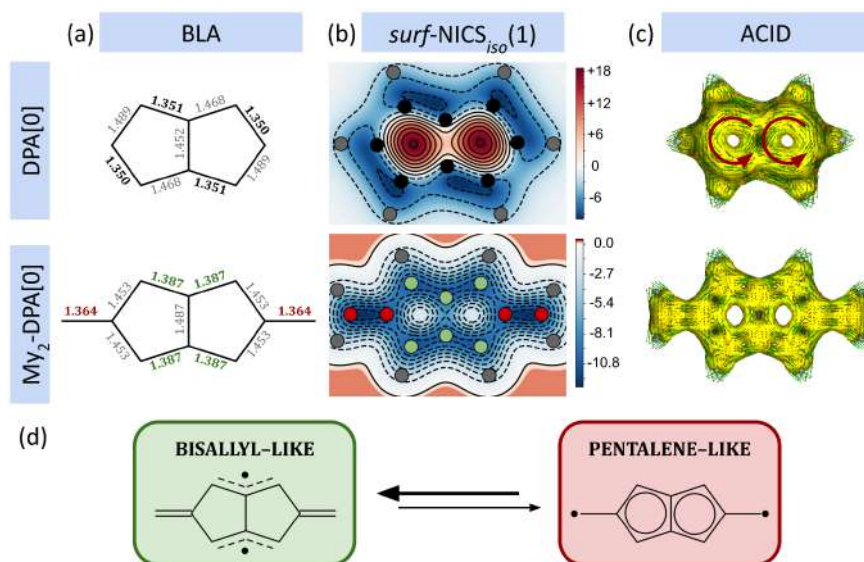
when going from CAS(2,2) to DDCI. While DDCI predicts an antiferromagnetic interaction of the unpaired electrons (AFM, $\Delta E_{\text{ST}} < 0$), CAS(2,2) wrongly renders a ferromagnetic interaction (FM, $\Delta E_{\text{ST}} > 0$). Notably, the correct relative energy of the spin states is recovered for CAS(4,4) and CAS(10,10), where a higher degree of electron correlation is allowed by expanding the active space. Overall our results highlight the impact of electron correlation on the ΔE_{ST} gap for these types of systems.

Thermodynamic stability analysis of expanded pentalene couplers. The thermodynamic stability of diradicals belonging to **R₂-DPA[n]** and **R₂-DADPA[n]** series has been assessed by means of Bond Length Alternation analysis (BLA), and aromaticity criteria based on the induced effects on the molecule as a response to an external magnetic field, such as NICS and ACID. In the first part of this subsection we focus on the simplest compound of the expanded pentalene series, *i.e.*, **My₂-DPA[0]**, to illustrate the overall behavior. Further evidence for the rest of the cases are provided later in the main text and in **Section S{5} of the SI**.

The bare **DPA[0]** moiety (*i.e.*, pentalene) can be identified as one of the most unstable couplers against dimerization reactions (among other processes) of the expanded pentalene series.^{Hafner1973} In line with previous theoretical work,^{GarcaCuesta2006} as well as NMR studies on 1,3,5-tri-*tert*-butylpentalene,^{Hopf2013} the isolated **DPA[0]** adopts a C_{2h} configuration with alternate short and long C-C bonds (see BLA Figure {F3}a). This configuration, which results from a pseudo-Jahn-Teller distortion of the D_{2h} configuration with an even distribution of C-C distances,^{Toyota1996} shows a localized π system that helps attenuate the antiaromatic character of the compound.^{MolesQuintero2022} The alternating C-C bond pattern of the **DPA[0]** moiety drastically changes when **My** units are connected to **DPA[0]** coupler and form the **My₂-DPA[0]** diradical. Specifically, **My₂-DPA[0]** features two local π -conjugated fragments arranged such that the system recovers a D_{2h} point group symmetry (see Figure {F3}a). Another significant structural feature of **My₂-DPA[0]** is a short C-C bond length of 1.364 Å between the **My** and **DPA[0]** blocks (see the bond lengths colored in red of Figure {F3}a). The BLA analysis reveals a severe distortion of the pentalene moiety, which indicates that the addition of these extra unpaired electrons can substantially influence the delocalization of electrons along the carbon scaffold (see 1.387 Å distances colored in green).

To assess the (anti)aromatic character of **My₂-DPA[0]**, we analyzed the chemical shielding at the center of the five-membered ring of the diradical, using the bare **DPA[0]** unit as a reference. In an external magnetic field, aromatic systems produce a diatropic electron current, resulting in molecular shielding.^{Minkin1994,Merino2004} Conversely, antiaromatic systems induce a

paratropic current, causing deshielding specially at the center of carbon rings. We have indirectly gauged current's tropicity and assessed the (anti)aromatic character of the system by means of NICS. Following the methodology of Lampkin *et al.*,^{Lampkin2020} we extended the calculation of $\text{NICS}_{\text{iso}}(1)$ ^{NOTE2} beyond the ring's centers, encompassing the entire molecular plane instead. This approach, named herein as *surf-NICS*_{iso}(1), not only offers an illustrative NICS map, but can also hint at average resonance structures of planar polycyclic hydrocarbons. Additionally, we have also calculated the scalar current field by means of ACID to provide a direct measure of the electron current tropicity as complementary evidence to our analyses.



{F3} Figure 3: (a) BLA analysis for bare DPA[0] and My₂-DPA[0] diradical, with bolded distances indicating shorter values and gray distances representing larger values. (b) *surf-NICS*_{iso}(1), where the color gradient ranges from red to blue, indicating regions from high deshielding to high shielding in ppm, respectively. The position of all carbon atoms is colored in black for DPA[0], while carbon atoms in My₂-DPA[0] are colored according to BLA. The position of hydrogen atoms is colored in gray in both compounds. (c) ACID results for DPA[0] and My₂-DPA[0] are shown for comparison. Red arrow indicates the overall current. (d) Average resonance structures for My₂-DPA[0], with bisallyl-like structures (green) and pentalene-like structures (red), emphasizing the former being more likely. Note that the analyses for DPA[0] and My₂-DPA[0] have been carried out for their respective ground electronic states (closed-shell singlet and triplet, respectively).

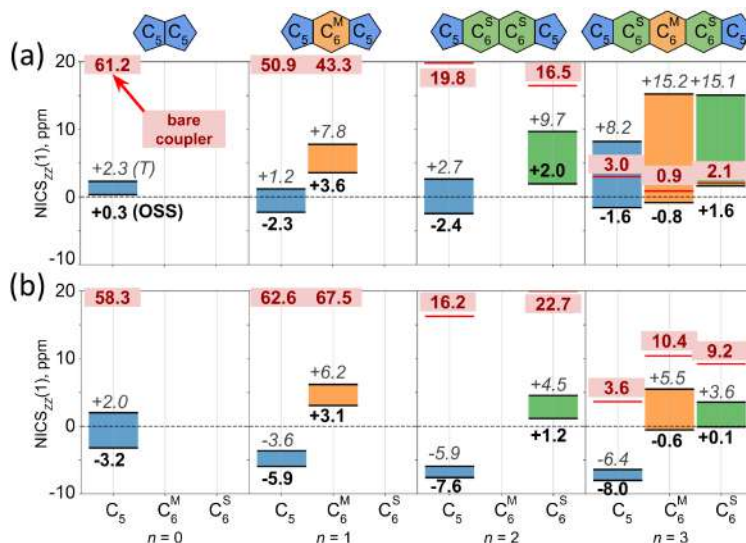
The *surf-NICS*_{iso}(1) map for the bare DPA[0] unit shows a significant positive NICS_{iso}(1) of ca. 18 ppm at the center of both 5-member rings (see Figure {F3}b), indicating a severe deshielding at

each center and, thus, a strong antiaromatic character. Moreover, the shared region between both 5-member rings features a lower deshielding compared to the ring centers, denoting a localized paratropicity where the direction of the currents in both rings counterbalance. ACID plots further support this interpretation, showing pronounced paratropic currents in both **DPA[0]** rings (see red arrows in Figure {F3}c). In line with the BLA results discussed above, *surf*-NICS_{iso}(1) also reveals alternant shielded regions along the carbon scaffold, revealing localized π - π interactions and further evidence of the characteristic C_{2h} pseudo-Jahn-Teller distortion of the bare **DPA[0]**. In contrast, *surf*-NICS_{iso}(1) reveals a residual negative NICS_{iso}(1) of 0.05 ppm at the ring centers of **My₂-DPA[0]**, which indicates a considerable dampening of the antiaromatic character compared to its precursor. The *surf*-NICS_{iso}(1) map depicts two local π -conjugated regions (see Figure {F3}b), each consisting of three carbon atoms, which resemble two connected allyl moieties (*i.e.*, $\cdot\text{CH}_2\text{-CH=CH}_2$, see atoms colored in green). There is also a significantly shielded region between **My** and **DPA[0]** carbon atoms, further proving the strong interaction between the open-shell cores and the coupler. The ACID results for **My₂-DPA[0]** show a weaker and disordered electron current compared to **DPA[0]** that supports the mitigated antiaromaticity of **My₂-DPA[0]** (refer to Figure {F3}c). This current also encompasses the carbon atom of the **My** core, revealing the strong participation of the **My** open-shell cores to the π -system. Overall, our findings suggest that the electronic structure of the **DPA[0]** in **My₂-DPA[0]** resembles that of a non-aromatic, bisallyl-like diradical rather than an antiaromatic, pentalene-like (see Figure {F3}d, green and red, respectively). Overall, our results unveil a significant dampening of the antiaromatic character of **My₂-DPA[0]** compared to **DPA[0]** and, consequently, the former can be expected to have an enhanced thermodynamic stability based on aromaticity criterias.

We have also performed similar BLA and ACID analyses for **DPM** and **PDM** cores, as detailed in **Section S{5} and S{X} of the SI**, respectively. The results for **DPM₂-DPA[0]** and **PDM₂-DPA[0]** align well with our initial conclusions for **My₂-DPA[0]**. Notably, both sterically protected derivatives display the characteristic bisallyl-like character and display pronounced C–C double-bonds between each core and the **DPA[0]** coupler. Such a robust core–coupler interaction provides further insights to explain why **DPM** and **PDM** units do not quench significantly the ΔE_{ST} gap compared to **My₂-DPA[0]** (as observed in Figure {F2}b). Namely, promoting a bisallyl-like electronic structure is preferred over diluting the spin density to the terminal phenyl rings of **DPM** and **PDM** and, thus, the core choice barely affects the topology of the SOMOs and the ΔE_{ST} gap.

In order to assess the thermodynamic stability of the diradicals compared to their bare precursors along **My₂-DPA[n]** and **My₂-DADPA[n]** series, we conducted a systematic NICS

evaluation (see Figure {F4}) at the center of each 5-member (C_5 , colored in blue) and both 6-member ring types (middle and side, *i.e.*, C_6^M and C_6^S , respectively colored in orange and green). Particularly, we here focused on the ZZ component of the magnetic tensor, located at 1 Å above the molecular plane, designated as $NICS_{zz}(1)$. In comparison to $NICS_{iso}(1)$ used for *surf*- $NICS_{iso}(1)$ maps, the major contributions to the chemical shielding at $NICS_{zz}(1)$ arise from the π -system{Chen2005} and, thus, it is a more appropriate indicator of the (anti)aromaticity of planar molecules.



{F4} **Figure 4.** $NICS_{zz}(1)$ results for diradicals ($R=My$) and bare couplers from (a) $DPA[n]$ and (b) $DADPA[n]$ series across the number of fused $[n]$ acenes, n . $NICS_{zz}(1)$ values for bare couplers are highlighted in red within each ring. For diradicals, $NICS_{zz}(1)$ values are distinguished between triplet (*italic*) and singlet (*bold*) states. A color-coded $NICS_{zz}(1)$ gap between these states is presented for clarity: blue, orange, and green correspond to rings C_5 , C_6^M , and C_6^S , respectively.

For the My_2 - $DPA[n]$ series, the $NICS_{zz}(1)$ gap between triplet (T) and open-shell singlet (OSS) states progressively widens (see Figure {F4}a). This indicates that the OSS electronic structure becomes increasingly more aromatic relative to the T solution across all its rings. In turn, the $NICS_{zz}(1)$ of the bare $DPA[n]$ coupler (emphasized in dark red at Figure {F4}a) decreases notably upon advancing through the series. That is, My_2 - $DPA[0]$ (2.3 ppm) and My_2 - $DPA[1]$ (1.2/7.8 ppm) diradicals exhibit larger $NICS_{zz}(1)$ values compared to their respective bare $DPA[0]$ (61.2 ppm) and $DPA[1]$ (50.9/43.3 ppm) precursors. Conversely, the $NICS_{zz}(1)$ for My_2 - $DPA[2]$ approaches those of $DPA[2]$ and even exceeds them in the case of My_2 - $DPA[3]$ (see color code for comparison between

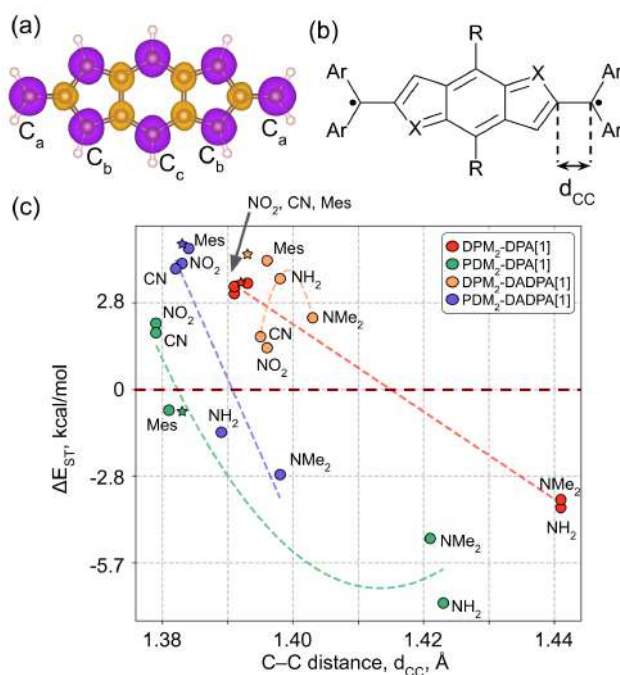
numerical values in Figure {F4}). Essentially, the antiaromatic dampening of diradicals compared to their bare precursors is clearly mitigated upon progressing through the series. The NICS_{zz}(1) results are in line with the BLA study carried out for the whole **My₂-DPA[n]** set (see **Section S{5} of the SI**). We found that the characteristic alternate pseudo-Jahn-Teller distortion from symmetric D_{2h} to C_{2h} point group spotted for the bare **DPA[0]** disappears when progressing through the series. However, the bisallyl-like character of the corresponding **My₂-DPA[0]** diradicals is preserved as n increases.

In the diaza-series, *i.e.* **My₂-DADPA[n]**, a systematic reduction in NICS_{zz}(1) is generally observed in comparison to their all-carbon analogs (for instance, see in Figure {F4}b how 0.3 ppm in **My₂-DPA[0]** at OSS becomes -3.2 ppm when **My₂-DADPA[0]** is considered). This reduction may be attributed to the incorporation of nitrogen atoms into the molecular scaffold, which potentially allows for C=N imine interactions. Moreover, the presence of nitrogen atoms could enhance the electron density in the π -system, shifting the $4n$ electronic configuration towards a $4n+2$ configuration. Notably, the difference in NICS_{zz}(1) values between the bare units and their diradical derivatives is more pronounced in the diaza series rather than in the all-carbon analogs. This could suggest an increased propensity of **My₂-DADPA[n]** series to form diradicals.

In light of our results, the NICS_{zz}(1) and BLA trends for the bare units align with other similar studies.{Stanger2019} That is, adding higher-order acenes within the (diaza)pentalene moiety progressively dilutes its antiaromatic character. However, an unanticipated inverse relationship appeared: while the bare precursors are anticipated to grow more thermodynamically stable, and thus less antiaromatic as n is increased, the affinity for forming diradicals decreases. Accordingly, for **DPA[2]** and **DPA[3]** a smaller driving force to form the diradicals can be expected. Overall, **R₂-DPA[0]** and **R₂-DPA[1]** (and their diaza-analogs) stand out as the most promising candidates of the series for the successful diradical formation based on aromaticity criteria.

Diradical kinetic stability enhancement through coupler functionalization. Our previous analyses indicate that diradicals with fewer fused acenes display larger ΔE_{ST} gaps and attenuated antiaromatic character compared to the bare precursors. Notably, the higher thermodynamic stability of bare **DPA[1]** and **DADPA[1]** over their zero-acene analogs{REF} might also ease synthetic viability, such as via core-coupler condensation reactions.{REF} However, **R₂-DPA[1]** and **R₂-DADPA[1]** diradicals possess unprotected sites with large spin density along the carbon scaffold (see highlighted C_{a,b,c} atoms in purple in Figure {F5}a), rendering them vulnerable to reduction reactions and potentially compromising the thermal resilience of their open-shell character. Since

the use of bulky **DPM** or **PDM** cores already offers steric protection to C_a and C_b sites, we focus on the functionalization of the remaining vulnerable C_c sites (see Figure {F5}b). Our selection of protective functional groups is based on a two-fold approach. Electron-withdrawing (EW) groups such as $-\text{NO}_2$ and $-\text{CN}$, as well as electron-donating (ED) groups like $-\text{NH}_2$ and $-\text{NMe}_2$, are selected to achieve a higher degree of spin density delocalization. In contrast, groups like mesityl ($-\text{Mes}$) are considered to provide steric hindrance against undesired reactions. {Konishi2017, Dressler2018}

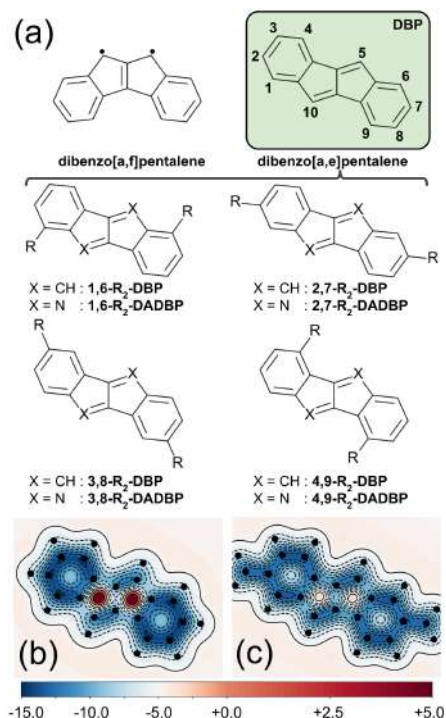


{F5} **Figure 5.** (a) Spin density of $\text{My}_2\text{-DPA}[1]$ at a $0.03 e/\text{\AA}^3$ isosurface, with purple and orange representing positive and negative density regions, respectively. (b) Scheme reflecting the coupler functionalization strategy applied to $\text{DPA}[1]$ and $\text{DADPA}[1]$ -based diradicals, where R represents various functional groups (NO_2 , CN , Mes , NH_2 , or NMe_2). (c) The distribution of ΔE_{ST} gaps (y -axis) is shown relative to the functionalization group (R) and the x -axis corresponds to the C-C bond length indicated in (b). Specific ΔE_{ST} gaps for $\text{DPM}_2\text{-DPA}[1]$, $\text{PDM}_2\text{-DPA}[1]$, and their diaza-substituted analogs ($\text{DPM}_2\text{-DADPA}[1]$ and $\text{PDM}_2\text{-DADPA}[2]$) are highlighted in red, green, orange, and purple, respectively, with a colored dashed line serving as a visual guide to the eye.

Figure {F5}c depicts the distribution of the ΔE_{ST} values with respect to the C-C distance of the core-coupler building blocks (labeled as d_{CC} in Figure {F5}b) for all functionalized $\text{R}_2\text{-DPA}[1]$ and $\text{R}_2\text{-DADPA}[1]$ diradicals, only using **DPM** and **PDM** cores. Interestingly, ED substitutions ($-\text{NH}_2$ and

-NMe₂) generally pull the ΔE_{ST} gaps in favor of AFM interactions in all cases but for **DPM₂-DADPA[1]** (compare with the unfunctionalized diradicals, marked by stars in Figure {F5}c). This underscores the role of ED groups in attenuating the antiaromaticity of the bare **DPA[1]** coupler by serving as alternative electron source to the π -system, thereby reducing the contribution of the core unpaired electrons and, thus, decreasing the ΔE_{ST} . In contrast, EW substitutions (**-NO₂** and **-CN**) retain the FM interactions of the diradical. Interestingly, the initially negative ΔE_{ST} gap in unfunctionalized **PDM-DPA[1]** becomes positive upon functionalization with either **-NO₂** and **-CN**. Clearly, EW groups inflict the opposite effect compared to ED functionalization in this case. Namely, EW groups pull out the electron density from the coupler, forcing a higher contribution of the open-shell core electrons to the π -system. Further evidence comes from variations in the core-coupler C-C bond lengths. In this vein, **R₂-DPA[1]** shows a shorter bond length when the diradical is substituted with EW functional groups. In contrast, the C-C bond length of the core-coupler bridge in ED-substituted diradicals displays a larger value. Alternatively, mesityl (**-Mes**) functionalization results in ΔE_{ST} gaps akin to those of unfunctionalized diradical analogs in all cases.

ΔE_{ST} gap in dibenzoannulated pentalene diradicals. In this subsection, the focus is shifted to derivatives of dibenzopentalene (see Figure {F6}a), which comprises a subset of pentalene-based antiaromatic compounds that have recently gained recognition for their enhanced thermodynamic stability and potential applications.{Hermann2021,Konishi2017} Similarly to the **DPA[n]** and **DADPA[n]** series, the stability of dibenzopentalenes arises from the attenuation of the antiaromatic character through the annulation of two benzene rings to the pentalene unit. Notably, dibenzo[a,e]pentalene and its isomer dibenzo[a,f]pentalene have been subject of intensive characterization in recent studies,{Hermann2021,Xu2012,Qiu2014,Wassy2019} revealing that dibenzo[a,f]pentalene intrinsically possesses an open-shell character prior to functionalization with open-shell cores.{Konishi2017} Thus, the discussion here is limited to dibenzo[a,e]pentalene and its diaza- derivative, 5,10-diaza-dibenzo[a,e]pentalene, referred hereafter as **DBP** and **DADBP**, respectively. Following the approach employed in previous subsections, our study has entailed the double functionalization of **DBP** and **DADBP** units with **My**, **DPM**, and **PDM** open-shell cores at diverse positions (collectively referred to as **R** in Figure {F6}a). Although 28 different isomers can be conceived, we have selected four representative cases for an in-depth characterization of their magnetic properties.



{F6} Figure 6. Structures and surf-NICS_{1s0}(1) contour plots of selected dibenzo[a,e]pentalene isomers. (a) Chemical structures display bare dibenzo[a,f]pentalene (left) and dibenzo[a,e]pentalene isomers (right, green). Below, some of the possible diradical derivatives of dibenzo[a,f]pentalene (DBP, X=CH) and diaza-dibenzo[a,f]pentalene (DADBP, X=N) are depicted. Surf-NICS_{1s0}(1) contour plots for (b) the bare DBP coupler and (c) 2,7-My₂-DBP diradical. The color gradient in (b) and (c) ranges from red to blue, indicating regions from high deshielding to high shielding (in ppm), respectively.

According to both SGT and EH methods, **2,7-R₂-DBP** and **4,9-R₂-DBP** diradical isomers are predicted to feature a pair of non-bonding, non-disjoint SOMOs (refer to **Section S{2} of the SI**). Hence, these isomers are deemed as diradicals with FM interactions ($\Delta E_{ST} > 0$). In contrast, the SOMOs of **1,6-R₂-DBP** and **3,8-R₂-DBP** isomers, although still non-disjoint, cannot be classified as NBMOs, which renders them prone to exhibit negative ΔE_{ST} gaps. Their inclusion in this study, however, have served as an evaluation of the consistency of the SGT and EH models compared to the DFT and CWF approaches. The PBE0/6-311G* ΔE_{ST} gaps for both **My₂-DBP** and **My₂-DADBP** are summarized in Table **{T2}** and validated with specific CAS and DDCI calculations (see **Section S{4} of the SI**), while Mulliken Population analyses are reported in **Section S{3} of the SI**. In agreement

with their molecular orbital topologies, both **2,7-My₂-DBP** and **4,9-My₂-DBP** isomers exhibit large ΔE_{ST} gaps of *ca.* 3.5 and 3.0 kcal/mol, respectively. As in **R₂-DPA[n]**, these gaps significantly exceed thermal energy at room temperature ($RT \approx 0.6$ kcal/mol), which place them as robust FM diradicals. Consistent with these findings, their diaza-derivatives, *i.e.*, **2,7-My₂-DADBP** and **4,9-My₂-DADBP**, display even larger ΔE_{ST} gaps compared to their **DBP** analogs. In contrast, the **1,6-My₂-DBP** and **3,8-My₂-DBP** isomers display considerably smaller ΔE_{ST} gaps, with **3,8-My₂-DBP** undergoing a severe singlet-triplet gap inversion for both **DBP** and **DADBP** couplers. Extending upon the previous framework, we have evaluated the ΔE_{ST} gap for the isomer having a 2,7-connectivity with **DPM** and **PDM** open-shell cores to further enhance the kinetic stability (thermal and time resilience) of the most promising diradical (see last two rows of Table {T2}). While all these diradicals display positive ΔE_{ST} values, our results reveal that **PDM** derivatives specially retain a robust ΔE_{ST} gap, thus positioning **2,7-PDM₂-DBP** and **2,7-PDM₂-DADBP** as the most promising realizable candidates among this set.

Isomer	DBP	DADBP
1,6-My₂	0.09	2.37
2,7-My₂	3.46	5.06
3,8-My₂	-2.23	-14.85*
4,9-My₂	3.07	3.58
2,7-DPM₂	1.60	1.72
2,7-PDM₂	3.03	3.68

{T2} **Table 2.** ΔE_{ST} gaps (in kcal/mol) for **My₂-DBP** and **My₂-DADBP** structural isomers. Values correspond to structures depicted in Figure 6, including ΔE_{ST} gaps for sterically protected **2,7-DPM₂** and **2,7-PDM₂** analogs in the last two rows. Negative ΔE_{ST} gaps are highlighted in red. Asterisk denote calculations using the closed-shell singlet state, otherwise, the open-shell singlet state is utilized.

In line with the methodologies applied to the **DPA[n]** series in previous subsections, *surf*-NICS_{iso}(1) and BLA analyses were performed on **My₂-DBP**-based compounds. For the **2,7-My₂-DBP** isomer, our findings parallel those from the **My₂-DPA[n]** series, *i.e.* the C-C bond length between the **My** cores and the **DBP** coupler is relatively short, indicative of a significant core participation in the π -system (see Section S{5} of the SI). Unlike the **My₂-DPA[n]** series, however, the BLA analysis reveals that the pentalene moiety in the bare **DBP** undergoes less severe structural deformations upon open-shell functionalization. Furthermore, *surf*-NICS_{iso}(1) analysis (see Figure {F6}b) indicates that the shielding values in the 5-member rings of **DBP** are far less negative compared to isolated **DPA[0]**, confirming a dampened antiaromatic character of the former (further

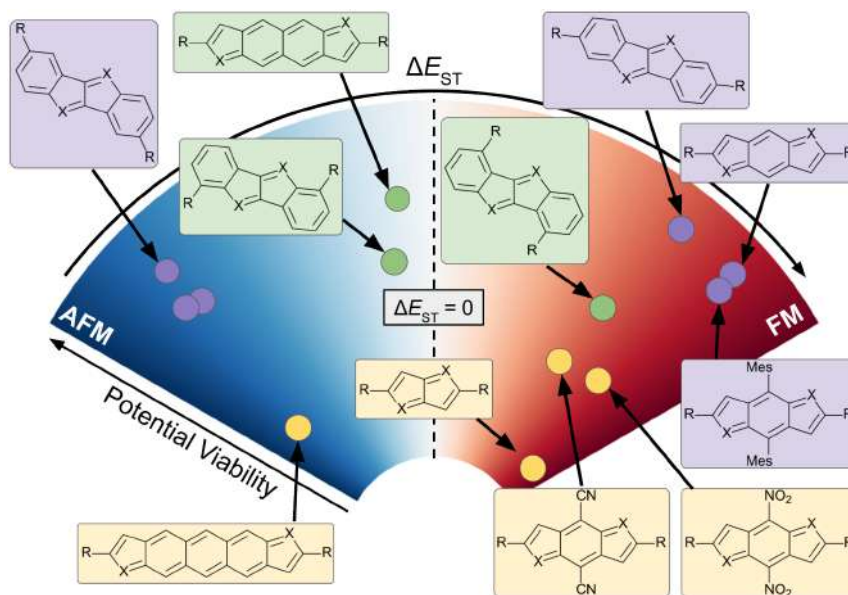
corroborated by NICS_{zz}(1) results reported in **SI Section S{6}**). Starting from an inherently less antiaromatic base structure such as **DBP** may reduce the driving force for extensive electronic structure deformations to mitigate antiaromaticity upon diradical formation, such as the emergence of the bisallyl-like average resonance form observed in **R₂-DPA[n]**. Yet, the NICS_{iso}(1) values at the center of the 5-member rings in the diradical are substantially lower than in bare **DBP** (see Figure {F6}c). Accordingly, the **2,7-R₂-DBP** diradical emerges as a potentially stable and viable candidate fulfilling both thermodynamic and kinetic criteria.

Promising pentalene-like derivatives as FM diradicals. In view of the large number of molecules considered in this study, a classification has been devised based on their potential viability and their magnetic ΔE_{ST} gap. The viability criteria was established based on three key factors: (i) whether or not the bare coupler or a closely related analogue has already been synthesized, (ii) the thermodynamic stability derived from the (anti)aromaticity of the diradical as well as that related to the bare coupler, and (iii) potential steric hindrance concerns between the bare coupler and the open-shell cores when the radical is formed, specifically when R = **DPM** and **PDM**. The results are graphically summarized in Figure {F7}, where a radial diagram is employed to categorize the molecules according to their viability and ΔE_{ST} . The angular component and color gradient are utilized to represent ΔE_{ST} values. The left region, colored in varying shades of blue, contains diradicals possessing AFM interactions between unpaired electrons. Conversely, the right region, coded in shades of red, encompasses diradicals with strong FM interactions. The radial axis reflects the potential viability of each molecule based on the aforementioned criteria. A discrete color scheme further differentiates specific cases according to their viability, ranging from unlikely (yellow) to possible (green) and feasible (purple).

Overall, our investigation of pentalene-based diradicals can be clearly parceled into two primary categories: dicyclopenta-[*n*]acene (**DPA[n]** and **DADPA[n]**) and dibenzo[a,e]pentalene (**DBP** and **DADBP**) derivatives. For the **DPA[n]** and **DADPA[n]** series, our findings reveal that compounds with $n \geq 2$ generally display AFM interactions, irrespective of their synthetic feasibility. The **DPA[0]** and **DADPA[0]** members of the series exhibit the largest ΔE_{ST} gaps, but the inherent antiaromatic instability of their precursors undermines their viability. The most favorable balance between a robust triplet ΔE_{ST} gap and viability is expected for **DPA[1]** and **DADPA[1]** derivatives, particularly those with diaza- substitutions. Those compounds are regarded to possess an enhanced thermodynamic stability due to the formation of imine (C=N) groups. {REF} Although these specific precursors have not been experimentally identified as stable, closely related derivatives have been reported in the literature. {REF} Notably, the functionalization of the existing derivatives is mainly

based on steric protection, which is unlikely to substantially perturb the topology of the SOMOs, thereby preserving their magnetic ΔE_{ST} gap (as we have shown above for mesityl substitutions). For **DBP** and **DADBP** derivatives, robust FM interactions are primarily observed in 2,7- and 4,9-substitutions, while 1,6- and 3,8-substitutions have been proven to be less effective. However, 4,9- configuration may face synthetic challenges due to the potential steric hindrance when functionalized by bulky **DPM** and **PDM** open-shell cores, thus rendering **2,7-DBP** and **2,7-DADBP** the most promising candidates devised for this set. As with the **DPA[1]** and **DADPA[1]** compounds, protected derivatives have been reported,^[REF] reinforcing the case for their viability. Finally, it is worth noting that our aromaticity analysis based on NICS and ACID methods firmly reveal that the diradical derivatives show an enhanced thermodynamic stability compared to the bare coupler, which further strengthens the potential viability.

In summary, we believe that the most promising candidates for achieving room-temperature robust FM diradicals across this study are the **2,7-PDM₂-DBP** and **2,7-PDM₂-DADBP** derivatives and diaza-substituted **PDM₂-DPA[1]** and **PDM₂-DADPA[1]**, due to their significant ΔE_{ST} gap and potential synthetic viability.



{F7} Figure 7. Viability and ΔE_{ST} gap summary of key pentalene-based diradicals explored across this work. The scheme qualitatively arranges structures based on potential viability (radial axis) and

ΔE_{ST} gap (angular axis). The background gradient transitions from blue (left) to red (blue), signifying the spectrum from most AFM to most FM interactions, respectively. The chemical structures are colored in yellow (unlikely), green (possible), or purple (probable) to indicate their viability.

3. Conclusions

This work offers a computational study of various π -conjugated organic diradicals, based on a modular approach to assemble these diradicals from two open-shell cores covalently linked to an antiaromatic coupler. We primarily focus on pentalene-like couplers displaying an attenuated antiaromatic character, specifically targeting the dicyclopenta- $[n]$ acene series (**DPA** $[n]$), dibenzo[a,e]pentalene (**DBP**), and their diaza-derivatives (**DADPA** $[n]$ and **DADBP**, respectively). Our findings reveal a set of viable compounds for practical applications with robust singlet-triplet ΔE_{ST} gaps, achieving a 100% triplet state population at room temperature.

We observe that diradicals presenting the most substantial singlet-triplet gaps, ΔE_{ST} , feature spatially restricted SOMOs, which are confined within the same region of the antiaromatic moiety. Accordingly, a decreasing ΔE_{ST} gap is found with the introduction of high-order $[n]$ acenes in radicals derived from **DPA** $[n]$ and **DADPA** $[n]$ couplers, resulting in a singlet-triplet energy crossover for compounds with $n > 2$.

Our results from diradicals based on **DBP** and **DADBP** couplers highlight the fundamental role of the connectivity between the open-shell cores and the coupler in shaping the MO topology and, thus, the ΔE_{ST} gap. Core-coupler connectivity is found to critically influence the topology of the SOMOs, setting the stage for tailoring the magnetic interactions in these types of systems. Accordingly, achieving a pair of non-bonding, non-disjoint SOMOs has been proved crucial to design robust FM building-blocks based on **DBP** and **DADBP** couplers.

Moreover, BLA, NICS and ACID results indicate a substantial dampening of the antiaromatic character of the diradicals compared to their bare precursors. Notably, we observe that the unpaired electrons of **R₂-DPA** $[n]$ diradicals considerably alter the electronic structure of the bare **DPA** $[n]$ units, shifting their associated π -conjugated systems from an antiaromatic pentalene-like to a non-aromatic bisallyl-like electronic structure.

Overall, the large ΔE_{ST} gaps together with a dampened antiaromatic character provide a solid foundation for further exploration into thermally stable organic diradicals. The insights gained from

this study not only contribute to the body of knowledge on organic electronics and spintronic materials but also suggest a more pragmatic role for pentalene beyond academic settings. Among all the diradicals explored across this work, our results stand **2,7-PDM₂-DBP** and **2,7-PDM₂-DADBP** derivatives, as well as diaza-substituted **PDM₂-DPA[1]** and **PDM₂-DADPA[1]**, as the most promising candidates for further study and application.

4. Computational Details

The molecular orbital structure of all open-shell compounds in this study was pre-screened using SGT methods, serving as a computationally-light approximation to evaluate the non-bonding and non-disjoint topology of frontier MOs. Regarding SGT method, eigenvalues and eigenvectors were derived from direct diagonalization of the adjacency matrix (A), represented solely by the carbon scaffold of the diradical. Elements of A , namely A_{ij} , were designated as either 0 (disconnected) or 1 (connected), based on a distance threshold criterion of 2.0 Å. This cutoff served as the appropriate value to render the atomic connectivity matrix according to the usual Lewis representation. Subsequently, the eigenstates of A were visually inspected by projecting the sign of the eigenvectors of interest onto the molecular structure (**Section S{1} of the SI**).

DFT calculations were performed by means of PBE0,{Ernzerhof1999} B3LYP{Becke1992, Lee1988} and ω -PBE{REF} exchange-correlation functionals as implemented in the Gaussian09 package,{gaussianRef} using an unrestricted formalism for all the open-shell calculations. A 6-311G* split-valence basis set with integrated polarization functions{Hariharan1973, Francl1982} has been employed, and MO symmetry not imposed during the calculation (NoSymm keyword). Only PBE0 ΔE_{ST} results are reported in the main text as a representative method that compares well with DDCI results. However, B3LYP and ω -PBE results are reported in **Section S{4.1} of the SI**. The ground state structures of diradicals were determined by a five-step protocol: (1) Geometry optimization in the triplet state, starting from a distorted diradical structure to avoid highly-symmetric local minima structures. (2) Geometry optimization of the Broken-Symmetry open-shell singlet state, guessing wave-function and atomic positions from the previous triplet (T) state. (3) Evaluation of the $\Delta_{ad}E_{ST} = (E_{S,opt} - E_{T,opt})$ gap to determine whether the triplet or singlet spin multiplicity state is the ground state (note that the subscript *ad* stands for adiabatic). Single point calculations at the opposite spin alignment were performed by taking the ground state structure and evaluating the direct singlet-triplet gap, ΔE_{ST} . For example, if $\Delta_{ad}E_{ST} > 0$ the triplet is the ground state and a new $\Delta E_{ST} = E_{T,opt}^S - E_{T,opt}^T$ would be evaluated. (4) A validation step was performed by repeating the last step with the structure of the excited spin state but with

opposite spin alignment. (5) Steps 2-4 were repeated for the closed-shell singlet (CSS) structure. Throughout the study, the CSS electronic structure was generally observed to be higher in energy compared to the open-shell singlet, exceptions being explicitly marked with an asterisk (*) along the text when reported. All DFT evaluations of ΔE_{ST} gaps were determined within the Broken-Symmetry (BS) approximation.^{BrokenSymm} The singlet-triplet gaps were calculated according to $\Delta E_{ST} = 2(E_S^{BS} - E_T)/(1+S_{ab})$,^{BrokenSymmEnergy} where S_{ab} represents the overlap between SOMOs and E_S^{BS} and E_T corresponds to the energies of the BS singlet and triplet states, respectively, calculated using the optimized geometry at the triplet state (referred to as $E_{T,opt}^S$ and $E_{T,opt}^T$ formerly in step 3). Due to significant overlap between SOMOs in all diradicals examined in this work, a value of $S_{ab} = 1$ has been assumed in all ΔE_{ST} calculations, simplifying the former expression to $\Delta E_{ST} = (E_S^{BS} - E_T) = \Delta E_{ST}^{BS}$. The accuracy of this approximation has been validated by means of DDCI calculations (see below).

The CASSCF and DDCI calculations were performed using the ORCA (version 4.2.1) package,^{orcaRef} a cc-pVTZ basis set^{REF} and the optimized structures at the PBE0 level. All DDCI calculations were carried out using the orbitals resulting from state-average (average between the lowest-energy singlet and lowest-energy triplet) CASSCF(10,10) calculations. The ΔE_{ST} gaps at the DDCI level (MR-DDCI3 keyword) reported in the article were obtained using the following values for the TPre and Tsel parameters: $1 \cdot 10^{-4}$ and $1 \cdot 10^{-10}$. As it may be seen in **Section S{4} of the SI**, these values ensure a converged and accurate value for the ΔE_{ST} gaps.

NICS calculations for diradicals were also performed using Gaussian09 package, utilizing the unrestricted PBE0 formalism with Gauge-Including Atomic Orbitals^{Cheeseman1996} (PBE0-GIAO/6-311G*) and an ultrafine grid for integral evaluations. These calculations were carried out at the ground state spin configuration of each diradical, or at the closed-shell singlet in the bare antiaromatic couplers. $NICS_{iso}(1)$ and $NICS_{zz}(1)$ were computed as the trace and ZZ component of the magnetic tensor, respectively, each with a reverse sign at the center of each ring, and positioned 1 Å above the molecular plane by means of inserting a “ghost” atom, X, at the target position. *surf*- $NICS_{iso}(1)$ mapping was performed similarly, employing a multiple evaluation of the magnetic tensor in a 150x150 grid of equidistant points in all the cases.

References

ID	Text Ref.	URL, DOI or Note
1	Borden1977	https://doi.org/10.1021/ja00456a010
2	Borden1994	https://doi.org/10.1021/ar00040a004
3	Hoffmann1963	https://doi.org/10.1063/1.1734456
4	Dias2003	https://doi.org/10.1021/ci034056z
5	Iwamura1990	https://doi.org/10.1016/S0065-3160(08)60046-5
6	Mallion1978	https://doi.org/10.1080/00268977800101451
7	BaranacStojanovi2019	https://doi.org/10.1039/c8cp07875k
8	Usuba2021	https://doi.org/10.1002/chem.202103122
9	Esser2022	https://doi.org/10.1055/a-1740-7139
10	Konishi2017	https://doi.org/10.1021/jacs.7b05709
11	Sprachmann2023	https://doi.org/10.1021/jacs.2c10501
12	Wassy2019	https://doi.org/10.1021/acs.joc.9b01195
13	Wssner2021	https://doi.org/10.1021/jacs.1c05251
14	Wssner2022	https://doi.org/10.1021/acs.orglett.1c03900
15	Hopf2013	https://doi.org/10.1002/anie.201307162
16	Kawase2010	https://doi.org/10.1002/anie.201003609
17	Wu2017	https://doi.org/10.1002/anie.201704668
18	Hermann2018	https://doi.org/10.1039/c8tc00970h
19	Tobe2014	https://doi.org/10.1002/tcr.201402077
20	Nishida2012	https://doi.org/10.1002/chem.201200591
21	Liu2015	https://doi.org/10.1002/chem.201502184
22	Dai2014	https://doi.org/10.1002/chem.201405652
23	Ren2016	https://doi.org/10.1039/c6tc01808d
24	Shimizu2022	https://doi.org/10.1002/ange.202205729
25	Rajca2011	https://doi.org/10.1021/ja200708b

ID	Text Ref.	URL, DOI or Note
26	Gallagher2019	https://doi.org/10.1021/jacs.9b00558
27	Makino2008	https://doi.org/10.1039/b714906a
28	Motomura2011	https://doi.org/10.1039/c1cp20773c
29	Chen2020	https://doi.org/10.1002/adv.201903766
30	Hong2022	https://doi.org/10.1002/ejoc.202101343
31	Xhu2021	https://doi.org/10.1021/jacs.1c06677
32	Jhang2023	https://doi.org/10.1021/jacs.2c13159
33	Hafner1973	https://doi.org/10.1002/anie.197305751
34	GarcaCuesta2006	https://doi.org/10.1002/cphc.200500327
35	Toyota1996	https://doi.org/10.1021/jp952459z
36	Lampkin2020	https://doi.org/10.1002/anie.202008362
37	Scipy	10.1038/s41592-019-0686-2
38	Levi2009	https://doi.org/10.1021/ja809930f
39	Kawase2009	https://doi.org/10.1002/chem.200802471
40	Hanida2021	https://doi.org/10.1002/anie.202109003
41	Tanaka1988	https://doi.org/10.1002/anie.198810611
42	Karas2023	https://doi.org/10.1002/anie.202307379
43	Chase2010	https://doi.org/10.1002/anie.201006312
44	Chase2012	https://doi.org/10.1021/ja303402p
45	Konishi2021	https://doi.org/10.1246/cl.200650
46	Yuan2016	https://doi.org/10.1021/acs.joc.6b01480
47	Dressler2018	https://doi.org/10.1038/s41557-018-0133-5
48	Li2018	https://doi.org/10.1021/jacs.8b11346
49	Horii2022	https://doi.org/10.1246/cl.210809
50	Zhang2017	https://doi.org/10.1039/c6qo00867d
51	ElBakouri2020	https://doi.org/10.1021/jacs.9b12435
52	Xu2021	https://doi.org/10.1021/jacs.1c06677

ID	Text Ref.	URL, DOI or Note
53	Mizuno2023	https://doi.org/10.1021/jacs.3c07356
54	Frederickson2017	https://doi.org/10.1021/acs.accounts.7b00004
55	MolesQuintero2022	https://doi.org/10.1002/anie.202209138
56	Schmidt2021	https://doi.org/10.1039/d0cc06810a
57	Dressler2020	https://doi.org/10.1016/j.chempr.2020.02.010
58	Barker2019	https://doi.org/10.1021/jacs.9b11898
59	Tobe2018	https://doi.org/10.1007/s41061-018-0189-0
60	Pozun2013	https://doi.org/10.1021/ja406002n
61	Schleyer1996	https://doi.org/10.1021/ja960582d
62	Geuenich2005	https://doi.org/10.1021/cr0300901
63	Minkin1994	ISBN-13:978-0471593829
64	Merino2004	https://doi.org/10.1002/chem.200400457
65	Chen2005	https://doi.org/10.1021/cr030088+
66	Stanger2019	https://doi.org/10.1002/cphc.201900952
67	Hermann2021	https://doi.org/10.1002/anie.202016968
68	Xu2012	https://doi.org/10.1021/ol3017353
69	Qiu2014	https://doi.org/10.1039/c3cc49418g
70	2020SciPy	10.1038/s41592-019-0686-2
71	orcaRef	10.18637/jss.v071.i10
73	Ernzerhof1999	https://doi.org/10.1063/1.478401
74	Becke1992	https://doi.org/10.1103/PhysRevA.38.3098
75	Lee1988	https://doi.org/10.1103/physrevb.37.785
76	Cheeseman1996	https://doi.org/10.1063/1.471789
77	Hariharan1973	https://doi.org/10.1007/bf00533485
78	Francl1982	https://doi.org/10.1063/1.444267
79	BrokenSymmEnergy	10.1021/jp9711757
80	BrokenSymm	10.1063/1.440939

ID	Text Ref.	URL, DOI or Note
86	Group7	<i>Konishi2017, Tobe2018, Dressler2018, Barker2019, Dressler2020, Xu2021, MolesQuintero2022, Mizuno2023</i>
87	Group6	<i>Levi2009, Kawase2009, Chase2010, Konishi2017, BaranacStojanovi2019</i>
88	Group5	<i>Tobe2014, Frederickson2017, Chen2020, Konishi2021, Hong2022</i>
	Group4	<i>Shimizu2022, Rajca2011, Gallagher2019</i>
	Group3	<i>Kawase2010, Wu2017, Zhang2017, ElBakouri2020, Sprachmann2023</i>
	Group2	<i>Kawase2010, Chase2012, Nishida2012, Liu2015, Dai2014, Ren2016, Yuan2016, Hermann2018, Schmidt2021, Horii2022</i>
	Group1	<i>Usuba2021, Esser2022, Konishi2017, Wassy2019, Wssner2021, Wssner2022</i>
	NOTE1	<i>In a two-state system (singlet and triplet), the temperature-dependent Boltzmann population analysis of the triplet state, $p_{\text{Triplet}}(T)$ in % , is defined as $p_{\text{Triplet}}(T) = 100 \cdot (3/Z)$, where 3 stands for the spin multiplicity and Z is the system's partition function, calculated as $Z(T) = 3 + \exp(-\Delta E_{\text{ST}}/RT)$, where the energy origin is set in the triplet state energy. With a ΔE_{ST} value of 5.7 kcal/mol, $Z(T)$ approaches $3 + 6.5 \cdot 10^{-5}$ at room temperature, leading to $p_{\text{Triplet}}(T=298.15\text{K}) \approx 100\%$.</i>
	NOTE2	<i>We here used $\text{NICS}_{\text{iso}}(1)$ instead of other versions of NICS such as $\text{NICS}_{\text{zz}}(1)$. As $\text{NICS}_{\text{iso}}(1)$ considers the average of the trace of the magnetic tensor, σ-interactions are also considered. This provides a consistent picture of surf-NICS with respect to the BLA analyses, where σ-interactions also play a significant role in determining the bond lengths.</i>

Supporting Information
Rational design of organic
diradicals with robust high-spin
ground state based on
antiaromatic building blocks

Supporting Information: Rational Design of Organic Diradicals with Robust High-Spin Ground State Based on Antiaromatic Building Blocks

Raul Santiago, M. Àngels Carvajal, Ibério de P. R. Moreira, Stefan T. Bromley, Jordi Poater, Mercè Deumal, Jordi Ribas-Ariño**

E-mail: raul.sant.1972@gmail.com & j.ribas@ub.edu

S1. Description of the Spectral Graph Theory Method

In this section we focus on outlining the Spectral Graph Theory (SGT) methodology employed to pre-screen the frontier Molecular Orbital (MO) topology of the pentalene-like diradicals studied throughout this study. The SGT methodology is tightly bonded to quantum chemistry, specially to Huckel Molecular Orbital (HMO) theory.^{REF} Thus, we will leverage the interpretability of HMO to showcase the potential of SGT in order to extract meaningful insights, assess the MO topology and forecast the singlet-triplet gap (ΔE_{ST}) of fully-conjugated (poly-)cyclic hydrocarbons.

In HMO, the Hamiltonian of the system, H , is expressed in its matrix representation on the basis of p_z^i atomic orbitals (AOs) of the carbon scaffold.^{REF} Diagonal elements, h_{ii} , are assigned to α , accounting for the energy of the isolated p_z^i AOs. In turn, the off-diagonal elements, h_{ij} , take a value of either β or 0, serving as the resonance integral between adjacent p_z^i and p_z^j AOs or an absence of interaction, respectively. The Huckel MOs can be directly obtained by solving the associated eigenvalue equation, $\hat{H}C = CE$, where E and C are the matrices representing the MO energies and coefficients, respectively. The HMO Hamiltonian bears a clear relationship with the adjacency matrix, A , of the graph associated with the carbon scaffold. More precisely, A is identical to \hat{H} when choosing $\alpha = 0$ and $\beta = 1$. The choice of α leads only to a shift of the sign and value of the MO energies, but $\beta = 1$ maintains the same topological atom connectivity and, thus, A yields the exact same MOs than a given \hat{H} with the same value of β for all the interacting atoms. Following the

Aufbau and Pauli's exclusion principles, the fully-occupied, partially-occupied, and unoccupied MOs can be straightforwardly obtained from the SGT perspective after diagonalizing the adjacency matrix associated with the carbon atoms of the system.

In order to assess the open-shell character of planar fully-conjugated systems, attention must be redirected towards the two frontier (potentially) singly-occupied MOs, namely Ψ_a and Ψ_b , and also to their associated energies, E_a and E_b . In non-degenerate cases, $E_a \neq E_b$, the *Aufbau* principle and Pauli's exclusion principle are sufficient to anticipate a closed-shell ground state in a first approximation. However, in the degenerate case, *i.e.* $E_a = E_b$, Hund's rule of maximum multiplicity favors a parallel alignment of the two electrons each in one MO. Despite two frontier MOs can "accidentally" be degenerate, here we have only focused on the diradicals possessing degenerate orbitals by design. At the HMO level of theory two orbitals are degenerate if (a) these are (anti-)bonding and symmetry related or (b) these are non-bonding Molecular Orbitals (NBMOs). As the sought degeneracy must occur at the frontier orbitals, we have restricted our exploration to the latter case. In this regard, either HMO or SGT can be employed to evaluate the energy gap between both frontier MOs, ΔE_{ab} , and also to visually assess their non-bonding topology.

Despite the Hund's rule predicting a triplet ground state for the degenerate case, further examination of the singly-occupied Molecular Orbitals (SOMOs) topology can be done to anticipate the "robustness" of such a triplet with respect to the singlet excited state. In this vein, various models for molecular magnetism{KAHN,HOFFMANN} emphasize the linear relationship between ΔE_{ST} and the exchange integral, K , and its critical importance when both SOMOs are degenerate. The exchange integral is formally defined as the following two-centers two-electrons electron repulsion integral:

$$K = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_2 - \mathbf{r}_1|^2} d\mathbf{r}_1 d\mathbf{r}_2$$

where $\rho(\mathbf{r}_i) = \Psi_a(\mathbf{r}_i)\Psi_b(\mathbf{r}_i)$ corresponds to the overlap density and \mathbf{r}_i stands as the coordinates of the i -th electron. The exchange integral attains its maximum value when both electrons are close in space, *i.e.* $|\mathbf{r}_i - \mathbf{r}_j| \rightarrow 0$. Thus, the magnitude of K is usually governed by the extrema of the $\rho(\mathbf{r}_i)$. Accordingly, diradicals displaying a pair of SOMOs that share significant non-null density regions, while still presenting NBMOs, are prone to present ferromagnetic interactions ($\Delta E_{ST} > 0$).

Given the specific NBMO topology of the SOMOs considered herein, several approximations can be done to grasp the trends of K at a low computational cost. First, bearing in mind the localized nature of the $\{p_z^i\}$ basis set, the overlap density can be approximated in the graph space as $\rho = \mathbf{c}_a \circ \mathbf{c}_b$, where \circ corresponds to the Hadamard (*i.e.* point-wise) product between the MO coefficients of both Ψ_a and Ψ_b SOMOs. In other words, the overlap density is assumed to be significantly non-null only at the atom centers. Analogously, K can be approximated as

$$K \approx \sum_i \sum_{j \in \nu(i)} \frac{\rho_i \rho_j}{D_{ij} + \epsilon}$$

where the first summation in the discretized version of the integral runs along all atoms i (vertices) and the second summation runs along all j neighbors of i , $\nu(i)$. Moreover, D_{ij} corresponds to the distance between nodes i and j , while ϵ is a small non-physical parameter that prevents divergence for $j = i$, as $D_{ii} = 0$ in this case. Since only NBMOs are considered, the products $\rho_i \rho_j$ tend to 0 for $i \neq j$, as for a given non-null ρ_i component, its contiguous ρ_j neighbor should be zero by definition. Thus, the expression of K can be further approximated as

$$K \approx \sum_i \frac{\rho_i^2}{\epsilon} \propto \sum_i \rho_i^2 = |\rho|^2$$

The assumptions considered this far lead to a clear relationship between the SOMOs' (non-)disjoint character and K . That is, when the SOMOs Ψ_a and Ψ_b can be localized to two different sets of atoms by any linear combination between them (*i.e.* disjoint set), all the components of the overlap density are 0 and, likewise, the exchange integral and ΔE_{ST} are zero. However, when Ψ_a and Ψ_b cannot be localized to different sets of atoms by any linear combination (no-disjoint set) some components of the overlap density must be non-null ($\rho_i \neq 0$), thus resulting also in non-null K and a $\Delta E_{\text{ST}} > 0$. The full set of pairs of orthonormal linear combinations of the SOMOs can be obtained by incorporating the parameter θ :

$$\psi_a^\theta = \cos(\theta)\Psi_a + \sin(\theta)\Psi_b$$

$$\psi_b^\theta = \sin(\theta)\Psi_a - \cos(\theta)\Psi_b$$

The task in hand is thus to minimize $\rho(\theta)=\varphi_a(\theta)\varphi_b(\theta)$ and $K(\theta) = |\rho(\theta)|^2$ in the interval $[0, \pi/2)$. Accordingly, if $K(\theta_{\min}) \neq 0$ the SOMOs comprise a non-disjoint set and, thus, could be considered excellent candidates to present ferromagnetic interactions.

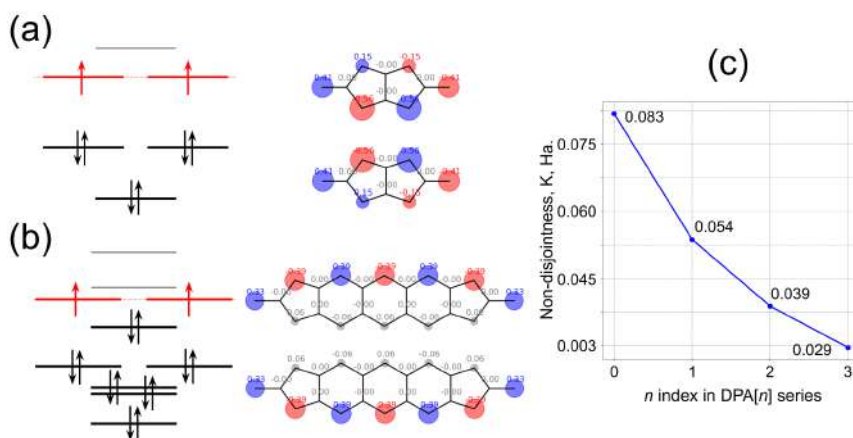
We employed this theoretical framework to pre-screen the non-bonding molecular orbitals (NBMOs) and non-disjoint character of all pentalene-like molecules explored in this work. Due to the simplicity of the method, the execution time for a moderately sized molecule is lower than the order of a second. Moreover, despite the evident limitations of this approach, our *Ab Initio* results and conclusions interestingly correlate well with such a simplistic procedure, showcasing its utility in the preliminary analysis of open-shell character in planar fully-conjugated systems. In a lesser note, SGT is unable to differentiate between the effect of different types of atoms in the set up defined so far. In this regard, the method is limited to all-carbon systems. If the incorporation of a hetero-atom does not significantly alter the MO topology, however, the SGT approach can still provide correct trends.

S2. Spectral Graph Theory Results

In the following section the molecular orbital representation of the diradicals derived from **DPA**[*n*], **DBP** are shown. As SGT method is not able to differentiate between hetero-substitutions only the full-carbon scaffolds are considered here.

S2.1. **My**₂-DPA[*n*] singly-occupied Molecular Orbitals

The energy diagram for **My**₂-DPA[0] and **My**₂-DPA[3] are shown in Figure {FS1}a and {FS1}b as the lowest and highest order dicyclopenta-[*n*]acene considered in this work. As illustrated, the SOMOs of both compounds are predicted degenerate at SGT level (see levels highlighted in red). Moreover, the pair of SOMOs for both diradicals presenting the lowest exchange integral, $K(\theta_{\min})$, comprise a non-disjoint set (see Figure {FS1}c-d), specially displaying a significant non-null overlap density on the **My** carbon atoms, thus anticipating a positive ΔE_{ST} gap in both cases. The SGT-based SOMOs for all compounds of the **My**₂-DPA[*n*] series present similar MO topology. However, as illustrated in Figure {FS1}e we have observed a decrease in *K* as increasing the acene size (*n*), possibly attributed to the higher delocalization of the SOMOs as progressing through the series. Accordingly, the SGT results indicate a lower ΔE_{ST} gap for larger values of *n*.

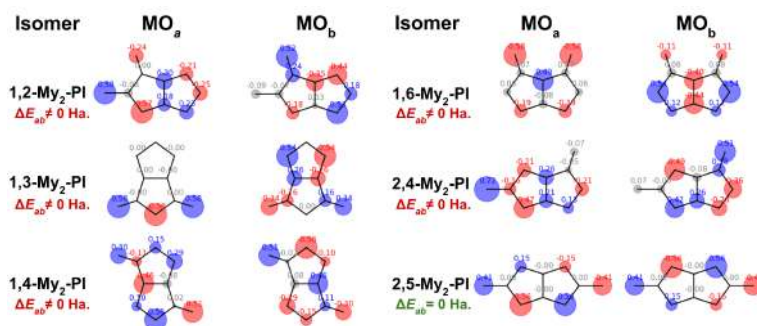


{FS1} **Figure S1.** Schematic representation of SGT results for $My_2-DPA[n]$ series. The energy levels, occupation and the most non-disjoint linear combination of SOMO projected onto the carbon scaffold for (a) $My_2-DPA[0]$ and (b) $My_2-DPA[3]$ compounds are illustrated. For the pair of two most non-disjoint SOMOs see inset coefficients of the corresponding linear combination. (c) Approximate exchange integral (K), computed as explained in Section S1, for the first 4 ($n=0,3$) $My_2-DPA[n]$ diradicals.

S2.2. MO topology of the structural isomers of $My_2-DPA[0]$

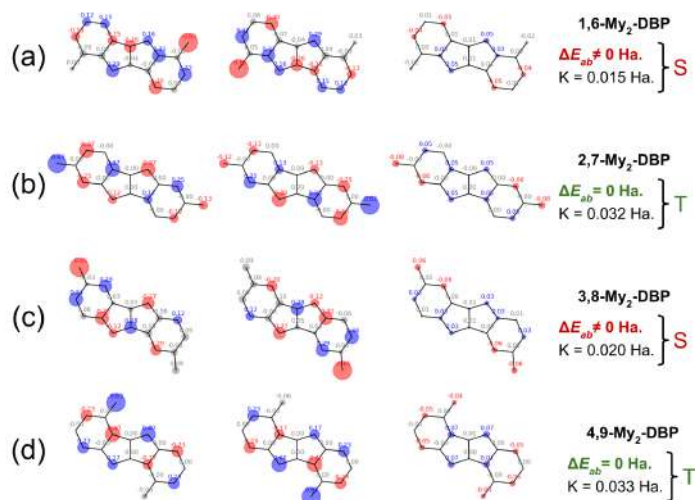
Figure {FS2} shows the frontier orbitals for all the possible structural isomers of $My_2-DPA[0]$. Among them, only **2,5- $My_2-DPA[0]$** present a set of two nonbonding frontier orbitals. Accordingly, **2,5- $My_2-DPA[0]$** is the only isomer displaying degenerate SOMOs and, thus, it is the only structural isomer for which a significant ΔE_{ST} gap can be predicted.

Interestingly, different scenarios are found across the rest of isomers. For example, **1,2-**, **1,4-** and **2,4- $My_2-DPA[0]$** compounds possess two (anti-)bonding MOs, while for **1,3-** and **1,6- $My_2-DPA[0]$** one of MOs present a non-bonding character (see MO_3). This diversity illustrates the key importance of atomic connectivity in designing diradicals with specific properties. Although not shown here, the results presented here for **DPA[0]** couplers extend to the whole **DPA[n]** series.



{FS2} **Figure S2.** SOMOs topology and ΔE_{ab} (gap between SOMOs 'a' and 'b') for all the structural isomers of $My_2\text{-DPA}[0]$. While all the compounds are found to possess a pair of non-disjoint SOMOs, the only isomer displaying a gap of $\Delta E_{ab} = 0$ is $2,5\text{-}My_2\text{-DPA}[0]$.

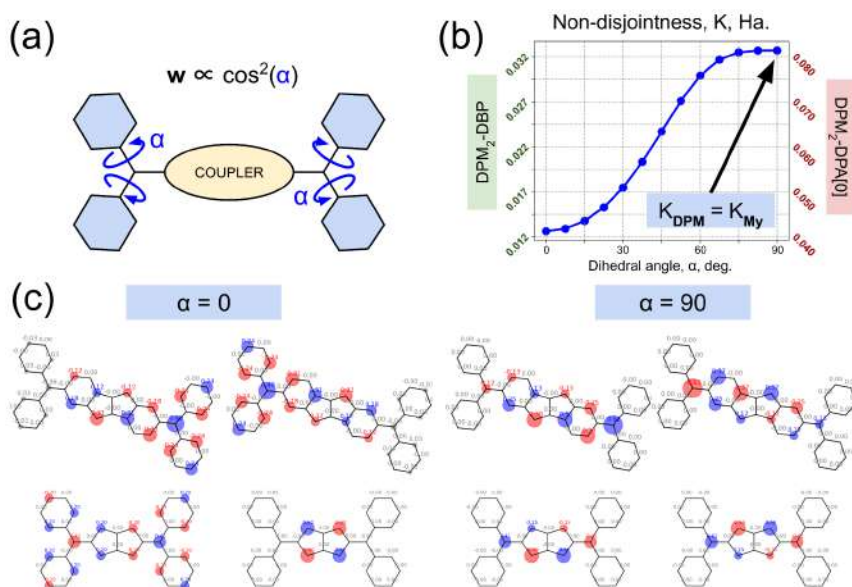
S2.3. $My_2\text{-DBP}$ singly-occupied Molecular Orbitals



{FS3} **Figure S3.** SOMOs topology, overlap density and ΔE_{ab} (gap between maximally disjoint SOMOs) for all the structural isomers of $My_2\text{-DBP}$. While all the compounds are found to possess a pair of non-disjoint SOMOs, the only isomers displaying a gap of $\Delta E_{ab} = 0$ are $2,7\text{-}My_2\text{-DBP}$ and $4,9\text{-}My_2\text{-DBP}$.

Figure {FS3} shows the summary of SGT results for the different structural isomers of **My₂-DBP** diradicals. We have observed a non-disjoint set of MOs in all the cases (see MOs in the first and second columns and the overlap density at the third one), thus leading to a substantial overlap density and K in all the cases. Moreover, a degenerate pair of SOMOs is predicted for **2,7-** and **4,9-My₂-DBP** diradicals (see Figure {FS3}b,d) and, thus, a substantial ΔE_{ST} gap can be anticipated in both cases. Our results show, however, that **1,6-** and **3,8-My₂-DBP** compounds possess non-degenerate MOs (see Figure {FS3}a,c) and, thus, a negative ΔE_{ST} gap is predicted to this end. Bearing in mind the substantial K , more accurate calculations must be carried out to determine the spin alignment of the ground state.

S2.4. Beyond methylenyl cores: DPM₂-DPA[0] and DPM₂-DBP results



{FS4} **Figure S4.** Evaluation of the disjointness of SOMOs by means of the parameterization of the C-C interaction, w in (a), to simulate a phenyl dihedral twist. (a) Schematic representation of the parameterization, where w is assumed proportional to the cosine squared of the dihedral angle. (b) Non-disjointness resolved for various angles. (c) Maximally disjoint SOMOs at the flat ($\alpha=0$) and perpendicular ($\alpha=90$) phenyl conformations for My₂-DPA[0] and 2,7-My₂-DBP diradicals.

The variations of ΔE_{ST} gap of **DPM** and **PDM** substituted diradicals with respect to the **My** cases can be explained by changes in the dihedral angle between the bare coupler and the terminal aryl groups (see α in Figure {FS4}a). The work of Alc3n *et al.*{REF} disclose a correlation between the π - π overlap (S_{π}) of atoms centered at the phenyl moieties and the neighboring atom with respect to their dihedral angle, particularly $S_{\pi}=\cos^2(\alpha)$. Within the generalized Wolfsberg-Helmholtz approximation the resonance integral (β) is linearly proportional to S_{π} .

In order to consider this kind of conformational effects within the SGT formalism, we have extended the SGT capabilities to weighted adjacency matrices. We have substituted particular values of $\beta=1$ for a selected weight (w from 0 to 1 since it is inferred from S_{π}) serving as a fictitious twist of the phenyl rings.

We have computed K for various values of α from 0° to 90° for **DPM₂-DPA[0]** and **DPM₂-DBP** diradicals in order to assess the effect of such dihedral twist to the ΔE_{ST} gap (see Figure {FS4}b). Interestingly, when the phenyl rings and the coupler are fully in-plane ($\alpha = 0^{\circ}$), the SOMOs are delocalized towards the phenyl moieties and, thus, the net overlap density drops (see Figure {FS4}c). When the phenyl rings are completely orthogonal to the coupler's plane, however, the phenyl rings and the coupler are essentially disconnected, thus leading to the same topology of the SOMOs as in the corresponding **My₂-DPA[0]** and **My₂-DBP** (see Figure {FS1}a for a visual comparison).

Although other polarization effects, such as electron density pulling of the **Cl** atoms can also play a significant role in **PDM₂-DPA[0]** and **PDM₂-DBP** diradicals, our results prove that the conformational effect associated to the twisting angle of the phenyl moieties can roughly triplicate the exchange integral, K , and thus, the ΔE_{ST} gap at SGT level of theory.

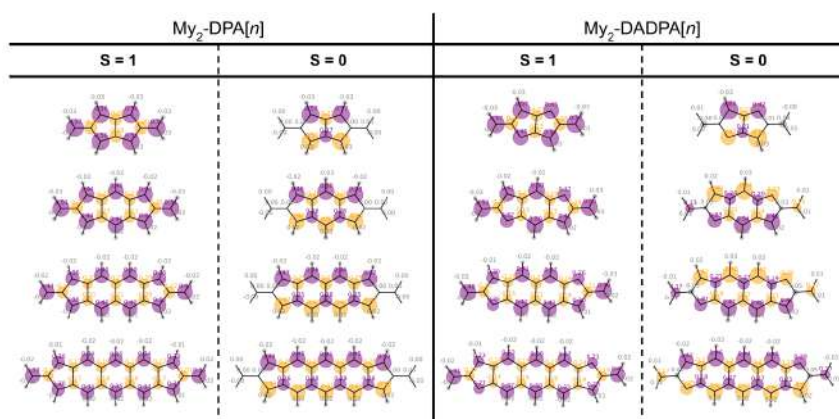
S3. Mulliken spin population and ERIs

In the following section the Mulliken spin population and some relevant molecular orbital integrals for the gap are shown. The population analysis has been carried out in the same calculation as the optimization calculations in Gaussian09 packages. However, Gaussian09 package does not provide a ready-to-use access to Coulomb, Exchange or other two-electron electron repulsion integrals (ERIs) at the MO basis. For this reason, the MO integrals have been evaluated separately in pySCE, using the atomic positions of the ground state of each diradical as obtained

from PBE0/6-311G* in Gaussian09. We employed the same PBE0/6-311G* level of theory for pySCF calculations.

S3.1. Mulliken spin population of $\text{My}_2\text{-DPA}[n]$ and $\text{My}_2\text{-DADPA}[n]$ diradicals

Figure {FS5} shows the Mulliken spin population (MSP) of $\text{My}_2\text{-DPA}[n]$ and $\text{My}_2\text{-DADPA}[n]$ diradicals at the triplet ($S=1$) and singlet ($S=0$) spin multiplicities. Overall we found that the spin distribution at a given n is substantially different for $S=0$ and $S=1$ (see positive and negative MSP distribution colored in purple and yellow, respectively). Particularly, for the $\text{My}_2\text{-DPA}[n]$ series we found that the MSP distribution at $S=0$ at the **My** cores and its adjacent atoms is null. However, for $S=1$, the carbon atom of the **My** core possesses a large spin population. Overall, the MSP analyses indicate that unpaired electrons are substantially more delocalized in the triplet than in the singlet spin configuration. Although for $\text{My}_2\text{-DADPA}[n]$ series the **My** core displays a non-null MSP, its magnitude is considerably lower compared to the triplet state.

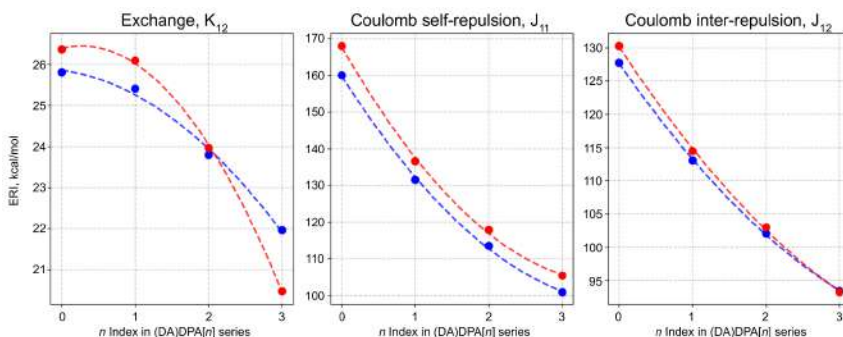


{FS5} **Figure S5.** Atom-resolved Mulliken spin population (MSP) for $\text{My}_2\text{-DPA}[n]$ and $\text{My}_2\text{-DADPA}[n]$. First and second columns correspond to $S=1$ and $S=0$ spin multiplicities for $\text{My}_2\text{-DPA}[n]$ while third and fourth columns correspond to $\text{My}_2\text{-DADPA}[n]$. Positive and negative populations are indicated in purple and yellow, respectively. The marker size is in agreement with its magnitude.

S3.2. ERIs of $\text{My}_2\text{-DPA}[n]$ and $\text{My}_2\text{-DADPA}[n]$ diradicals

Figure {FS6} illustrates exchange (K), Coulomb self-repulsion (J_{11}) and Coulomb inter-repulsion (J_{12}) electron repulsion integrals (ERIs) along the $\text{My}_2\text{-DPA}[n]$ and $\text{My}_2\text{-DADPA}[n]$

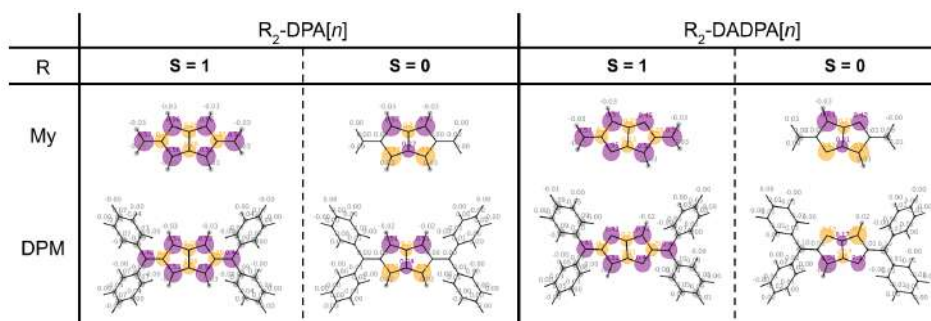
series (colored in blue and red, respectively). We have observed that all the ERIs decrease as the $[n]$ acene size increases. The trend is related to the increasing delocalization of the SOMOs as n increases. That is, for high values of n , the exchange integral, K , is lower than for smaller n 's and, thus, the ΔE_{ST} gap decreases as n increases. Our results also determine a larger slope for **My₂-DADPA** $[n]$ than for **My₂-DPA** $[n]$ series, in agreement with the slope registered for the corresponding ΔE_{ST} gaps, reported in the main text.



{FS6} **Figure S6.** Electron Repulsion Integrals (ERIs) for **My₂-DPA** $[n]$ and **My₂-DADPA** $[n]$ series, colored in blue and red, respectively. Dashed lines do not indicate continuity, instead, these are only included as a guide to the eye. The ERI type is documented on top of each plot.

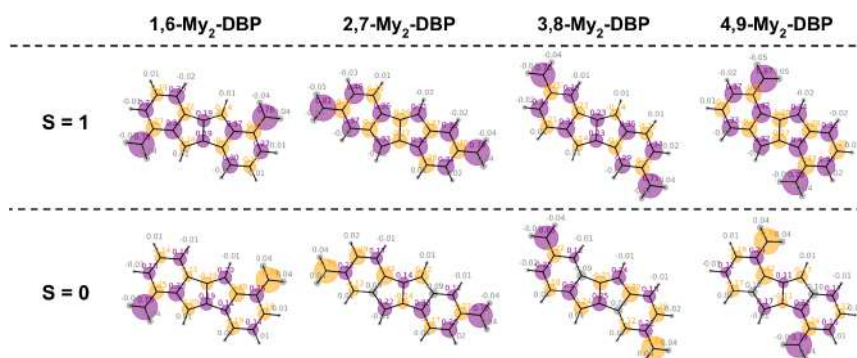
S3.3. Spin population of **My₂-(DA)DPA** $[n]$ and **DPM₂-(DA)DPA** $[n]$

Figure {FS7} illustrates the Mulliken spin population (MSP) for **My** and **DPM** cores. In agreement with the ΔE_{ST} gap results, we have only observed a small effect of the core's choice to the MSP in both $S=0$ and $S=1$ spin multiplicities. MSP results indicate that the **DPM** core does not critically quench the spin distribution and, thus, it does neither affect the ΔE_{ST} gap.

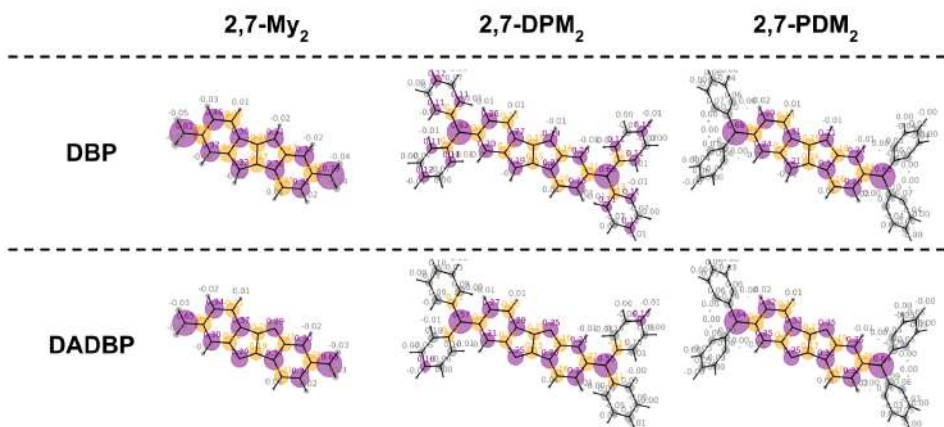


{FS7} **Figure S7.** Atom-resolved Mulliken spin population (MSP) for R_2 -DPA[n] and R_2 -DADPA[n]. First and second columns correspond to $S=1$ and $S=0$ spin multiplicities for R_2 -DPA[n] while third and fourth columns correspond to R_2 -DADPA[n]. First and second columns correspond to $R=My$ and $R=DPM$ cores. Positive and negative populations are indicated in purple and yellow, respectively. The marker size is in agreement with its magnitude.

S3.4. Mulliken spin population of R_2 -DBP and R_2 -DADBP diradicals



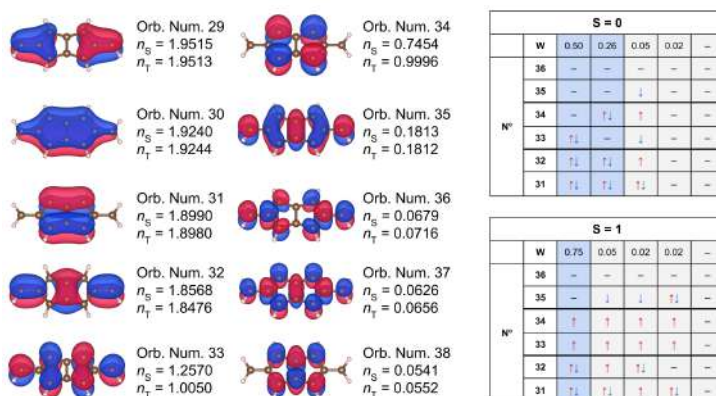
{FS8} **Figure S8.** Mulliken spin population (MSP) for various My_2 -DBP structural isomers. First and second rows correspond to $S=1$ and $S=0$ spin multiplicities. Positive and negative populations are indicated in purple and yellow, respectively. The marker size is in agreement with the magnitude of the MSP.



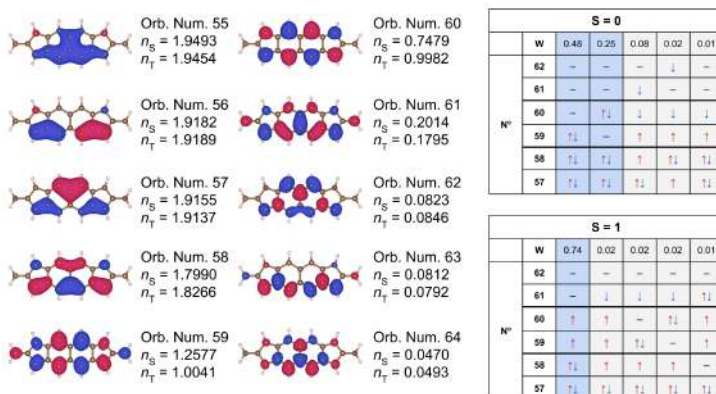
{PS9} **Figure S9.** Mulliken spin population (MSP) for different core substitutions ($R=My, DPM, PDM$) of 2,7- R_2 -DBP. First and second rows correspond to $S=1$ and $S=0$ spin multiplicities. Positive and negative populations are indicated in purple and yellow, respectively. The marker size is in agreement with the magnitude of the MSP.

S4. Further results related to ΔE_{ST} gaps

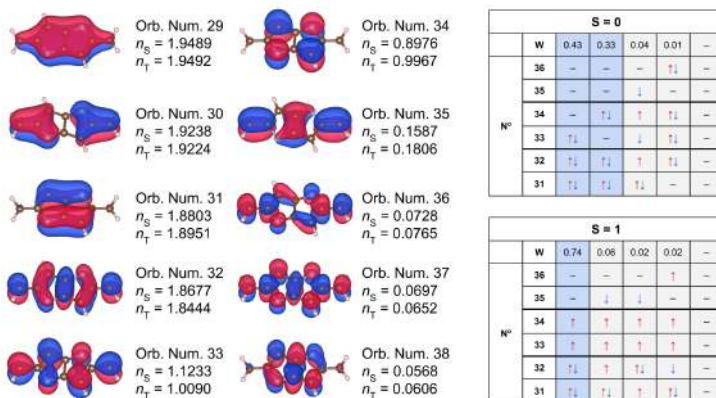
S4.4. CAS(10,10) Active MOs and Multi-referential CASCF weights



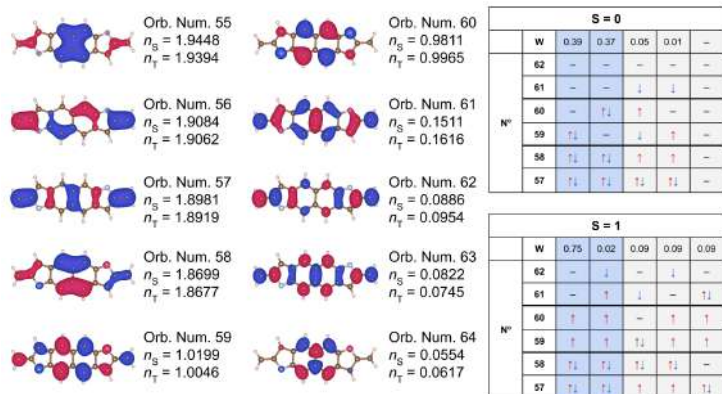
{FSX:SI_F10_01_My2DPA0.png} Figure SX. *TODO.*



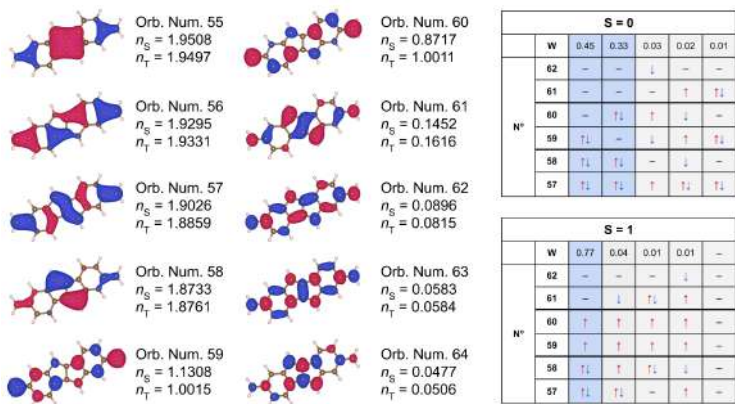
{FSX:SI_F10_02_My2DADPA0.png} Figure SX. *TODO.*



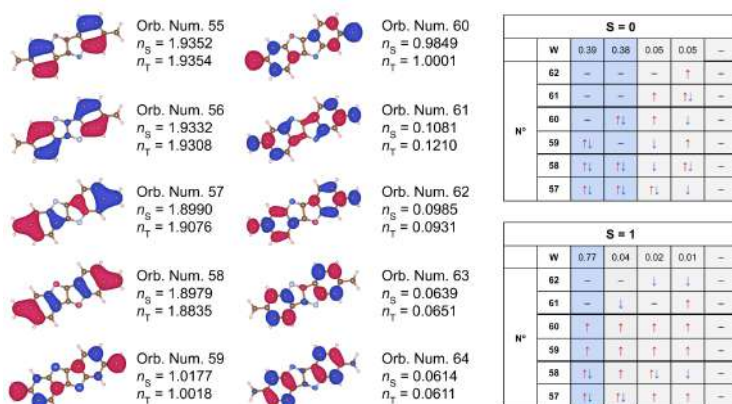
{FSX:SI_F10_03_My2DPA2.png} Figure SX. *TODO.*



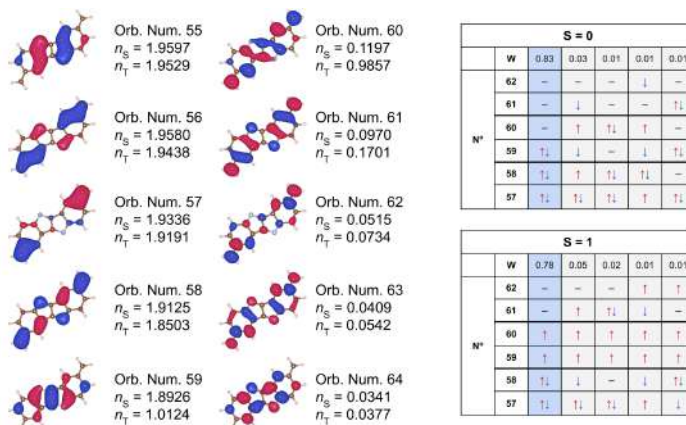
{FSX:SI_F10_04_My2DADPA2.png} Figure SX. *TODO.*



{FSX:SI_F10_05_My2DBP_01.png} Figure SX. *TODO.*



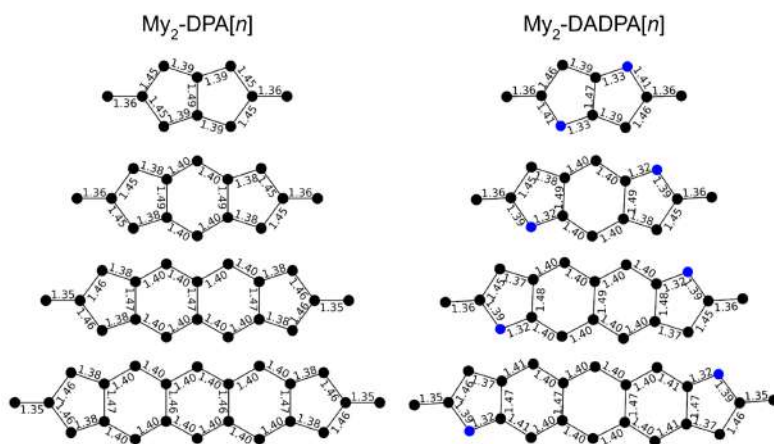
{FSX:SI_F10_06_My2DADBP_01.png} Figure SX. *TODO.*



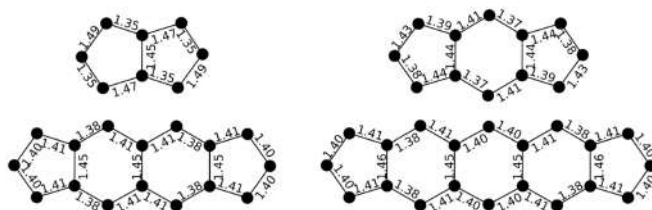
{FSX:SI_F10_07_My2DADBP_01.png} Figure SX. *TODO.*

S5. Further Bond Length Alternation Analyses

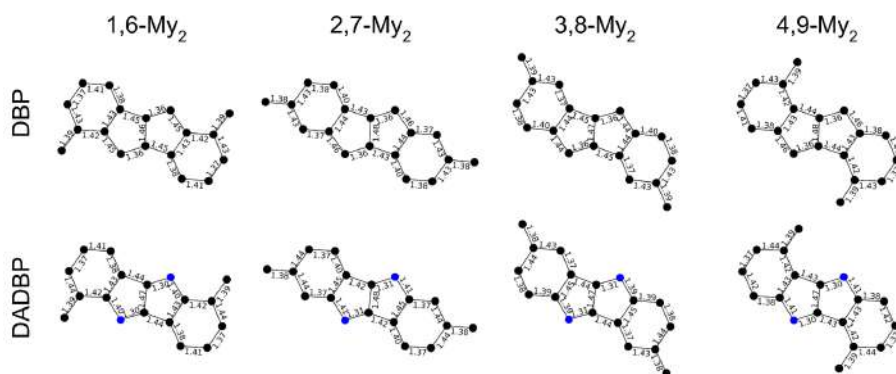
TODO.



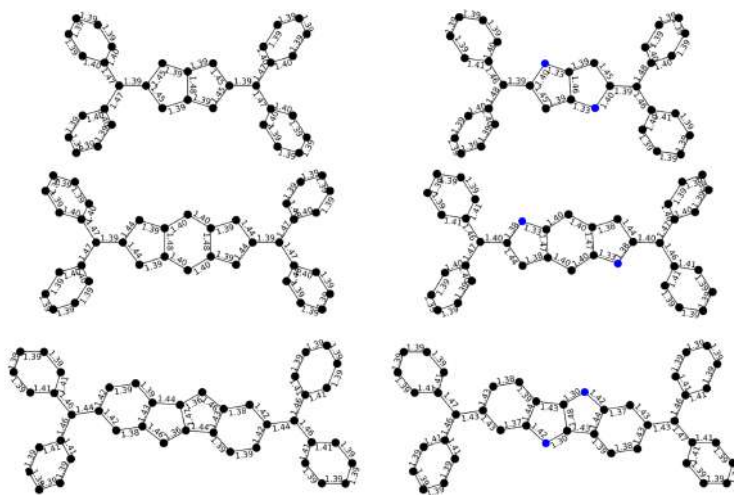
{FSX:SI_F12_DPA_BLA.png} Figure SX. *TODO.*



{FSX:SI_F12_DPA_BLA.png} Figure SX. *TODO.*

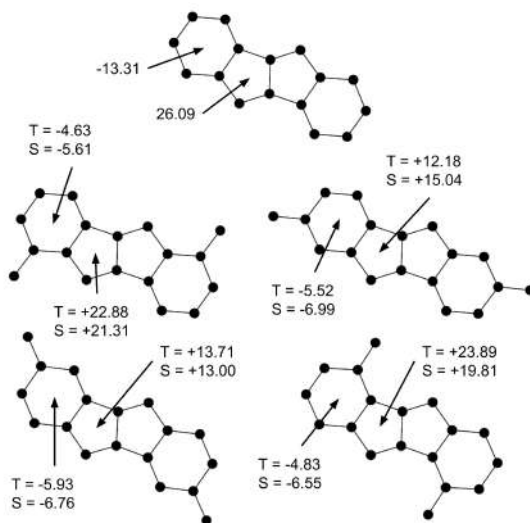


{FSX:SI_F13_DBP_BLA.png} Figure SX. *TODO.*



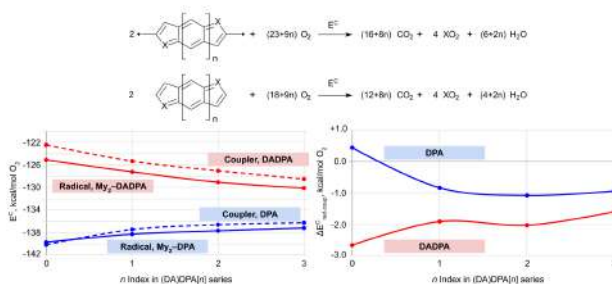
{FSX:SI_F14_DPA-radicals-BLA.png} Figure SX. *TODO.*

S6. Further NICS results for My₂-DBP isomers

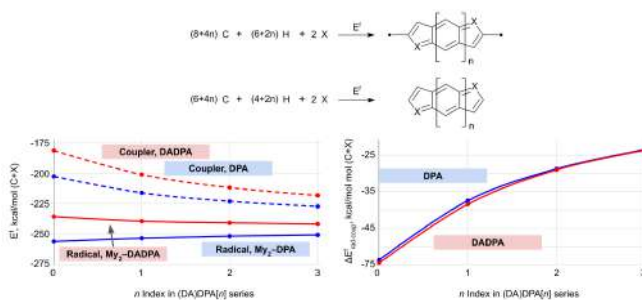


{FSX:SI_F15_DBP_NICS.png} Figure SX. *TODO.*

S7. Combustion and Formation Energies of (DA)DPA[n]



{FSX:SI_F09_01_combustion.png} Figure SX. *TODO.*



{FSX:SI_F09_01_formation.png} Figure SX. *TODO.*

References

ID	Text Ref.	URL, DOI or Note
1	Borden1977	https://doi.org/10.1021/ja00456a010
2	Borden1994	https://doi.org/10.1021/ar00040a004

Chapter 5

A new quantum-informed
representation for ML applications
in chemistry

5.1. Introduction

Over the past few years, machine learning (ML) has undergone a remarkable technological transformation, opening the door to a wide array of applications. These applications span across various domains, becoming integral parts of our daily lives.^{1,2} In technology, ML has made significant contributions to areas like natural language processing,³ autonomous driving systems,⁴ and advanced control systems.⁵ Additionally, in the scientific community, ML has made impactful strides in fields such as medical diagnosis,⁶ particle physics,⁷ nanotechnology,⁸ brain interface technologies,⁹ analysis of social media trends,¹⁰ robotics,¹¹ and various types of gaming and strategic board games.^{12,13} These methods are also having a tremendous impact in Quantum Chemistry (QC).¹⁴⁻¹⁹ Accordingly, several fields such as molecular electronics,^{20,21} excited states,²²⁻²⁴ low-cost discovery of new materials,^{25,26} or catalysis²⁷ are benefiting from new computational strategies combining QC and ML.

This development inspired researchers to approach challenges in chemical science with these tools, driven by the prospect that ML would revolutionize their respective fields in a similar way. Similarly, in the early stages of the work gathered in this chapter, the objective was to readily apply tested ML methods to predict magnetic exchange coupling, J_{AB} , as further elaborated below. This endeavor was perhaps started with a naive optimism, the full extent of which was not yet apparent.

Contrary to the usual experience when learning other subjects such as Quantum Mechanics, ML can take a rapid initial learning process. This is in part facilitated by accessible mathematical concepts, a plethora of ready-to-go, user-

friendly and free-of-tax software, together with abundant and curated datasets or tutorials. It thus usually follows that one can early feel proficient in the application of ML in practical scenarios. However, while such an initial confidence is encouraging and should be celebrated, it might also be symptomatic of a manifestation of the phenomenological Dunning-Kruger effect,²⁸ i.e., a cognitive bias where limited knowledge leads to an inflated self-perception (see the "Peak of Mt. Stupid" in Figure 5.1). Overall, this can be materialized as a deceptive sense of mastery in the early stages of the research.

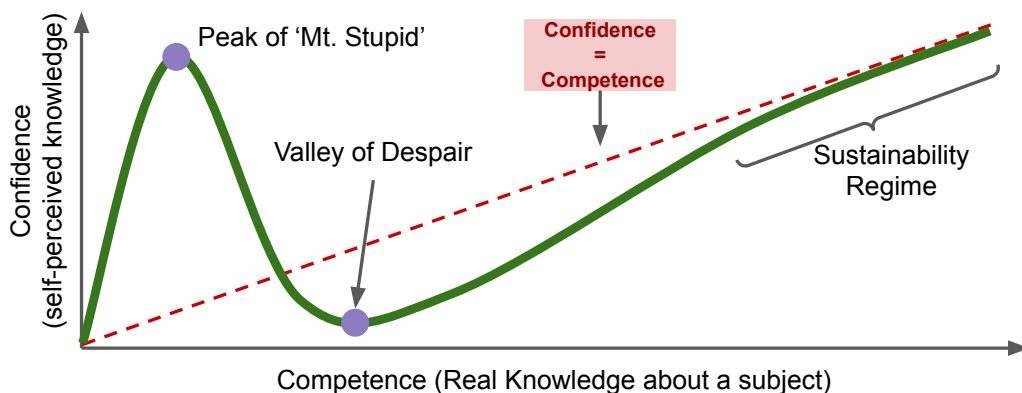


Figure 5.1: *Qualitative representation of the phenomenological Dunning-Kruger effect. As progressing along the x-axis (competence) the highlighted regions are the Peak of Mt. Stupid, Valley of Despair and Regime of Sustainability. Dashed red lines correspond to the identity function ($y=x$) where the ideal behavior, i.e., Confidence = Competence, is achieved.*

As progressing beyond the initial successes, the application of the still blurred ML concepts to the frontiers of novel science leads to what is known as the "Valley of Despair", as per Dunning-Kruger effect (see Figure 5.1). As it can be recognized from the name of this stage, it is marked by the realization that the complexity of ML vastly exceeds the current grasp, often leading to frustration towards the unsuccessful experiments and the inability to conceptually

identify the origin of such failure. The challenge here is not only to recognize the evident gap in knowledge but to actively seek to progressively fill it. Overall, understanding that mastery is not a rapid ascent, but a journey marked by continuous learning and adaptation, inevitably leads to the "Regime of Sustainability" (as illustrated in Figure 5.1). Unfortunately, recognizing the own stage in the Dunning-Kruger curve is intricate as, for example, the self-perception of being in the sustainability regime most likely places one in the stupidity peak. However, recognizing this fact subsequently hints that one is, indeed, falling to the despair well, and so on.

In short, we anticipated that existing ML techniques would readily extend to our project. However, as we delved deeper, the challenge proved more complex than anticipated. That is, what seemed straightforward at first instances revealed some small yet critical intricacies, which forced us to reevaluate our objectives. Instead of solely relying on established methods, we undertook the development of a new descriptor, one that was better suited to the particularities of the system we were studying. This was not just an extension of our original objectives, but a significant blend that broadened the scope and depth of our research.

Instead of a standard section on ML techniques, it has been thought to be more profiting to report the obstacles that this PhD candidate has overcome before completing the work presented in this chapter. While the following sections are not meant to be arranged chronologically, for the sake of the readership, they are intended to be organized such that the pitfalls can be recognized as the readers progress along the chapter. Accordingly, the following introductory sections are dedicated to outline the prevalent elements along our work. In this regard, **Section 5.1.1** offers a detailed description of the chemical species to which ML

has been applied. This is followed by **Section 5.1.2**, which provides an overview of the regression model (i.e., the core ML engines) and learning strategies used consistently throughout our work. Finally, **Section 5.1.3** delves into a brief discussion on the importance of chemical descriptors in ML applications within chemistry. These foundational elements set the stage for the deeper explorations and insights that are to follow. For further details on other introductory aspects related to our research, readers are directed to the *Introduction* and *Dataset* sections of the publication attached at the end of this chapter.

In the Results and Discussion section (**Section 5.2**), we conduct an in-depth analysis of the results obtained from applying ML models and descriptors to our chosen dataset. The initial part of this section, **Section 5.2.1**, focuses on the results from the direct utilization of existing ML models. This analysis reveals the first significant limitation: the inability of current descriptors of choice to predict intermolecular properties accurately. The next section, **Section 5.2.2**, discusses the outcomes after implementing necessary adjustments to our approach. Here, a second critical issue emerges, related to the fundamental accuracy and suitability of the descriptors when dealing with the prediction of properties deeply rooted in quantum mechanics. In **Section 5.2.3**, we describe the development of our novel descriptor, which was crafted to address the identified shortcomings and the challenges we faced. The results section concludes with **Section 5.2.4**, where a detailed comparison between the developed descriptor and the existing ones is conducted. We believe that this comparative study not only demonstrates the improvements made by our descriptor but also highlights the significant contributions of our research.

5.1.1. Selection of the dataset

As outlined in the previous section, the objective of this work is to train a ML model to predict magnetic exchange couplings, J_{AB} , between two organic open-shell spin-holding units A and B . When this project was in its initial stages, we found that the work conducted by Herrmann²⁹ and co-workers had already explored the prediction of J_{AB} interactions, primarily focusing on magnetic interaction in metal-organic complexes. Given that the central element of this PhD dissertation is restricted to fully-organic molecules and materials, the dataset used in their study was not directly applicable to our scope of research. Thus, we have generated three distinct datasets for our purpose, each serving a unique objective in exploring the relationship between structural representation and J_{AB} . Despite that in the publication all three of them are covered and tested, in this chapter we have chosen to review only the most relevant for a streamlined discussion.

Overall, we leveraged the previous experience of the research group to generate samples collecting structures and J_{AB} values of dimers of 1,3,5-trithia-2,4,6-triazapentalenyl radical (TTTA in short, see Figure 5.2) from their arrangement in the solid state. In this regard, on heating above 300 K, TTTA organizes into a paramagnetic, monoclinic phase (referred to as the high-temperature or HT phase), characterized by four columns of TTTA units (marked as "CX" in Figure 5.2a) that pile up due to labile π - π interactions³⁰ (note that "DX" stands for adjacent different pairs of TTTA radicals along the b axis, see Figure 5.2b). The regular stacks of radicals (i.e. with a uniform distance between adjacent radicals) observed in X-ray measurements of the HT phase were shown to result from a rapid intra-stack Pair Exchange Dynamics³¹ (PED), a thermally

activated process where TTTA radicals continually exchange adjacent neighbors within the stack, as depicted in Figure 5.2c. Accordingly, when the thermal energy overcomes the intermolecular energy, the PED process is triggered and the inter-planar distance between a given pair of TTTA molecules is observed

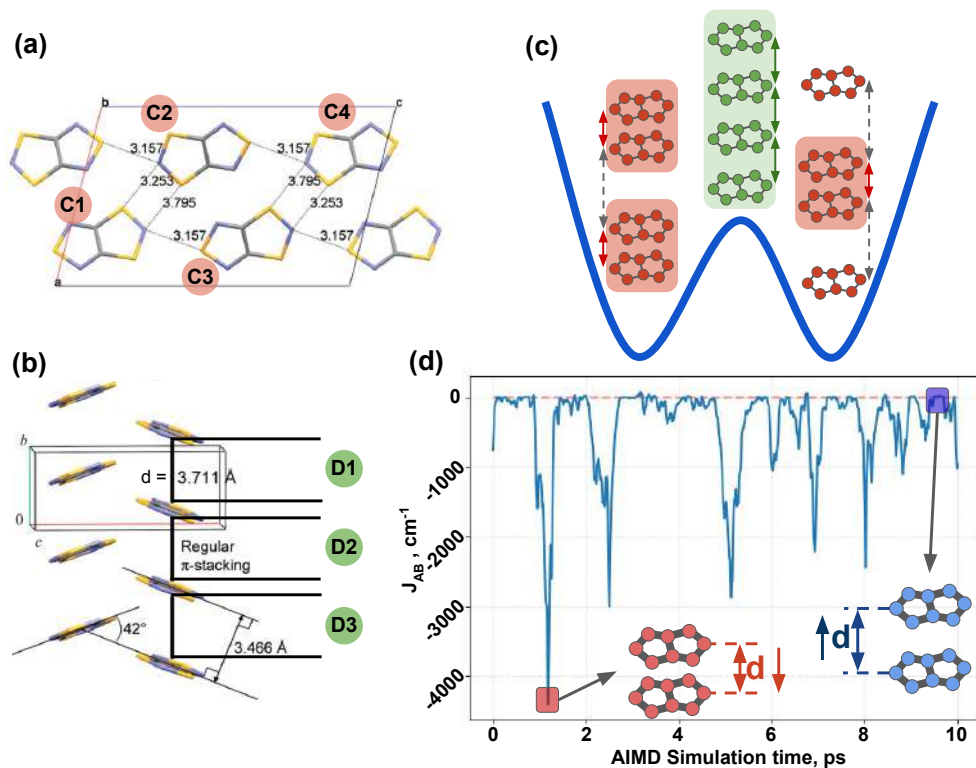


Figure 5.2: (a) Top ac -view and (b) side along b -axis view of the monoclinic polymorph unit cell of TTTA. Insets in (a) and (b) provide labels to identify the TTTA dimers throughout the unit cell. For instance $D2C1$ stands for the dimer $D2$ in (b) located at column $C1$ in (a). (c) Schematic depiction of the Pair Exchange Dynamics (PED) phenomenon, which can be understood as a dynamic interconversion between two degenerate dimerized stacks. The path along the Potential Energy Surface (blue) features a double well, illustrating that the dimerized TTTA π -stacks occupy a lower energy state than the equidistant columns. (d) Time resolved evolution of J_{AB} of two TTTA radicals along 10 ps. Insets highlight two regions with extremely different J_{AB} values: -4300 cm^{-1} (red) and $+20 \text{ cm}^{-1}$ (blue).

to rapidly oscillate between ca. 3.25 Å and ca. 4.5 Å. Such subtle structural changes lead to massive time- and temperature-dependent variations in the J_{AB} interaction between pairs of π -stacked TTTA units,³² ranging from weakly ferromagnetic (FM, ca. 50 cm⁻¹) to strongly antiferromagnetic (AFM, ca. -5000 cm⁻¹), as shown in Figure 5.2d.

Following the approach of previous works for large sample mining,^{20,33–35} we obtained the TTTA dimers from Ab Initio molecular dynamics (AIMD), carried out at two different temperature conditions, 250 K and 300 K (i.e. below and above the PED phenomenon fully come into effect). While the data at 300 K was readily available from another publication³¹ conducted by some of the co-authors of this project, the data at 250 K is a new addition. Importantly, each AIMD simulation had an independent 10 ps run, resulting in a total of 10,000 time-resolved structures that ergodically explore different regions of the thermally available configuration space. Regarding the data from the simulation at 300 K, we selected 30,000 configurations from the AIMD trajectory, which were chosen from three different dimers within the crystal structure (with 10,000 configurations derived from each dimer). Likewise, the data from the simulation at 250 K comprises a smaller set of 20,000 structures, selected following the same strategy as the 300 K data. Finally, it is important to mention that the J_{AB} values for each structure of the set have been obtained by means of DFT calculations of isolated pairs of radicals using the UB3LYP functional and a 6-31+G* basis set. This approach allowed for a detailed analysis of how J_{AB} varies upon progressing along the time-resolved configurations of TTTA pairs.

The inclusion of the two sets of TTTA dimers aims to evaluate the extrapolation capabilities of the ML engine, specifically its ability to generalize from

training on the 300 K data to making accurate predictions at 250 K. Overall, we consider that the TTTA dataset represents a challenging and realistic scenario, where atomic motion is naturally influenced by thermal fluctuations, allowing for the full range of structural degrees of freedom to contribute to data diversity.

5.1.2. Fundamental ingredients to learn chemical properties: chemical descriptors

In Quantum Chemistry (QC), the spatial arrangement of atoms, along with other parameters like atomic type, basis, charge, and spin multiplicity, serves as the foundational fingerprints to determine the properties of chemical species. This information is essential for Quantum Mechanics engines, such as HF or DFT, to accurately determine the electronic structure of a system, resulting in an accurate derivation of properties like total energy or J_{AB} (see Figure 5.3a).

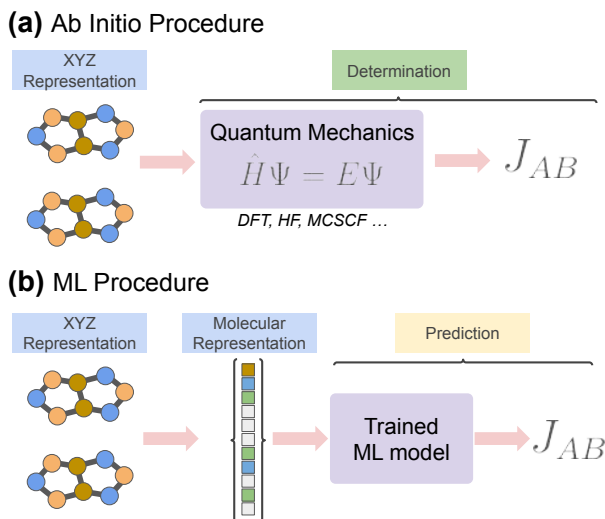


Figure 5.3: Comparative scheme of the determination of J_{AB} interactions: (a) Traditional Ab Initio methods employing quantum mechanics for direct computation of interaction values, and (b) Modern ML techniques utilizing a trained model for prediction of interaction values from molecular representations

However, the mere arrangement of atoms in space is not sufficient as input for ML models in chemistry. Instead, ML models require a more processed form of data, where chemical information is mapped, projected, or transformed into an invariant, fixed-size representation, typically in the form of a vector. Accordingly, each element embodies a specific attribute or feature of the chemical species, aligning better with the conditions in which ML algorithms operate with data (see Figure 5.3b, and further elaborated in **Chapter 2**). Overall, it is in this transformation where molecular descriptors play a crucial role. They act as a bridge, translating the complex information of a molecular structure into a uniform, fixed-size format that ML models can effectively learn from. However, unlike in other ML applications, molecules do not possess an intrinsic order, structure or size, which makes it difficult to elaborate a universal representation. As outlined in the work of Kermode³⁶ and co-workers, chemical descriptors must satisfy the largest amount of the conditions listed below to be considered appropriate:

- **Invariance:** descriptors should be invariant under symmetry operations, permutation of atoms, translation and rotations. This ensures that the descriptor accurately represents the inherent properties of molecules, independent of its orientation or spatial origin.
- **Sensitivity (local stability):** The descriptors should respond proportionally to minor changes in atomic positions, reflecting these alterations accurately. This sensitivity ensures that slight modifications in the structure are captured, enabling precise modeling and prediction.

- **Global Uniqueness:** Descriptors should represent a specific atomic environment, ensuring an injective mapping. This is critical for distinguishing between different molecular structures unambiguously.
- **Dimensionality:** The cardinality of the vector space spanned by the descriptors must be carefully balanced. It should be large enough to maintain uniqueness across different structures but not excessively larger.
- **Differentiability:** Descriptors should be derived from continuous functions that approximate linear behavior locally. This property ensures smooth transitions and gradients, facilitating more accurate computations in modeling and simulation.
- **Interpretability:** Ideally, the characteristics of the descriptor should correlate with structural or material properties, enabling straightforward interpretation of results. This interpretability bridges the gap between complex molecular data and tangible chemical properties.
- **Scalability:** Descriptors should be adaptable and generalizable to various systems or structures, without constraints on the number of elements, atoms, or properties involved. This scalability is essential for broad applicability in diverse chemical contexts.
- **Complexity:** The computational process to generate descriptors should be efficient, ensuring low computational complexity. This efficiency is crucial for enabling the descriptor's application in large-scale simulations and high-throughput screenings.
- **Discrete Mapping:** Regardless of the input atomic environment, descriptors should consistently map to a hyper-dimensional space of constant size. This consistency ensures that the feature set remains uniform across different molecules, simplifying comparative analyses.

In recent years, significant advancements have been made in developing reliable molecular descriptors satisfying most of the aforementioned principles that pursue a holistic view of representing molecules by considering various aspects of atomic arrangements and interactions. A notable category among these are descriptors based on three-dimensional structural information, often sustained under the principles of classical mechanics, such as electrostatic potentials or local density overlaps, which are thus usually referred to as Classical-Informed Representations (CIR). Examples of such descriptors include SOAP³⁷ (Smooth Overlap of Atomic Positions) and SLATM³⁸ (Spectrum of London and Axilrod-Teller-Muto). Other notable instances are the MBTR³⁹ (Many-Body Tensor Representation), BoB⁴⁰ (Bag of Bonds), and ACSF⁴¹ (Atom-Centered Symmetry Functions).

On the other hand, Quantum-Informed representations (QIR) are molecular descriptors that employ quantum mechanics principles for a deeper understanding of molecular systems. These descriptors, such as SPA^{HM}⁴² (Spectrum of Approximated Hamiltonian Matrices Representation), FJK⁴³ (Fock-Coulomb-Exchange), and MAOC⁴⁴ (Matrix of Orthogonalized Atomic Orbital Coefficients), capture intricate quantum mechanical properties of molecules. In short, SPA^{HM} represents molecules using the Hamiltonian matrix energy spectra, FJK combines Fock, Coulomb, and exchange matrices for detailed interatomic interactions, and MAOC focuses on the electronic structure via MO localization schemes.

Among the extensive list of molecular descriptors available, this work initially focused on two distinct types within the CIR category: BoB and SOAP. The selection of BoB is partly due to its status as one of the earliest appearing

descriptors in this category, marking a significant development in the field. Its simplicity and easy implementation further underscore its appeal. On the other hand, SOAP represents one of the most well-established descriptors in terms of the balance between computational complexity and performance, offering robust and reliable results. Its widespread availability, particularly in Python packages like Dscribe,⁴⁵ makes it also an accessible and practical choice. Accordingly, in the following subsections some brief details about their origin and mathematical formulation are provided.

5.1.2.1. Bag of Bonds

The Bag of Bonds⁴⁰ (BoB) is a variant of the Coulomb Matrix descriptor,⁴⁶ where instead of a matrix representation, atomic pairs are sorted into "bags" based on atomic types, with each bag containing the sorted Coulombic potentials of those atomic pairs. First, consider a molecule with N atoms, where each atom I is described by a nuclear charge Z_I and a position vector \mathbf{R}_I in the three-dimensional space. The elements M_{IJ} of the Coulomb matrix are defined by

$$M_{IJ} = \begin{cases} 0.5Z_I^{2.4} & \text{if } I = J, \\ \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} & \text{if } I \neq J \end{cases} \quad (5.1)$$

In the BoB representation, the $N(N - 1)/2$ unique off-diagonal elements of the Coulomb matrix are distributed into different "bags" corresponding to distinct pairs of atomic species. For example, all elements M_{IJ} with I corresponding to a carbon atom and J to a hydrogen atom are placed into the "C-H" bag. Then, for each bag, the elements are sorted in descending order of magnitude. The final BoB representation is the concatenation of these sorted bags. It should be noted that to handle molecules of different sizes, zero padding is often used to ensure

a consistent size of the feature vector across all molecules. Overall, BoB maintains a direct link to chemical intuition, as its elements can be associated with pairwise interactions between atoms. Importantly, this representation captures the invariance to atom indexing (unlike the Coulomb matrix).

5.1.2.2. Smooth Overlap of Atomic Positions

The Smooth Overlap of Atomic Positions^{37,45} (SOAP) descriptor represents a more sophisticated approach to encoding atomic geometries compared to BoB. It utilizes a local expansion of gaussian-smearred atomic density, employing orthonormal functions based on spherical harmonics and radial basis functions that guarantee continuity and differentiability. Essentially, the SOAP representation corresponds to a partial power spectrum, \mathbf{p} , where its components ($\rho_{nn'l}^{Z_A Z_B}$) are defined as follows:

$$\rho_{nn'l}^{Z_A Z_B}(\mathbf{r}) = \pi \sqrt{\frac{8}{2l+1}} \sum_m c_{nlm}^{Z_A}(\mathbf{r}) c_{n'lm}^{Z_B}(\mathbf{r}) \quad (5.2)$$

In this formulation, n and n' denote indices for different radial basis functions, extending up to a preset maximum, n_{\max} , specified by the user. Likewise, l represents the angular degree of the spherical harmonics, truncated at l_{\max} . Moreover, each component of the power spectrum can be attributed to a pair of chemical species, A and B , indicated in the previous expression as Z_A and Z_B , respectively. Note that the prefactor simply corresponds to a normalization value with respect to the angular component, l , which is shared by n and n' values.

Central to the SOAP representation are the calculation of the coefficients $c_{nlm}^{Z_I}(\mathbf{r})$, which are defined as the following inner products

$$c_{nlm}^{Z_A}(\mathbf{r}) = \iiint_{\mathcal{R}^3} g_n(r) Y_{lm}(\theta, \phi) p^{Z_I}(\mathbf{r}) dV \quad (5.3)$$

where $g_n(r)Y_{lm}(\theta, \phi)$ represents the basis set in which the Hilbert space is spanned, typically a combination of a radial basis function, $g_n(r)$, and real spherical harmonics, $Y_{lm}(\theta, \phi)$. This choice is not only for a parallelism with quantum chemistry, but also for a cheap evaluation of the inner products by means of analytical or numerical recipes. In addition, $p^{Z_I}(\mathbf{r})$ is referred to as the gaussian smoothed density for atoms with atomic number Z_I , which consists on the aggregation of the dampened atomic positions as:

$$p^{Z_I}(\mathbf{r}) = \sum_K e^{-\frac{1}{2\sigma^2}|\mathbf{r}-\mathbf{R}_K|^2} \quad (5.4)$$

where the summation ranges all the atoms K atoms with atomic number Z_I . Moreover, \mathbf{R}_K corresponds to the position of the K -th nuclei within the set, and σ is the standard deviation of the gaussian functions used to expand the atomic density.

By integrating these elements, SOAP coefficients project the spatial positions (smoothed by Gaussian functions) onto a basis set comprising products of isotropic radial functions and anisotropic angular ones. The use of a varied collection of n and l indices allows these projections to be captured in different many-body components of the power spectrum, effectively decorrelating their contributions and enriching the representation of molecular geometries.

As outlined in the preceding definitions, the utilization of the SOAP descriptor requires the user to pre-define three key parameters (among others), i.e., n_{\max} , l_{\max} and σ . While the optimal values for these parameters can vary depending

on the specific application, we chose $n_{\max} = 6$, $l_{\max} = 4$ for our applications, and conducted a performance assessment of SOAP across different values of σ (0.1, 1, 2, 3 and 10). Moreover, it is crucial to recognize that SOAP can be computed at arbitrary spatial positions, denoted as \mathbf{r} in the previous explanation, thus, classifying SOAP fundamentally as a local descriptor. Typically, these evaluations are carried out at each atomic center within a molecule. However, considering that J_{AB} corresponds to a global property, our application of SOAP required an averaging process across all atomic centers to derive a global representation. Specifically, the SOAP components utilized in our study have been defined as follows:

$$\rho_{nn'l}^{Z_A Z_B} = \frac{1}{n_{\text{at.}}} \sum_{s=0}^{n_{\text{at.}}} \rho_{nn'l}^{Z_A Z_B}(\mathbf{r}_s) \quad (5.5)$$

This approach is often referred to as outer averaged (or structural averaged) SOAP, which ensures that the local representations obtained from individual atomic centers, $\rho_{nn'l}^{Z_A Z_B}(\mathbf{r}_s)$, are effectively integrated together to reflect the global characteristics of the molecules, aligning with the intended application of SOAP in assessing J_{AB} .

5.2. Results and discussion

5.2.1. Preliminary results and model's transferability

In this section, we present the results from the application of BoB and SOAP descriptors to predict J_{AB} values for the TTTA dimer dataset. The initial experiment involved training a Kernel Ridge Regression (KRR), using a Radial Basis Function (RBF) kernel, as outlined in the Methodology chapter (**Chapter 2**). The model was trained and tested by means of a fraction of the TTTA dataset, comprising 10,000 configurations of a TTTA dimer and their associated

J_{AB} values. Moreover, we adhered to standard machine learning protocols, which included the following steps:

- **Feature Computation:** We computed BoB and SOAP features for all data samples.
- **Data Shuffling:** The consecutive, time-resolved AIMD data was shuffled to ensure randomness and decorrelation during the training process.
- **Dataset Splitting:** We divided the dataset into training and testing subsets, using varying proportions of the data (ranging from 5% to 50%) to optimize the hyperparameters γ (from the RBF kernel) and λ (from the ridge regularization term), see more details in the Methodology chapter (**Chapter 2**).
- **Model Evaluation:** The remaining data (test subset) was used to evaluate the prediction accuracy, specifically using the Mean Absolute Error (MAE) metric.

This methodology enabled us to generate a learning curve, providing insights into the performance of the KRR model and the effectiveness of BoB and SOAP descriptors in representing TTTA structures for the task of predicting J_{AB} values. In this regard, Figure 5.4a displays the learning curve derived from applying this methodology to the first dimer (D1) of the first column (C1) in the dataset generated by AIMD simulation at 300 K (denoted as HT-300K-D1C1), as it can be identified in Figure 5.2a,b.

The results revealed an expected trend: with a smaller training data size, specifically between 5% and 20%, the MAE decreases from approximately 40 to 10 cm^{-1} for predictions on the remaining test set. This trend indicates an im-

proving generalization capability of the model for J_{AB} predictions as the training data size increases. Notably, the learning curve reaches a plateau as the training data size ranges between 30% and 50%, with the MAE stabilizing below 5 cm^{-1} . This plateau indicates an optimal data size range for training, beyond which additional data does not significantly enhance the predictive capabilities of the KRR model. Note that the results depicted in Figure 5.4 belong to those obtained from the experiments utilizing the BoB descriptor. The analogous outcomes related to the SOAP descriptor were found to be very similar to those of BoB. Thus, for the sake of simplicity and to avoid redundancy, we have chosen not to include the SOAP results in the present figure.

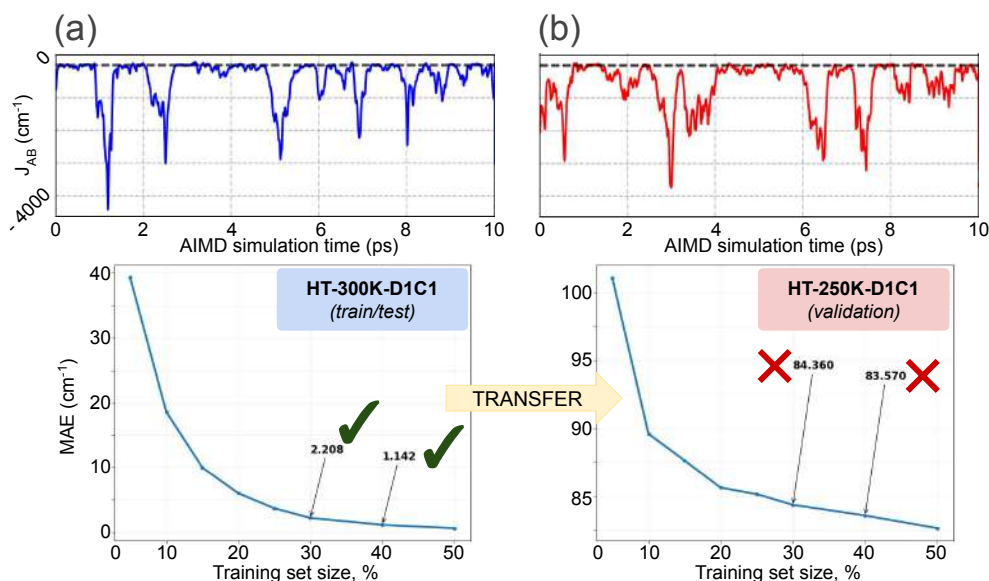


Figure 5.4: Time-resolved evolution of J_{AB} (top) and learning curve (bottom) for (a) first dimer of the first column of the AIMD simulation performed at 300 K (HT-300K-D1C1, blue) and (b) the same dimer at the simulation conducted at 250 K (HT-250K-D1C1, red). Note that the learning curve illustrate the Mean Absolute Error (MAE, in cm^{-1}) as a function of the training size (in % of data size). The data illustrated at (a) has been used to train and test the KRR model, while the data of (b) has been employed to validate the model and assess its transferability. The results collected correspond to those using the BoB descriptor.

Overall, our initial experiment demonstrated promising results, suggesting that both BoB and SOAP are capable of accurately *inform* the KRR to subsequently establish sample–sample correlations and predict J_{AB} values. The observed error margin of 5 cm^{-1} translates to a relative error of approximately 1.5% in relation to the mean values of the J_{AB} distribution ($\langle J_{AB} \rangle \approx 350\text{ cm}^{-1}$), underscoring the effectiveness of BoB and SOAP for the task in hand.

Building upon this apparent success, we extended our analysis to include J_{AB} values derived from TTTA data obtained from the AIMD simulations conducted at a lower temperature of 250 K, referred to as HT-250K-D1C1 in Figure 5.4b. This additional dataset served as a validation set, assessing the capacity of the pre-trained KRR mode for both interpolation and extrapolation. Intriguingly, while the learning curve exhibited a similar descending pattern to the initial experiment, a substantial shift can be observed for the MAE values. As the training set size increased from 5% to 50%, the MAE escalated to notoriously unacceptable levels, ranging from 100 to 80 cm^{-1} . This provides evidence of a significant issue of overfitting in the KRR model, suggesting that, while the model performed well on the training and highly similar testing sets, its ability to generalize to new, unseen data is clearly limited.

In view of the unsatisfactory results, our research was then focused on solving the issues related to the clear lack of HT-300K \rightarrow HT-250K model’s transferability. This involved evaluating the performance of alternative descriptors such as MBTR³⁹ and ACSF,⁴¹ conducted together with a thorough review of the DFT calculations used for extracting J_{AB} values. To refine our approach further, we implemented advanced hyper-parameter optimization techniques, such as meta-learning strategies like Bayesian Optimization Search^{47,48} (BOS) or cross-

validation schemes like the leave-p-groups-out cross-validation⁴⁹ (LPGO-CV), as detailed in **Chapter 2**. Additionally, we explored a variety of ML and Deep Learning models, including online versions of KRR,⁵⁰ Gaussian Process Regression⁵¹ (GPR), Random Forest⁵² (RF), and Deep Neural Networks⁵³ (NN). Finally, we also incorporated a pre-processing step consisting of dimensionality reduction techniques by means of Principal Component Analysis⁵⁴ (PCA), Kernel PCA,⁵⁵ and Neural Network-based Autoencoders.⁵⁶

Despite these extensive efforts, the outcomes remained unsatisfactory, whether these techniques were applied individually or, when possible, in combination. This persistent unsuccessful results led us to conclude that the problem was not directly tied to the ML model, the DFT data, the learning strategy, the descriptor, or its dimensionality. Instead, it appeared to be an inherent issue within our approach in learning *intermolecular* J_{AB} values by means of descriptors that are rather prepared to excel in representing *intramolecular* interactions.^{34,57} The reason is that, for the sake of better sensitivity, the majority of the descriptors emphasize the magnitude of short-range interactions to better capture the local atomic environment. As a result, these representations can underperform in cases where medium- or long-range (possibly intermolecular) interactions are crucial.^{58,59}

In order to provide a tangible grasp on this, consider a hypothetical descriptor, $\rho(\mathbf{d})$, that operates by mapping all the bond distances (\mathbf{d}) of a given molecular system as $\rho(\mathbf{d}) = \sum_i 1/d_i$, as depicted in blue in Figure 5.5. Accordingly, the descriptor displays a pronounced sensitivity to small perturbations, Δd , of short intramolecular distances, designated as d_1 and shown in red. In contrast, the same descriptor exhibits significantly reduced responsiveness for intermolecular

interactions (labeled as d_2 and colored in green) even with an equivalent variation of Δd . Overall, when the contribution from d_1 and d_2 are added, the net response of the descriptor is clearly dominated by the amplified d_1 contributions, masking the effect of d_2 . This visual representation serves to underscore the inherent limitations of descriptors, such as BoB or SOAP, in scenarios where the properties under examination are fundamentally driven by intermolecular interactions. These limitations in the standard versions of the descriptors can be addressed by decoupling intra- and intermolecular interactions in the representation,^{33,57} as demonstrated in the following sections.

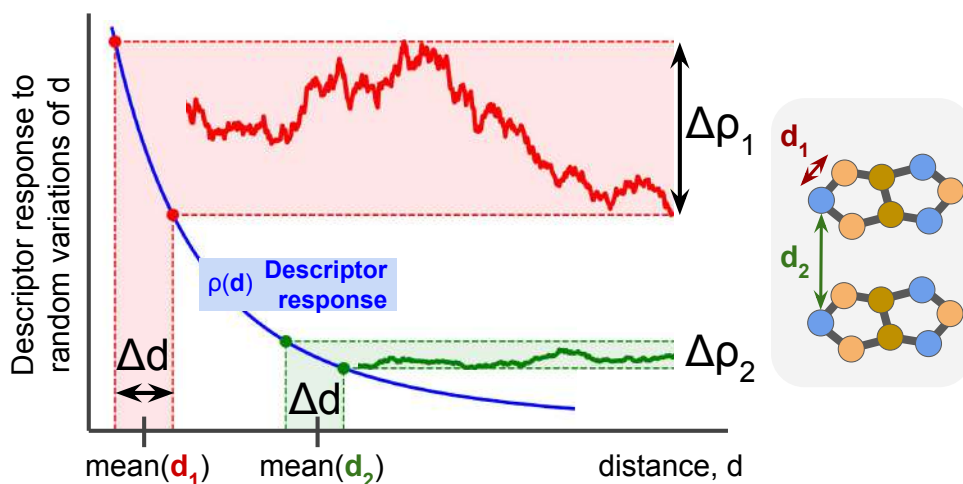


Figure 5.5: Schematic representation of the usual sensitivity of the descriptor in short range (red) and medium or long range (green). Note that the descriptor response (illustrated as $\rho(\mathbf{d})$, blue) to a variation of distance, Δd , for an intermolecular distance (red), d_1 , is several times more sensitive than the same magnitude of variation for intermolecular distances (green), d_2 .

5.2.2. Decoupling of intra- and intermolecular components

In this subsection, we introduce the formalism to separate intra- and intermolecular components in BoB and SOAP representations. We also outline an automated method for detecting molecules in multi-moiety systems, which has

been crucial for efficiently processing large datasets like TTTA. Subsequently, we detail the necessary modifications for integrating this decoupling approach into Kernel-based models, such as KRR. Finally, we present the results from applying this scheme to the TTTA dataset, focusing only on the SOAP descriptor for convenience.

5.2.2.1. Decoupling of BoB

In the BoB representation, a decoupling scheme can be straightforwardly implemented from the discussion structured around equation 5.1. Accordingly, the separation of intra- and intermolecular components can be accomplished by partitioning each "bag" into two distinct sets of the same atomic pair types, like C-H. This division is achieved by analyzing the atomic pairs contributing to the bag. If both atoms, C and H, are in the same molecule, their interaction is classified as an intramolecular component, thus allocated in the intramolecular C-H bag. Conversely, if they are in different molecules, the interaction is allocated in a separate intermolecular bag. Hence, repeating this process for every atomic pair, a double set of bags can be created, which can be subsequently arranged in descending order of magnitude, as in the non-decoupled BoB descriptor.

5.2.2.2. Decoupling of SOAP

Unlike the BoB representation, SOAP encapsulates 3-body terms,⁵¹ making the intra/inter decoupling strategy used for BoB (which stem from 2-body interactions) inappropriate. Consequently, we adopted a decoupling approach for SOAP, heavily influenced by the one proposed by Cersonsky⁶⁰ and co-workers. In this regard, consider x_a and x_b as separate sub-systems within a dimer, X (note that x_i is a shortcut notation to identify the abstract objects correspond-

ing to each system, which can include coordinates, species, etc.). While their respective SOAP representations, ρ_{x_a} and ρ_{x_b} , capture intramolecular details, the representation of the dimer, ρ_X , includes both intra- and intermolecular information, which are not straightforwardly divisible. Notably, the intramolecular components of the dimer, ρ_X^I , can be calculated as the sum of the components from x_a and x_b :

$$\rho_X^I = \rho_{x_a} + \rho_{x_b} \quad (5.6)$$

Subsequently, the intermolecular components, referred to as ρ_X^i , can be derived by subtracting the intramolecular components from the dimer representation:

$$\rho_X^i = \rho_X - \rho_X^I \quad (5.7)$$

Overall, this process effectively separates the components while maintaining the 3-body nature of the SOAP encoding at the expense of calculating three times the SOAP representation for each chemical system.

5.2.2.3. Automatic detection of moieties in chemical systems

It is essential to emphasize that the decoupling formalism for BoB and SOAP relies on identifying isolated moieties within a dimer. Thus, the decoupling algorithm must accurately discern interactions between atom pairs. Given the large size of datasets (e.g., TTTA), which typically contain hundreds to thousands of samples, automating this process is thus critical. In this study, we investigated two graph theory-based approaches to meet this challenge.

In this regard, a graph $\mathcal{G} = (V, E)$ can be represented by vertices V and edges E , akin to atoms and bonds in molecules to some extent. By setting a

cutoff distance for atomic distances, these can be mapped to \mathcal{G} as connected or disconnected. Through an iterative processing, the graph adjacency matrix, A , can be constructed, where each element a_{IJ} is either 1 or 0, reflecting the presence or absence of a connection. Building upon these concepts, graph theory provides robust techniques for detecting sub-graphs (or moieties) within a graph (or dimer), with the matrix A being a key component. Among various methods, depth-first search⁶¹ and spectral clustering-based⁶² (SC) techniques are popular choices. While both approaches were implemented during this study, the SC-based algorithm proved superior, offering greater efficiency, reduced computation times, and more consistent results, especially for graph sizes equivalent to the atom counts in TTTA dimers.

In order to illustrate the performance of the SC-based strategy, consider the Laplacian matrix, L , which is derived as $L = D - A$, where D is the degree matrix and A is the adjacency matrix. Notably, the matrix D is diagonal and is easily computed from A , where each diagonal element (d_{II}) represents the number of edges connected to vertex v_I (*i.e.*, $d_{II} = \text{deg}(v_I)$). Now, the eigen-decomposition of L is expressed as:

$$L = U\mathcal{L}U^\dagger \quad (5.8)$$

where U comprises the eigenvectors of L in its columns, and \mathcal{L} is a diagonal matrix containing the eigenvalues, λ_I . Spectral clustering theory reveals that the number of eigenvalues equaling zero, $\lambda_K = 0$, corresponds to the number of sub-graphs (or moieties) in the entire graph. Moreover, as L is positive semidefinite, it must satisfy that $\lambda_I \geq 0$ and, thus, the process of selecting those with $\lambda_K = 0$ reduces to collect the lowest degenerate eigenvalues. Subsequently, the nodes (or atoms) of each moiety can be identified by examining the eigenvectors, u_K , linked

with $\lambda_K = 0$. Interestingly, the components of these eigenvectors are found to display the following pattern:

$$u_{KJ} = \begin{cases} \frac{1}{\sqrt{N_I}} & \text{if } v_K \in \mathcal{G}_I, \\ 0 & \text{otherwise} \end{cases} \quad (5.9)$$

or, in words, the J -th node (or atom) of u_K is $1/\sqrt{N_I}$ if vertex v_K is part of the I -th sub-graph (\mathcal{G}_I), and 0 otherwise. In this equation, N_I denotes the total number of nodes in the I -th sub-graph. Building upon that, the algorithm to detect moieties from a multi-moiety system consists of the following steps:

- Build the adjacency matrix A based on the distances of all atoms with the rest and a threshold value serving as a distance cutoff (2.0 Å in our case).
- Calculate the Laplacian matrix L as $L = D - A$, where D is easily obtained from A .
- Diagonalize L to obtain the eigenvectors u_K and their associated eigenvalues λ_K .
- Identify the number of eigenvalues that satisfy $\lambda_K = 0$ and extract their associated eigenvectors.
- Group together all the nodes with non-zero components belonging to each sub-graph, \mathcal{G}_I , based on the values of their associated eigenvectors.

5.2.2.4. Modification of the RBF kernel for decoupled descriptors

In this subsection we present the formalism with which the previous decoupling schemes, outlined for BoB and SOAP, can be applied within the usual KRR learning ensembles. The usual procedure simply consists of ignoring the intramolecular contributions in favor of the intermolecular components.^{33,57} How-

ever, neglecting the intramolecular components might be a rather severe simplification, since these can be correlated to the intermolecular property to some extent. For this reason, we derived a mathematical formalism able to accommodate fine-tuned contributions of intra- and intermolecular components in kernel-based ML models such as KRR, specially focused in the RBF kernel employed herein.

As discussed in the Methodology chapter (**Chapter 2**), the RBF kernel is formally defined as:

$$\kappa(\rho_1, \rho_2; \gamma) = e^{-\gamma \|\rho_1 - \rho_2\|^2} \quad (5.10)$$

where, ρ_1 and ρ_2 represent two data samples characterized by an arbitrary descriptor, with γ being the hyperparameter that influences the decay of the RBF kernel, and the symbol $\|\cdot\|$ denotes the Euclidean distance.

While the standard RBF definition is adequate for the direct applications of chemical descriptors, the complexity of multi-moiety systems elaborated above requires a more tailored approach. In this context, the RBF kernel can be adapted to address intra- and intermolecular components separately. In this regard, when the representation ρ can be divided into intramolecular (ρ^I) and intermolecular (ρ^i) components, as demonstrated with BoB and SOAP, the Euclidean distance can be split according to the following decomposition:

$$\|\rho_1 - \rho_2\|^2 = \|\rho_1^I - \rho_2^I\|^2 + \|\rho_1^i - \rho_2^i\|^2 \quad (5.11)$$

which allows to separately identify the intra- and intermolecular contributions to the total distance. It thus follows that the RBF kernel can be adapted to the former decomposition as a product of its intra- and intermolecular contributions by means of the following operations:

$$\begin{aligned}\kappa(\rho_1, \rho_2) &= e^{-\gamma^I \|\rho_1^I - \rho_2^I\|^2} e^{-\gamma^i \|\rho_1^i - \rho_2^i\|^2} \\ &= \kappa(\rho_1^I, \rho_2^I; \gamma^I) \kappa(\rho_1^i, \rho_2^i; \gamma^i)\end{aligned}\tag{5.12}$$

Here, γ is substituted by two hyperparameters, γ^I and γ^i , each controlling the decay of the RBF kernel for the intramolecular and intermolecular components, respectively.

Overall, employing distinct hyperparameters for intra- and intermolecular components allows a more refined modeling of molecular interactions in multi-moiety systems. This approach enables the ML model to independently adjust the influences of γ^I and γ^i . However, it is important to balance the benefits of this added flexibility against the challenges it introduces in training the model: a larger amount of hyperparameters not only increase the complexity of model fitting, but also increases the risk of overfitting. In this situation, the risk is deemed justifiable, particularly since intermolecular interactions are proved to be masked otherwise. Additionally, it is worth noting that the formalism developed here is not limited to the RBF kernel. It can be adapted for use with other types of kernels, such as the Laplacian kernel or those based on cosine similarity, among others.

5.2.2.5. Analysis of decoupled SOAP in TTTA dataset

In this subsection, we inspect the results of decoupling SOAP features in the initial picoseconds of the AIMD simulation for the HT-300K-D1C1 dimer of TTTA. In this regard, Figure 5.6a illustrates the intramolecular components ($\text{SOAP}_{\text{intra}}$), with each curve representing the evolution of a specific feature over the time interval. Interestingly, the time-resolved J_{AB} evolution, depicted in

Figure 5.6b, shows no significant correlation with $\text{SOAP}_{\text{intra}}$. This observation can be attributed to thermally-activated bond vibrations that, while not greatly affecting the J_{AB} determination, predominantly introduce noise into the ML model. Conversely, the intermolecular components ($\text{SOAP}_{\text{inter}}$), as demonstrated in Figure 5.6c, visually exhibit a greater correlation with J_{AB} .

This finding underscores the effectiveness of the decoupling scheme and uncover the source of the overfitting issues encountered in our initial experiment. Specifically, in the first experiment involving a fraction of the TTTA dataset (i.e. 10,000 structures of HT-300K-D1C1), the KRR model was inadvertently trained to interpolate random noise resulting from the unrefined application of SOAP and BoB descriptors, where the intramolecular components dominate. This led to a model that, although (surprisingly) effective in fitting the training data, was essentially overfitting by learning from the noise. When this KRR model was applied to predict J_{AB} values for the HT-250K-D1C1 dimer, the limitations of its capabilities inevitably showed up. The absence of correlations that the model had been forced to learn resulted in poor generalization when validated with new unseen data.

However, while a previous examination of $\text{SOAP}_{\text{intra}}$ and $\text{SOAP}_{\text{inter}}$ components readily suggests that $\text{SOAP}_{\text{inter}}$ is a more adequate representation for predicting intermolecular J_{AB} values, a closer look at the evolution of $\text{SOAP}_{\text{inter}}$ responses in certain regions manifests some notable miss-alignments with respect to the evolution of J_{AB} . For instance, this lack of correlation becomes apparent when examining the descriptor response ($\Delta\rho$) to changes in J_{AB} over the AIMD simulation time (t): when TTTA dimers undergo abrupt structural changes that shift their associated J_{AB} from paramagnetic to strongly AFM regimes (large

$\Delta_{12}J$, t_1 vs. t_2 in Figure 5.6b) SOAP fails to accurately represent these variations (small $\Delta_{12}\rho_{\text{SOAP}}$, Figure 5.6c), causing completely opposite structures to appear incorrectly similar. Likewise, J_{AB} values at $t = t_2$ and $t = t_3$ are almost identical in both cases, but SOAP indicates an abrupt change ($\Delta_{23}\rho_{\text{SOAP}}$), even larger than the change associated with the extremely contrasting regimes at $t = t_1$ and $t = t_2$. The primary explanation for this discrepancy is illustrated

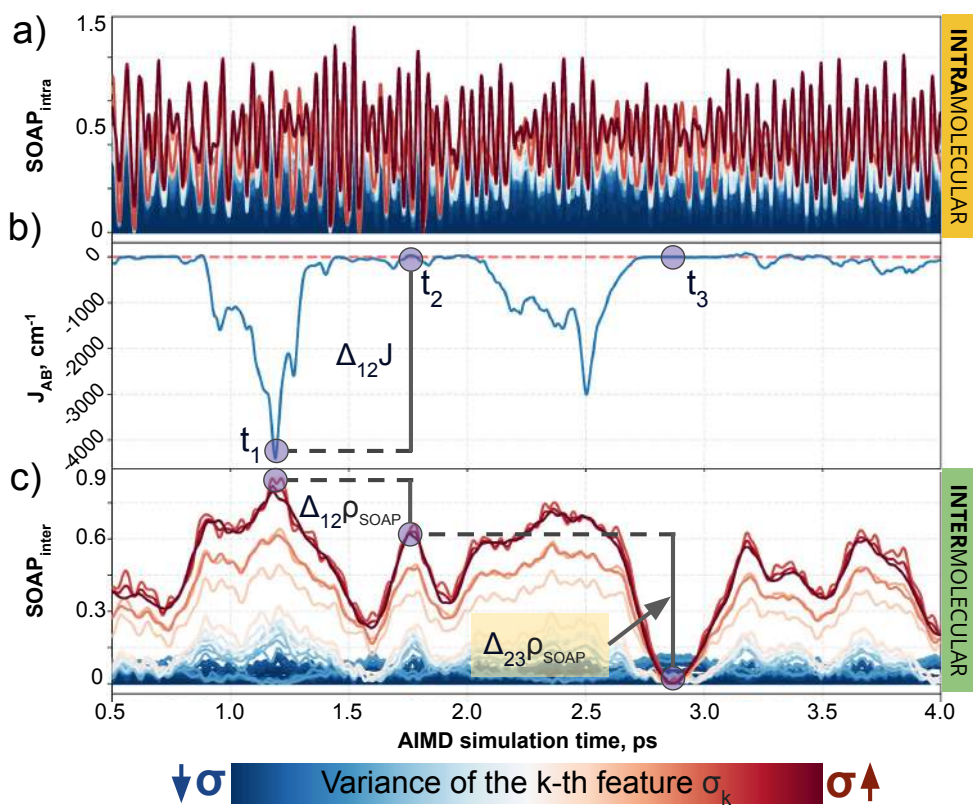


Figure 5.6: Evolution of (a) intra- and (c) intermolecular SOAP features (colored by variance value) of a TTTA dimer along the structures explored during the 0.5–4.0 ps time interval of AIMD, as well as (b) the associated time-resolved evolution of J_{AB} in this interval. The blue circles in (b) and (c) indicate three regions with specific values of J_{AB} , where one can compare the change associated with J_{AB} and SOAP.

in Figure 5.7, which displays the structure and SONOs of the TTTA dimer at $t = t_2$ and $t = t_3$, see d_{AB} in the figure. Accordingly, the TTTA moieties at $t = t_2$ are notably closer than at $t = t_3$. Nevertheless, the overlap between the SONOs is nearly zero in both cases (compare to the shared isosurface between TTTA SONOs of $t = t_1$ in Figure 5.7), resulting in a similar J_{AB} .⁶³ Hence, this situation is particularly ill-defined for SOAP, which fails to capture the subtleties rooted in the MO topology due to its lack of electronic structure information, potentially transferring this miss-information to KRR models trained upon it.

This analysis underscores the inherent limitations of CIR descriptors, such as BoB and SOAP, in differentiating TTTA dimer conformations with respect to J_{AB} . Revisiting the classification of descriptors from earlier discussions, it becomes evident that transitioning from CIR to QIR descriptors might be a correct strategic move in this situation. Essentially, the primary advantage of QIR compared to CIR-type descriptors lies in its capability to encode the electronic state of a system, covering aspects such as electronic structure, charge, and spin multiplicity. This might suggest that QIR can track the nuances emerging from the SONOs of TTTA, thus properly representing the intricacies deeply rooted in the electronic structure of open-shell dimers, like J_{AB} , where CIR-type descriptors fall short.

As shown with BoB and SOAP, the possibility of decoupling atomic interactions depends on the initial approach to encode molecular information. For instance, BoB relies on 2-body Coulomb potentials, which permits its decoupling based on distance or connectivity criteria. However, none of the existing QIR-type descriptors are compatible with decoupling schemes, akin to those used for BoB and SOAP. For example, QIR descriptors such as SPA^HM and MAOC,

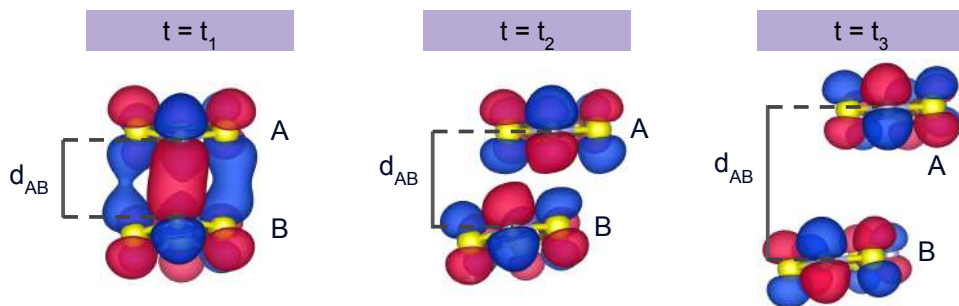


Figure 5.7: Singly-Occupied Natural Orbitals (SONOs) of the HT-300K-D1C1 dimer of TTTA at three different configurations: $t = t_1$, $t = t_2$, and $t = t_3$ in agreement with the notation of Figure 5.6.

based on eigen-decomposition of the Hamiltonian, encompass many-body interactions, potentially resulting from contributions of atoms in different moieties. Consequently, the development of methods that combine the advantages of QIR descriptors for fine-grained representations and the separability of some CIR are sought. Such advancements would pave the way for more precise predictions of intermolecular properties, leveraging the strengths of QIR while ensuring its practical applicability.

Herein, we introduce the *Molecular Orbital Decomposition and Aggregation* (MODA) as a new QIR descriptor, the first of its kind that allows decoupling strategies. The mathematical foundation of MODA, along with its unique capability to support decoupling schemes, are thoroughly explored in the following section.

5.2.3. Derivation of Molecular Orbital Decomposition and Aggregation

The derivation of MODA is visually illustrated in Figure 5.8. The general idea stems from extracting quantitative measurements of pairwise interatomic interactions from a selected subset of MOs. Beginning with the atomic coordinates of

the molecule (see Figure 5.8, step 1), we progress to calculate the MOs, the core mathematical object to assemble the representation (see Figure 5.8, step 2). As usually, those can be expanded as a linear combination of atomic orbitals (AOs) like:

$$|\psi_k\rangle = \sum_{\forall i} c_{ik} |\chi_i\rangle \quad (5.13)$$

where c_{ik} is the coefficient for the atomic orbital $|\chi_i\rangle$ within the MO $|\psi_k\rangle$. This representation enables the explicit formulation of the one-particle reduced density matrix, D , in the AO basis, defined as the sum of outer products of the MO coefficient vectors:

$$D = \sum_{\forall k} n_k |c_k\rangle \langle c_k| = \sum_{\forall k} D_k \quad (5.14)$$

where n_k corresponds to the occupation number of the k -th MO and D_k corresponds to the partial density matrices associated to each MO. The density matrix is fundamental in calculating the bond orders, which is a chemically-intuitive metric of strength about the interactions between pairs of atoms. Among the different approaches to quantify the bond order, the Mayer's definition⁶⁴ starts by defining the population matrix $P = DS$, where S corresponds to the overlap matrix, also in the AO basis. Now, the bond order (\mathcal{B}_{AB}) between two atoms, A and B , is extracted by the trace operation over the population matrix P :

$$\mathcal{B}_{AB} = Tr(P_{AB}P_{BA}) \quad (5.15)$$

where $P_{AB} = P_{BA}^T$ are sub-matrices of P that include contributions from AOs centered on atom A and B . Exploiting the decomposition of the density matrix into orbital contributions, as stated in equation 5.14, the bond order can be further fragmented as a sum of individual contributions from each MO, $|\psi_k\rangle$ as

$$\mathcal{B}_{AB} = \sum_{\forall k} \text{Tr} (P_{AB}^k P_{BA}^k) = \sum_{\forall k} \mathcal{B}_{AB}^k \quad (5.16)$$

In this expression, \mathcal{B}_{AB}^k signifies the partial bond order contribution, reflecting the distinct role of each orbital in the bond order (see Figure 5.8, steps 3 and 4). Building upon this approach, the contributions to the partial bond can be decomposed even further. Specifically, we can examine the contributions to the bond order stemming from the different atomic orbitals of $|\psi_k\rangle$, characterized by their principal (n), azimuthal (l) and magnetic (m) quantum numbers. To this end, the orbital-specific partial bond order, here represented as $\mathcal{B}_{AB}^{k,nn'l}$, can be calculated following the same principles applied so far. In this notation n and n' correspond to the principal quantum number of atoms A or B , respectively.

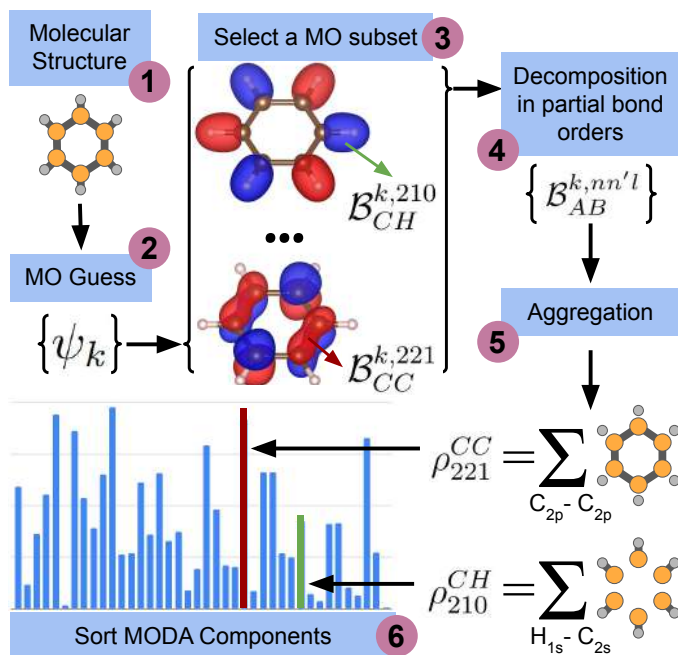


Figure 5.8: Scheme of the workflow to compute the features of MODA. The figure shows the process that produces MODA features from the molecular structure (step 1) to the final representation (step 6).

It is within this context that we aggregate these contributions to construct the components of the MODA descriptor. Formally, MODA consists of the aggregation of such orbital-specific partial bond orders in many-body batches to guarantee the translational, permutational and rotational invariance of the representation (see step 5 in Figure 5.8). Accordingly, the MODA components, $\{\rho_{nn'l}^{Z_A Z_B}\}$, can be defined as:

$$\rho_{nn'l}^{Z_A Z_B} = \sum_{\forall k} \sum_{A \in Z_A} \sum_{B \in Z_B} \mathcal{B}_{AB}^{k,nn'l} \quad (5.17)$$

here A and B refer to all possible atoms in a given molecule with atomic numbers Z_A and Z_B , respectively. That is, each term in MODA aggregates the partial bond orders of all A - B interactions coming from specific atomic orbitals identified by n , n' and l quantum numbers, allowing all possible values of the magnetic quantum number, m (see step 5 in Figure 5.8) for rotational invariance. Our choice of using one azimuthal quantum number while incorporating two principal quantum numbers is inspired by the level of detail present in the standard SOAP representation. Nevertheless, alternative versions of MODA could delve deeper into distinguishing components by incorporating another quantum number, l' . This would allow to differentiate, for example, σ - σ from σ - π interactions. However, this extra level of detail would also come at the cost of enlarging the representation size, which we believe to be unnecessary for TTTA, where all the contributions to the SONOs come from π -MOs.

Although equation 5.17 collects the contribution of all molecular orbitals (i.e., the range of k in the outermost summation), MODA can be tuned to include only a subset, \mathcal{S} , of orbitals, denoted as $k \in \mathcal{S}$. This is beneficial to avoid the computation of bond orders between orbitals with negligible contributions (*e.g.*,

core orbitals or unoccupied orbitals), or when the property under examination arises primarily from a *specific* group of molecular orbitals. In this regard, in what follows only SONOs have been selected for further experiments, as those are the precise elements from which J_{AB} stem.

Moreover, like in other QIR descriptors such as SPA^HM and MAOC, MODA does not require the explicit calculation of the electronic structure through the usual computationally-taxing self-consistent field (SCF) methods. Instead, MODA representation can be constructed using well-established "guess" Hamiltonians,⁶⁵ such as the Superposition of Atomic Densities (SAD),⁶⁶ the Superposition of Atomic Potentials (SAP),⁶⁷ or the extended Hückel method⁶⁸ (EH), which are the usual starting points for SCF methods. The use of these "guess" Hamiltonians provides a computationally-light, yet powerful and simple platform to accelerate the calculation in QIR descriptors. In line with other works in this field,⁴² our implementation of MODA employs the SAD guess, which is a good compromise between computational efficiency and quality of the approximate electronic structure. Finally, MODA components can be calculated in any AO basis set, spin multiplicity or oxidation state. Note that, resembling the impact of basis sets in quantum chemistry computations, larger basis sets lead to better MODA representations and, thus, potentially to better predictions. It is thus reasonable to consider that both the level of theory (SAD, SAP, EH, etc.) and the basis set (e.g. STO-3G, 6-31G, cc-aug-pvdz, etc.) are hyperparameters of the MODA representation.

5.2.4. Intra-/intermolecular decoupling strategy in MODA

As MODA batches result from the aggregation of pairwise interactions, it is compatible with intra-/intermolecular decoupling schemes. This capability can

be incorporated directly into the formalism through a minor modification of equation 5.17. In this regard, let δ_{AB} be a Kronecker delta parameter, taking the value of 1 when atoms A and B belong to the same molecule, and 0 otherwise. Now, MODA components ascribed to intra- and intermolecular interactions (with super-index I and i , respectively) can be defined as

$$\begin{aligned}\rho_{nn'l}^{Z_A Z_B, I} &= \sum_{\forall k} \sum_{A \in Z_A} \sum_{B \in Z_B} \mathcal{B}_{AB}^{k, nn'l} \delta_{AB} \\ \rho_{nn'l}^{Z_A Z_B, i} &= \sum_{\forall k} \sum_{A \in Z_A} \sum_{B \in Z_B} \mathcal{B}_{AB}^{k, nn'l} (1 - \delta_{AB})\end{aligned}\tag{5.18}$$

That is, the intramolecular components only add up the bond orders resulting from the same molecular entity. Likewise, intermolecular components aggregate those not previously classified as intramolecular terms.

5.2.5. Comparison of decoupled CIR and QIR descriptors

In this subsection, we examine the performance of MODA with respect to established CIR descriptors like BoB and SOAP. To facilitate a straightforward comparison, we only display the intermolecular MODA features of the HT-300K-D1C1 TTTA dimer in Figure 5.9, across an identical time interval used for the SOAP analysis in Figure 5.6. Notably, MODA features demonstrate a more pronounced correlation with the exchange coupling interaction J_{AB} than the corresponding SOAP intermolecular components. This is particularly evident when we inspect the sensitivity of the descriptor ($\Delta\rho$) to shifts in J_{AB} across the same time-resolved structural configurations previously examined in SOAP.

For instance, when the TTTA dimers experience sudden configurational distortions that alter J_{AB} from a paramagnetic to an antiferromagnetic (AFM)

coupling (noted as a significant $\Delta_{12}J$ between time steps t_1 and t_2 in Figure 5.9b), MODA adequately reflects these variations (observed as a substantial $\Delta_{12}\rho$ in Figure 5.9a). Furthermore, MODA consistently reflects the small J_{AB} variations between times t_2 and t_3 , demonstrating a reliable descriptor response where SOAP has been previously found to fall short.

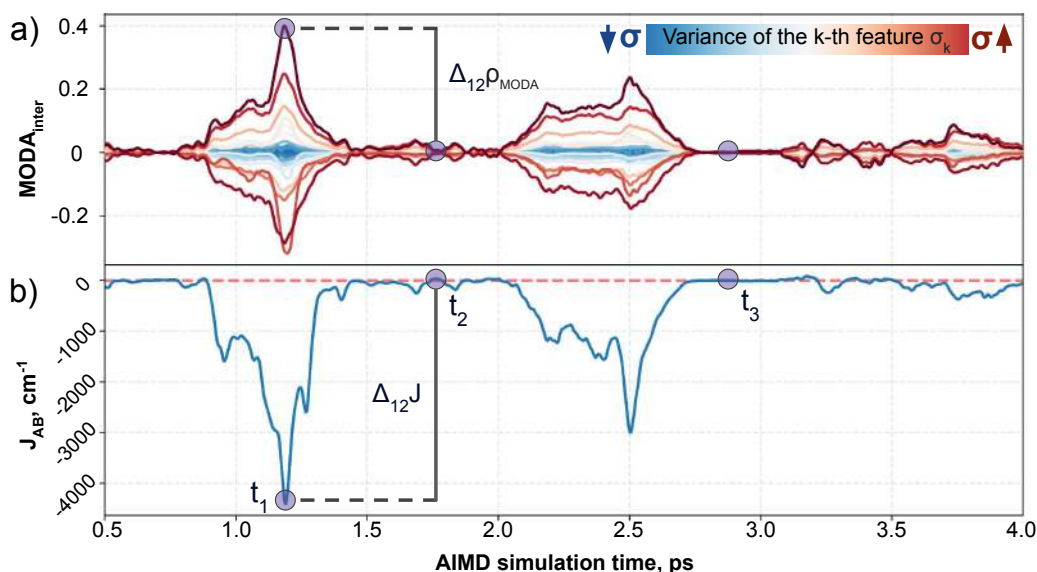


Figure 5.9: Evolution of (a) intermolecular MODA features (colored by variance value) of a TTTA dimer along the structures explored during the 0.5–4.0 ps time interval of AIMD, as well as (b) the associated time-resolved evolution of J_{AB} in this interval. The blue circles in (a) and (b) allow for comparison between variations in J_{AB} and variations in the features of MODA and SOAP (see Figure 5.6).

The qualitative correlation observations discussed so far for SOAP and MODA can be quantified by means of a variance–covariance analysis (see Figure 5.10). In this regard, we observe for MODA that the features bearing a larger variance (σ_1) simultaneously possess the highest covariance with J_{AB} (Figure 5.10a). That is, those features exerting the greatest influence on the overall molecular descriptor (high variance), are also those presenting the highest joint variability with J_{AB}

(high covariance). This stark disparity is evident when comparing the outcomes of MODA with those of SOAP and BoB, as depicted in Figure 5.10b-c. For example, the SOAP component with the greatest influence (σ_1 in Figure 5.10b) displays a level of covariance with J_{AB} comparable with the sixth-ranked feature of MODA (see $\sigma_1^{\text{SOAP}} = 0.098$ vs. $\sigma_6^{\text{MODA}} = 0.088$ in Figure 5.10a). Likewise, BoB components present the lowest variance-covariance relations among the ones inspected here, not only for the lower magnitude in covariance compared to MODA, but also because of the uneven distribution of covariance attributed to their features sorted variance-wise. Furthermore, the cumulative covariance of the 20 most significant features in SOAP and BoB (338.2 and 162.1, respectively) are substantially lower than for MODA, which is 598.2. Overall, these results set the stage to conclude that MODA, being a separable QIR, is better suited to predict J_{AB} than other separable CIR-type descriptors like BoB and SOAP.

However, it is important to recognize that variance-covariance analysis inherently assumes linear relationships between features and J_{AB} , potentially oversimplifying the complex, often non-linear nature of these relationships. In this regard, we examined the MAE metric resulting from training a KRR model, coupled with the separable RBF kernel. In line with the initial experiments conducted in this chapter, we used various portions of the HT-300K dataset for training, testing, and fine-tuning the optimal γ^l , γ^i , and λ hyperparameters. Likewise, the performance of these models was then assessed using data from the HT-250K AIMD trajectory, serving as a validation set. Moreover, we used the leave-p-groups-out cross-validation scheme (LPGO-CV) to prevent potential overfitting issues, as detailed in **Chapter 2**.

For all the descriptors, the test learning curves (dashed lines in Figure 5.11) still exhibit a reduction in MAE beyond a training size of 40%. However, the validation curves (solid lines) reach a saturation plateau at 25% regardless of the descriptor in use. This indicates that each model achieves an adequate training size around 25%, and increasing the data size beyond this point only results in KRR models with similar extrapolation and interpolation capabilities. Concerning the accuracy of the different descriptors, KRR models based on BoB and SOAP display MAEs around 40 cm^{-1} and 30 cm^{-1} , respectively. In contrast, the KRR model utilizing MODA representation outperforms both SOAP and BoB, registering a substantially lower MAE under 20 cm^{-1} . While the test curve for BoB and SOAP (dashed red and dashed green lines in Figure 5.11) reaches the lowest MAE, their respective validation curves (solid red and solid green lines) register substantially higher MAEs in both cases. In contrast, test and validation curves employing MODA (dashed and solid purple lines in Figure 5.11, respectively) show a joint evolution along the entire range, which indicates that KRR model trained upon MODA does not present as much overfitting as in

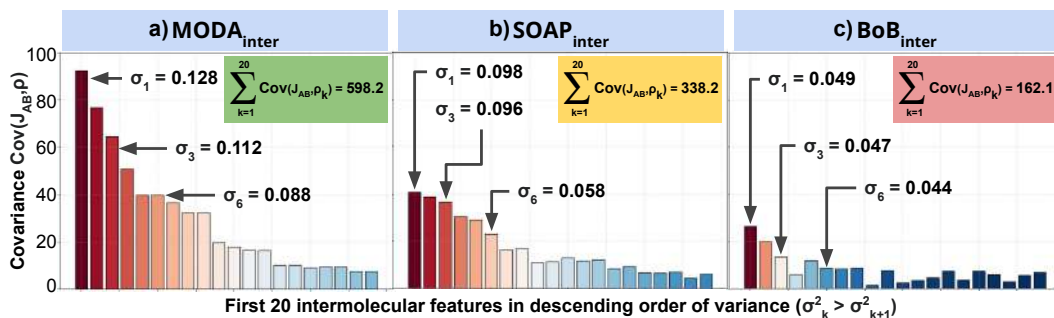


Figure 5.10: Variance-covariance analysis for intermolecular components of (a) MODA, (b) SOAP and (c) BoB with J_{AB} . The x-axis on each plot (and the color of each bar) corresponds to the variance of the 20 intermolecular features with highest variance, while the y-axis corresponds to the covariance of the feature with J_{AB} . Insets show the cumulative covariance of all the plotted 20 features.

BoB and SOAP. These differences reinforce the superior performance of MODA in predicting J_{AB} values in the TTTA dataset, resulting in a better choice for capturing the underlying relationships in the data and generating more accurate predictions of J_{AB} .

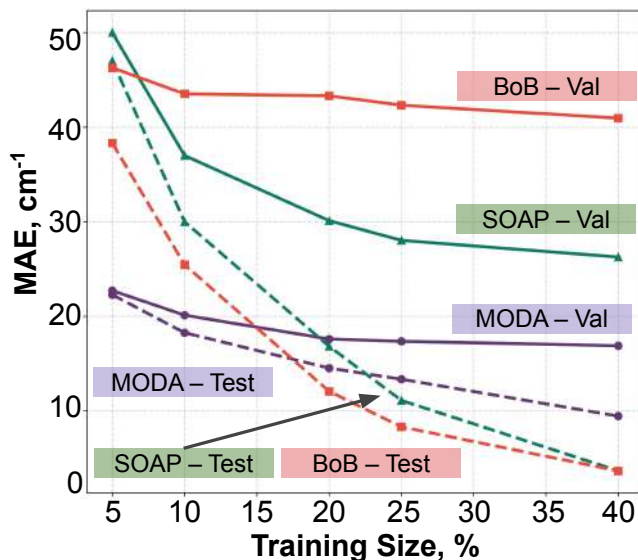


Figure 5.11: KRR learning curves obtained from BoB (red), SOAP (green) or MODA (purple) descriptors, using the leave p -groups-out strategy. Solid and dashed lines indicates the data source: HT-300K (test) and HT-250K (validation), respectively.

Moreover, it is important to note that the MAE of 20 cm^{-1} recorded for MODA corresponds to a 10% error margin with respect to the median of the J_{AB} distribution (ca. 350 cm^{-1}), a substantial enhancement compared to the 15% and 21% errors represented by MAEs of 30 cm^{-1} and 40 cm^{-1} for SOAP and BoB, respectively. This improvement is particularly impactful in practical applications. For instance, in molecular magnetism applications the accurate J_{AB} predictions are crucial, as they serve as intermediate steps in calculating thermodynamic properties like magnetic susceptibility. It is well-established that

the prediction of such macroscopic observables from first-principles calculations is highly sensitive to the accuracy in reproducing J_{AB} , especially when J_{AB} is close to 0 cm^{-1} .⁶⁹ Therefore, the enhanced precision of J_{AB} prediction offered by MODA can be of critical important in practice.

As final remarks, it is important to mention that, although the structure of this chapter might lead to deduce that applying MODA may require significant user supervision for effective learning (emphasizing human rather than machine learning), it should be understood that the extensive process discussed here is intended to provide a fundamental understanding of MODA's working mechanisms and its differences compared to other descriptors. Thus, the application of MODA in pre-trained models, or even in the training process for new applications, is designed to be straightforward and automatic, following the insights and computational recipes presented here. Based on our results, we believe that using MODA only requires exploring the basis set and the quality of the electronic structure guess for a proper performance, without requiring exhaustive analysis of the dataset nuances. Moreover, note that the analysis of the performance of MODA has been restricted to J_{AB} interactions of pairs of TTTA radicals, nothing precludes the application of MODA for the prediction of other properties (such as energy or electronic couplings) in other datasets.

5.3. Conclusions

In this chapter, we have conducted an exploration, employing ML methodologies to enhance the prediction of complex intermolecular molecular J_{AB} interactions. The primary contribution has been the introduction and development of a new QIR descriptor referred to as Molecular Orbital Decomposition and Aggregation (MODA).

MODA stands as an innovative integration of Quantum Chemical insights and Machine Learning techniques. It is engineered to effectively capture and represent the nuanced interactions within the electronic structure. Moreover, MODA is particularly tailored to be compatible with the splitting of intramolecular and intermolecular components, which is a fundamental characteristic to predict intermolecular magnetic exchange couplings (J_{AB}) with enhanced accuracy. Accordingly, the analytical rigor applied in this study has unveiled the superior performance of MODA over conventional CIR descriptors such as Bag of Bonds (BoB) and Smooth Overlap of Atomic Positions (SOAP). Moreover, the results discussed for MODA when coupled with KRR has provided a deeper insight into the behavior of these descriptors, highlighting MODA's reduced tendency towards overfitting. This superiority is not merely in terms of the outcomes related to KRR, but also in its proved capacity to consistently reflect variations in J_{AB} across diverse molecular configurations of TTTA.

In summary, we believe that MODA emerges as an innovative, versatile, and effective tool combining Quantum Chemistry and Machine Learning. Its introduction can mark a significant advancement in the field, which could potentially inspire the development of new descriptors and also emerge as a useful tool for enhancing the discovery of new molecules and materials by means of fast evaluations. Finally, it is important to note that the insights gained through this research not only aid our understanding of molecular systems but also contribute to establish the foundations of ML applications in the PhD candidate's research group.

References

- (1) Y. LeCun, Y. Bengio and G. Hinton, *Nature*, 2015, **521**, 436–444.
- (2) J. Schmidhuber, *Neural Networks*, 2015, **61**, 85–117.
- (3) A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, *Attention Is All You Need*, 2017.
- (4) E. Yurtsever, J. Lambert, A. Carballo and K. Takeda, *IEEE Access*, 2020, **8**, 58443–58469.
- (5) T. Morimoto and Y. Hashimoto, *Control Engineering Practice*, 2000, **8**, 555–567.
- (6) F. Klauschen, K.-R. Müller, A. Binder, M. Bockmayr, M. Hägele, P. Seegerer, S. Wienert, G. Pruneri, S. de Maria, S. Badve, S. Michiels, T. Nielsen, S. Adams, P. Savas, F. Symmans, S. Willis, T. Gruosso, M. Park, B. Haibe-Kains, B. Gallas, A. Thompson, I. Cree, C. Sotiriou, C. Solinas, M. Preusser, S. Hewitt, D. Rimm, G. Viale, S. Loi, S. Loibl, R. Salgado and C. Denkert, *Seminars in Cancer Biology*, 2018, **52**, 151–157.
- (7) P. Baldi, P. Sadowski and D. Whiteson, *Nature Communications*, 2014, **5**.
- (8) P. Leinen, M. Esders, K. T. Schütt, C. Wagner, K.-R. Müller and F. S. Tautz, *Science Advances*, 2020, **6**.
- (9) B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe and K.-r. Muller, *IEEE Signal Processing Magazine*, 2008, **25**, 41–56.
- (10) B. Perozzi, R. Al-Rfou and S. Skiena, Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2014.
- (11) D.-O. Won, K.-R. Müller and S.-W. Lee, *Science Robotics*, 2020, **5**.
- (12) D. Ferrucci, A. Levas, S. Bagchi, D. Gondek and E. T. Mueller, *Artificial Intelligence*, 2013, **199-200**, 93–105.
- (13) D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel and D. Hassabis, *Nature*, 2016, **529**, 484–489.

- (14) J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K. R. Müller and A. Tkatchenko, *Chemical Reviews*, 2021, **121**, 9816–9872.
- (15) B. Huang and O. A. von Lilienfeld, *Chemical Reviews*, 2021, **121**, 10001–10036.
- (16) A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé and A. Laio, *Chemical Reviews*, 2021, **121**, 9722–9758.
- (17) A. Nandy, C. Duan, M. G. Taylor, F. Liu, A. H. Steeves and H. J. Kulik, *Chemical Reviews*, 2021, **121**, 9927–10000.
- (18) H. J. Kulik, T. Hammerschmidt, J. Schmidt, S. Botti, M. A. Marques, M. Boley, M. Scheffler, M. Todorović, P. Rinke, C. Oses, A. Smolyanyuk, S. Curtarolo, A. Tkatchenko, A. P. Bartók, S. Manzhos, M. Ihara, T. Carrington, J. Behler, O. Isayev, M. Veit, A. Grisafi, J. Nigam, M. Ceriotti, K. T. Schütt, J. Westermayr, M. Gastegger, R. J. Maurer, B. Kalita, K. Burke, R. Nagai, R. Akashi, O. Sugino, J. Hermann, F. Noé, S. Pilati, C. Draxl, M. Kuban, S. Rigamonti, M. Scheidgen, M. Esters, D. Hicks, C. Toher, P. V. Balachandran, I. Tamblyn, S. Whitelam, C. Bellinger and L. M. Ghiringhelli, *Electronic Structure*, 2022, **4**.
- (19) N. Fedik, R. Zubatyuk, M. Kulichenko, N. Lubbers, J. S. Smith, B. Nebgen, R. Messerly, Y. W. Li, A. I. Boldyrev, K. Barros, O. Isayev and S. Tretiak, *Nature Reviews Chemistry*, 2022, **6**, 653–672.
- (20) M. Deffner, M. P. Weise, H. Zhang, M. Mücke, J. Proppe, I. Franco and C. Herrmann, *Journal of Chemical Theory and Computation*, 2023.
- (21) K. P. Greenman, W. H. Green and R. Gómez-Bombarelli, *Chemical Science*, 2022, **13**, 1152–1162.
- (22) J. Westermayr and P. Marquetand, *Chemical Reviews*, 2021, **121**, 9873–9926.
- (23) P. O. Dral and M. Barbatti, *Nature Reviews Chemistry*, 2021, **5**, 388–405.
- (24) P. Marquetand, *Chemical Reviews*, 2022, **122**, 15996–15997.
- (25) H. Türk, E. Landini, C. Kunkel, J. T. Margraf and K. Reuter, *Chemistry of Materials*, 2022, **34**, 9455–9467.
- (26) S. Wengert, G. Csányi, K. Reuter and J. T. Margraf, *Journal of Chemical Theory and Computation*, 2022, **18**, 4586–4593.

- (27) J. T. Margraf, H. Jung, C. Scheurer and K. Reuter, *Nature Catalysis*, 2023, **6**, 112–121.
- (28) J. Kruger and D. Dunning, *Journal of Personality and Social Psychology*, 1999, **77**, 1121–1134.
- (29) M. P. Bahlke, N. Mogos, J. Proppe and C. Herrmann, *The Journal of Physical Chemistry A*, 2020, **124**, 8708–8723.
- (30) W. Fujita and K. Awaga, *Science*, 1999, **286**, 261–262.
- (31) S. Vela, F. Mota, M. Deumal, R. Suizu, Y. Shuku, A. Mizuno, K. Awaga, M. Shiga, J. J. Novoa and J. Ribas-Arino, *Nature Communications*, 2014, **5**.
- (32) S. Vela, M. Deumal, M. Shiga, J. J. Novoa and J. Ribas-Arino, *Chemical Science*, 2015, **6**, 2371–2381.
- (33) C.-I. Wang, M. K. E. Braza, G. C. Claudio, R. B. Nellas and C.-P. Hsu, *The Journal of Physical Chemistry A*, 2019, **123**, 7792–7802.
- (34) C.-I. Wang, I. Joanito, C.-F. Lan and C.-P. Hsu, *The Journal of Chemical Physics*, 2020, **153**, 214113.
- (35) O. Çaylak, A. Yaman and B. Baumeier, *Journal of Chemical Theory and Computation*, 2019, **15**, 1777–1784.
- (36) B. Onat, C. Ortner and J. R. Kermode, *The Journal of Chemical Physics*, 2020, **153**.
- (37) A. P. Bartók, R. Kondor and G. Csányi, *Physical Review B*, 2013, **87**.
- (38) B. Huang and O. A. von Lilienfeld, *Nature Chemistry*, 2020, **12**, 945–951.
- (39) H. Huo and M. Rupp, *Machine Learning: Science and Technology*, 2022, **3**, 045017.
- (40) K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller and A. Tkatchenko, *The Journal of Physical Chemistry Letters*, 2015, **6**, 2326–2331.
- (41) J. Behler, *The Journal of Chemical Physics*, 2011, **134**, 074106.
- (42) A. Fabrizio, K. R. Briling and C. Corminboeuf, *Digital Discovery*, 2022, **1**, 286–294.
- (43) K. Karandashev and O. A. von Lilienfeld, *The Journal of Chemical Physics*, 2022, **156**, 114101.
- (44) S. Llenga and G. Gryn'ova, *ChemRxiv*, 2023.

- (45) L. Himanen, M. O. J. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke and A. S. Foster, *Computer Physics Communications*, 2020, **247**, 106949.
- (46) M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, *Physical Review Letters*, 2012, **108**.
- (47) S. Daulton, D. Eriksson, M. Balandat and E. Bakshy, Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence, ed. J. Cussens and K. Zhang, PMLR, 2022, vol. 180, pp. 507–517.
- (48) L. Fang, X. Guo, M. Todorović, P. Rinke and X. Chen, *Journal of Chemical Information and Modeling*, 2023, **63**, 745–752.
- (49) M. Ojala and G. C. Garriga, *Journal of machine learning research*, 2010, **11**.
- (50) S. Van Vaerenbergh and I. Santamaría, *Regularization, Optimization, Kernels, and Support Vector Machines*, 2014, 477–501.
- (51) V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti and G. Csányi, *Chemical Reviews*, 2021, **121**, 10073–10141.
- (52) H. Tyrallis, G. Papacharalampous and A. Langousis, *Water*, 2019, **11**, 910.
- (53) L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie and L. Farhan, *Journal of Big Data*, 2021, **8**.
- (54) M. Greenacre, P. J. F. Groenen, T. Hastie, A. I. D’Enza, A. Markos and E. Tuzhilina, *Nature Reviews Methods Primers*, 2023, **3**.
- (55) K. E. Pilario, M. Shafiee, Y. Cao, L. Lao and S.-H. Yang, *Processes*, 2019, **8**, 24.
- (56) P. Li, Y. Pei and J. Li, *Applied Soft Computing*, 2023, **138**, 110176.
- (57) M. Rinderle, W. Kaiser, A. Mattoni and A. Gagliardi, *The Journal of Physical Chemistry C*, 2020, **124**, 17733–17743.
- (58) A. Grisafi and M. Ceriotti, *The Journal of Chemical Physics*, 2019, **151**, 204105.
- (59) M. F. Langer, A. Goëßmann and M. Rupp, *npj Computational Materials*, 2022, **8**.
- (60) R. K. Cersonsky, M. Pakhnova, E. A. Engel and M. Ceriotti, *Chemical Science*, 2023, **14**, 1272–1285.

- (61) T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to Algorithms*, MIT Press, London, England, 2nd edn., 2001.
- (62) W. E. Donath and A. J. Hoffman, *IBM Technical Disclosure Bulletin*, 1972, **15**, 938–944.
- (63) J. J. Girerd, Y. Journaux and O. Kahn, *Chemical Physics Letters*, 1981, **82**, 534–538.
- (64) A. J. Bridgeman, G. Cavigliasso, L. R. Ireland and J. Rothery, *Journal of the Chemical Society, Dalton Transactions*, 2001, 2095–2108.
- (65) S. Lehtola, *Journal of Chemical Theory and Computation*, 2019, **15**, 1593–1604.
- (66) J. H. V. Lenthe, R. Zwaans, H. J. J. V. Dam and M. F. Guest, *Journal of Computational Chemistry*, 2006, **27**, 926–932.
- (67) S. Lehtola, *Physical Review A*, 2020, **101**.
- (68) R. Hoffmann, *The Journal of Chemical Physics*, 1963, **39**, 1397–1412.
- (69) J. J. Novoa, M. Deumal and J. Jornet-Somoza, *Chemical Society Reviews*, 2011, **40**, 3182.

Publication #4
Unlocking the predictive power of
quantum-inspired representations
for intermolecular properties in
Machine Learning



Cite this: DOI: 10.1039/d3dd00187c

Unlocking the predictive power of quantum-inspired representations for intermolecular properties in machine learning†

Raul Santiago, * Sergi Vela,  Mercè Deumal  and Jordi Ribas-Arino 

The quest for accurate and efficient Machine Learning (ML) models to predict complex molecular properties has driven the development of new quantum-inspired representations (QIR). This study introduces MODA (Molecular Orbital Decomposition and Aggregation), a novel QIR-class descriptor with enhanced predictive capabilities. By incorporating wave-function information, MODA is able to capture electronic structure intricacies, providing deeper chemical insight and improving performance in unsupervised and supervised learning tasks. Specially designed to be separable, the multi-moieity regularization technique unlocks the predictive power of MODA for both intra- and intermolecular properties, making it the first QIR-class descriptor capable of such distinction. We demonstrate that MODA shows the best performance for intermolecular magnetic exchange coupling (J_{AB}) predictions among the descriptors tested herein. By offering a versatile solution to address both intra- and intermolecular properties, MODA showcases the potential of quantum-inspired descriptors to improve the predictive capabilities of ML-based methods in computational chemistry and materials discovery.

Received 20th September 2023
Accepted 13th November 2023

DOI: 10.1039/d3dd00187c

rsc.li/digitaldiscovery

Introduction

Machine Learning (ML) is having a tremendous impact in Quantum Chemistry (QC).^{1–6} Several research fields are benefiting from new computational strategies combining QC and ML, such as molecular electronics,^{7,8} excited states,^{9–11} low cost discovery of materials,^{12,13} or catalysis.¹⁴ Generally, ML models require data to be transformed into a fixed-size *representation*, usually in the form of a vector, where each element represents a specific attribute or feature. In chemistry, the construction of these elements (typically called *descriptors*) is particularly challenging due to the diversity and complexity of

chemical systems and their interactions. Unlike other applications, molecules do not possess an intrinsic order, structure or size, which makes it difficult to define a universal representation. In the last decade, significant efforts have been made to develop reliable descriptors.¹⁵ These can be classified in three different categories. The first one are cheminformatics descriptors, based on either string fingerprints^{16,17} or on descriptive properties that are easily obtainable by *a priori* knowledge, such as the number of aromatic rings or the molecular size.^{18,19} The second category comprises descriptors based on three-dimensional structural information, usually supplemented with parameters inherited from classical mechanics (e.g., electrostatic potentials, or local density overlaps). In essence, this category does not consider the principles of quantum mechanics and, thus, the descriptors belonging to this category can be referred to as classical-informed representations (CIR). Examples include SOAP²⁰ (Smooth Overlap of Atomic Positions), SLATM²¹ (Spectrum of London and Axilrod-Teller-Muto), many-body interaction descriptor,²² MBTR²³ (Many-Body Tensor Representation), BoB²⁴ (Bag of Bonds) or ACSF²⁵ (Atom-Centered Symmetry Functions). Finally, the third category are descriptors that utilize principles of quantum mechanics to represent the molecular systems and, hence, can be referred to as quantum-informed representations (QIR), with SPA^HM²⁶ (Spectrum of Approximated Hamiltonian Matrices Representation), FJK²⁷ (Fock-Coulomb-Exchange) and MAOC²⁸ (Matrix of Orthogonalized Atomic Orbital Coefficients) being the few existing ones.

Departament de Ciència de Materials i Química Física, Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Martí i Franquès, 1, 08028 Barcelona, Spain. E-mail: raul.sant.1972@gmail.com

† Electronic supplementary information (ESI) available: (1) Assessment of the effect of the basis set in MODA intermolecular components. (2) Decoupling formalism used to separate intra- and intermolecular components of BoB and SOAP; methodology based on spectral clustering for automatic detection of moieties in multi-moieity systems. (3) Detailed description of the TTTA dataset; two-sample Kolmogorov–Smirnov test for HT-300K and HT-250K AIMD simulations. (4) Results of agglomerative clustering for BoB, SOAP, and MODA with a varying number of clusters. (5) Formalism and strategy to calculate natural orbitals from the density matrix, including an extension for approximating multi-moieity NOs from linear combinations of NOs; representation of the SONOs for TTTA, THL, and PHYL included in the MODA representation. (6) Analysis of MODA and BoB results for intra- and intermolecular decoupling. (7) Description of the leave-p-groups-out cross-validation strategy used in the KRR predictions. (8) Notes on the performance of descriptors. See DOI: <https://doi.org/10.1039/d3dd00187c>



The main advantage of QIR over CIR descriptors is that the former can explicitly encode the electronic state of a system, including its electronic structure, charge and spin multiplicity. These attributes are necessary to discriminate, *e.g.*, radicals from closed-shell molecules, or neutral from charged species, particularly when the molecular geometry is not significantly affected (and thus, CIR does not significantly change). Although in their early stage of development, QIR descriptors are believed to have the potential to enhance the prediction of properties rooted in the electronic structure of a system. However, a common problem for both CIR and QIR approaches when combined with ML models is the prediction of intermolecular properties.^{29,30} The reason is that, for the sake of better sensitivity, the majority of the descriptors emphasize the magnitude of short-range interactions to better capture the local atomic environment. As a result, these representations can underperform in cases where medium- or long-range (possibly intermolecular) interactions are crucial.^{15,31} This issue can be addressed by decoupling intra- and intermolecular interactions in the representation, and simply ignoring the former in the construction of the descriptor.^{29,32} Although this approach has proven to be successful, neglecting the intramolecular components might be a rather severe simplification, since they can be correlated to the intermolecular property to some extent. For this reason, we here present a mathematical formalism able to accommodate fine-tuned contributions of intra- and intermolecular components in kernel-based ML models.

In general, the possibility to decouple atomic interactions depends on the strategy used to encode molecular information. For instance, BoB is based on 2-body Coulomb potentials and, hence, it can be decoupled using some distance or connectivity criterion. Contrarily, SPA^M and MAOC, being based on an eigen-decomposition of the Hamiltonian, are not compatible with such an approach, as the resulting eigen-states are related to many-body interactions, potentially gathering contributions from atoms in different moieties. The development of methods that combine the advantages of QIR descriptors for fine-grained representations and the separability of some CIR are sought, as would enable accurate predictions of intermolecular properties while maintaining the benefits of QIR approaches. The relevance of such methods is thus clear, as it would significantly extend the capabilities of ML models in predicting complex intermolecular properties. Herein, we introduce the Molecular Orbital Decomposition and Aggregation (MODA) as a new QIR descriptor, the first of its kind that allows decoupling strategies. As in SPA^M and MAOC, MODA does not require the calculation of self-consistent field (SCF) solutions. Instead, MODA representation can be constructed using well-established “guess” Hamiltonians,³³ such as the Superposition of Atomic Densities (SAD),³⁴ the Superposition of Atomic Potentials (SAP),³⁵ or the extended Huckel method³⁶ (EH), which are typically starting points in quantum chemistry. The use of these “guess” Hamiltonians provides a computationally-light, yet powerful and simple framework to develop QIR descriptors.

One adequate platform to test and develop new QIR descriptors and also evaluate the possibility to separate intra- and intermolecular components is molecular magnetism. In

particular, we have focused on the evaluation of magnetic exchange couplings, J_{AB} , which quantify the strength and character of the spin–spin interactions between two, A and B , spin carrying moieties. As it has been shown for datasets comprising di-copper complexes described by BoB, SOAP and MBTR, among others, the prediction of J_{AB} with non-linear regression models can be challenging.³⁷ Moreover, magnetic interactions often occur through-space between independent molecular units and, thus, J_{AB} becomes an intermolecular property. In the following sections, (a) we present the formalism to derive MODA and the strategy to decouple it in intra- and intermolecular components, (b) we discuss the modifications required to kernel-based ML models to accommodate the decoupled representations and, (c) we assess the performance of MODA and other descriptors at predicting magnetic exchange couplings, J_{AB} , as a representative example of an intermolecular property.

In this work, we implement decoupled versions of BoB and SOAP, and we show its advantages over the standard versions when intermolecular properties are targeted. Additionally, we prove that MODA outperforms the decoupled versions of BoB and SOAP in standard supervised regression methods and unsupervised classification models, such as Kernel Ridge Regression³⁸ (KRR) and Agglomerative Clustering³⁹ (AC). Overall, we will demonstrate that MODA is the best representation, among the descriptors here tested, to tackle the prediction of intermolecular J_{AB} interactions, while still being suitable for intramolecular ones.

Methodology

Mathematical formalism

The computation of MODA starts from the atomic positions, which are the only required input (see Fig. 1, step 1). Subsequently, molecular orbitals, the core mathematical object to assemble the representation, are calculated, and expanded as a linear combination of atomic orbitals (see Fig. 1, step 2):

$$|\psi_k\rangle = \sum_i c_{ik} |\chi_i\rangle \quad (1)$$

where c_{ik} are the coefficients describing the contribution of each atomic orbital, $|\chi_i\rangle$, to the molecular orbital, $|\psi_k\rangle$. From this set of coefficients, we formulate the density matrix (D) as a sum of partial density matrices associated to each molecular orbital (D_k).

$$D = \sum_k n_k |c_k\rangle \langle c_k| = \sum_k D_k \quad (2)$$

where n_k corresponds to the orbital occupation. We can measure the interaction strength between pairs of atoms in the system using the density matrix of the molecule (eqn (2)) by means of the Mayer's definition of bond order.⁴⁰ First, the matrix $P = DS$ is constructed, where the overlap matrix (S) is included to account for the non-orthogonality of the atomic orbital basis set. Subsequently, the bond order between two atoms (\mathcal{B}_{AB}) is determined by the trace of the product of P_{AB} and P_{BA} blocks, which contain the rows and columns of P associated



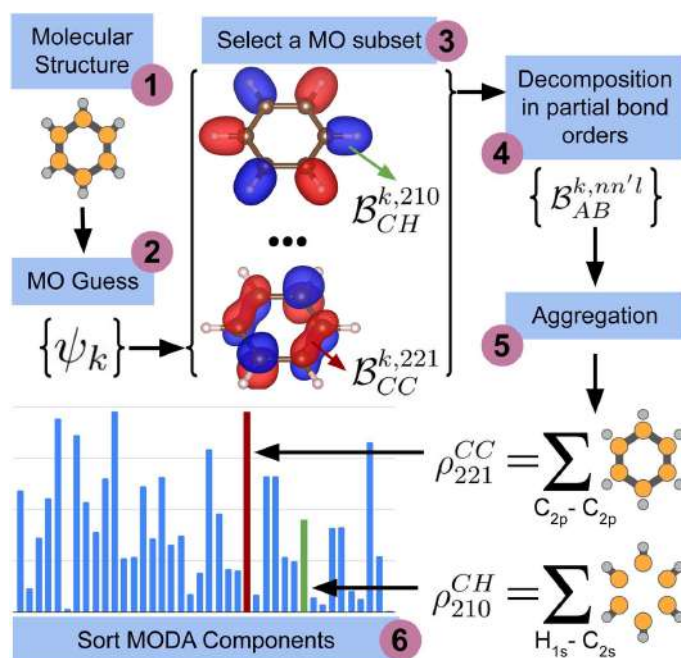


Fig. 1 Scheme of the workflow to compute the features of MODA. The figure shows the process that produces MODA features from the molecular structure to the final representation.

with the atomic orbitals centered at atoms A and B . Formally, the Mayer's bond order is defined as

$$B_{AB} = \sum_{i \in A} \sum_{j \in B} p_{ij} p_{ji} = \text{Tr}(P_{AB} P_{BA}) \quad (3)$$

where i and j run over all the basis functions centered at atoms A and B , respectively, and p_{ij} are the elements of the matrix P . Exploiting the decomposition of the density matrix into orbital contributions, as stated in eqn (2), the bond order can be expressed as a sum of partial bond orders (B_{AB}^k) each corresponding to the contribution of a specific molecular orbital, $|\psi_k\rangle$, to B_{AB} (see Fig. 1, steps 3 & 4).

$$B_{AB} = \sum_k \text{Tr}(P_{AB}^k P_{BA}^k) = \sum_k B_{AB}^k \quad (4)$$

The contributions to the partial bond can be decomposed even further. Specifically, we can examine the contributions stemming from the different atomic orbitals of $|\psi_k\rangle$. To this end, we introduce an orbital-specific partial bond order contribution, here represented as $B_{AB}^{k,nn'l}$. In this notation n and n' correspond to the principal quantum number of the basis functions of atoms A or B , while l corresponds to the azimuthal quantum number. Building upon that, the centerpiece strategy

to generate MODA components consists of the aggregation of such orbital-specific partial bond orders in many-body batches to guarantee the translational, permutational and rotational invariance of the representation. Mathematically, MODA representation corresponds to the set $\{\rho_{nn'l}^{Z_A Z_B}\}$, where each element is defined as (see step 5 in Fig. 1):

$$\rho_{nn'l}^{Z_A Z_B} = \sum_k \sum_{A \in Z_A} \sum_{B \in Z_B} B_{AB}^{k,nn'l} \quad (5)$$

here A and B refer to all possible atoms in a given molecule with atomic numbers Z_A and Z_B , respectively. That is, each term in MODA aggregates the partial bond orders of all A - B interactions coming from specific atomic orbitals univocally identified by n , n' and l quantum numbers, allowing all possible values of the magnetic quantum number, m (see step 6 in Fig. 1). Our choice of using one azimuthal quantum number while incorporating two principal quantum numbers is inspired by the level of nuance present in the standard SOAP representation. Nevertheless, alternative versions of MODA could delve deeper into distinguishing components by incorporating another quantum number, l' . This would allow to differentiate, for example, σ - σ from σ - π interactions. However, this extra level of nuance would also come at the cost of enlarging the representation size.



Intra-/intermolecular decoupling strategy in MODA

As MODA batches result from the aggregation of pairwise interactions, it can be adapted to decouple intra- and intermolecular interactions. This capability can be incorporated directly into our formalism through a minor modification of eqn (5). For instance, consider δ_{AB} as a parameter that takes a value of 1 when atoms A and B belong to the same molecule, and 0 otherwise. Accordingly, MODA components ascribed to intra- and intermolecular interactions (with super-index I and I , respectively) can be defined as

$$\begin{aligned}\rho_{m'l}^{Z_A Z_B, I} &= \sum_{\forall k} \sum_{A \in Z_A} \sum_{B \in Z_B} \delta_{AB}^{k, m'l} \delta_{AB} \\ \rho_{m'l}^{Z_A Z_B, I} &= \sum_{\forall k} \sum_{A \in Z_A} \sum_{B \in Z_B} \delta_{AB}^{k, m'l} (1 - \delta_{AB})\end{aligned}\quad (6)$$

Although eqn (5) and (6) consider the contribution of all molecular orbitals (see the range of k in the outermost summation), we can tune MODA components to include only a S subset of orbitals, denoted as $k \in S$. This is beneficial to avoid the computation of bond orders between orbitals with negligible contributions (e.g., core orbitals), or when the property under examination arises primarily from a specific group of molecular orbitals. For instance, here we have selected the singly-occupied natural orbitals (SONOs) to interpret our property of interest, J_{AB} . Accordingly, the described methodology can be applied to any set of orbitals, and allows the flexible selection of both the level of theory and the basis set, which should be considered as hyperparameters of the MODA representation.

In line with other works in this field,²⁶ our implementation of MODA employs the SAD “guess”, which is a good compromise between computational efficiency and quality of the approximate electronic structure. Note that, resembling the impact of basis sets in quantum chemistry computations, larger basis sets lead to better MODA representations and, thus, better predictions (see Section 1 in the ESI†).

Multi-moieties decoupling

Many ML models, such as Kernel Ridge Regression (KRR) or Agglomerative Clustering (AC), utilize the kernel trick^{41,42} to map the features space into samples space, allowing to use non-linear metrics in practice. The radial basis function (RBF) is a common kernel of choice, which can be defined as

$$\kappa(\rho_1, \rho_2; \gamma) = e^{-\gamma \|\rho_1 - \rho_2\|^2} \quad (7)$$

where ρ_1 and ρ_2 are vector representations of two data samples computed by an arbitrary descriptor, γ is a hyperparameter that determines the decay of the RBF kernel, and $\|\cdot\|$ is the euclidean distance. When the representation ρ can be split in intramolecular (ρ^I) and intermolecular features (ρ^I), the euclidean distance can be arranged as:

$$\|\rho_1 - \rho_2\|^2 = \|\rho_1^I - \rho_2^I\|^2 + \|\rho_1^I - \rho_2^I\|^2 \quad (8)$$

Consequently, the RBF kernel can be expressed as the product of intra- and intermolecular kernels, where γ is

replaced by two hyperparameters (γ^I, γ^I) that control the width of intra- and intermolecular components, respectively. The definition of the RBF kernel (eqn (7)) can then be combined with the former splitting (eqn (8)) as

$$\begin{aligned}\kappa(\rho_1, \rho_2) &= e^{-\gamma^I \|\rho_1^I - \rho_2^I\|^2} e^{-\gamma^I \|\rho_1^I - \rho_2^I\|^2} \\ &= \kappa(\rho_1, \rho_2; \gamma^I) \kappa(\rho_1, \rho_2; \gamma^I)\end{aligned}\quad (9)$$

The use of separate hyperparameters for intra- and intermolecular components provides a more nuanced approach to model molecular interactions of multi-moieties systems, as the ML model can calibrate the effect of γ^I and γ^I separately. Moreover, the user can aid in the interpretation of the model's predictions by simply inspecting $\gamma^I/(\gamma^I + \gamma^I)$. However, it is important to consider the trade-off between the added flexibility and the increased difficulty in training the model. As the number of hyperparameters increases, so does the effort required to properly fit the model to the data. Additionally, having a larger number of hyperparameters also increases the risk of overfitting, leading to poor generalization performance on unseen data. Therefore, it is crucial to balance the benefits of having more fine-tuned control of the interactions against the increased complexity and the risk of overfitting. We consider the former intra-/intermolecular splitting of the kernel as a reasonable balance between model complexity and interpretability. In addition, note that the here-derived formalism is not specific of the RBF kernel, instead, it can be adapted to a Laplacian kernel or cosine-similarity-based kernels, among other choices.

Finally, it is crucial to note that the formalism derived herein requires the detection of the intra- vs. intermolecular origin of the contributions. Consequently, the algorithm assigned to this task should be informed of the nature of the interaction occurring between each pair of atoms. Given that datasets customarily encompass hundreds or even thousands of data samples, it is crucial that this process is automated in practice to facilitate its efficient application. To meet this requirement, we have devised an algorithm anchored in spectral clustering elaborated in detail in the Section 2 in the ESI.†

Dataset

Next, we introduce all the datasets chosen to illustrate the performance of MODA, prior to present all results and discuss them. Three different datasets have been used in this study to investigate the connection between structural representation and J_{AB} . All three focus on purely organic radicals, since they undergo massive variations in J_{AB} as a result of subtle structural changes.⁴³ The first two datasets are computationally tailored to explore changes along a specific internal coordinate of the system to reduce the structural complexity, providing a simple and chemically intuitive platform to expose the differences between CIR and QIR descriptors. The third dataset comprises a challenging and experimentally realized scenario that includes changes in all structural degrees of freedom in order to test the performance of the evaluated descriptors under more complex conditions.



The first *ad hoc* dataset consists of 60 different conformations of phenalenyl (PHYL, see Fig. 2a) radical pairs, arranged at a fixed inter-planar distance of 3.50 Å and varying the twist in the stacked dimer, θ , from fully eclipsed to staggered configurations ($\theta = 0^\circ$ and 60° , respectively). This dataset provides insight into the influence of relative orientation and molecular orbital symmetry on intermolecular J_{AB} couplings.⁴⁴ The second *ad hoc* dataset, containing 180 different conformations of Thiele's diradical⁴⁵ (THIL, see Fig. 2b), spans the central phenyl ring's rotational angle from $\theta = 0^\circ$ to 180° (see θ in Fig. 2b). Notice that the substantial bulkiness of the terminal phenyl groups induces an out-of-plane twist to alleviate steric hindrance. As a result, the symmetry of the molecule in the planar conformation diminishes from D_{2h} to either D_2 or C_{2h} . Our dataset is based on the rotation of the central phenyl ring of the C_{2h} conformer along the axis connecting the two $\cdot\text{CPh}_2$ moieties. Note that this rotation implies a sweeping range from $\theta = 0^\circ$ to 180° without mirror symmetry at 90° (refer to Section 5 in the ESI† for a 3D representation of the THIL diradical). This dataset provides a valuable test case for analyzing variations of intramolecular J_{AB} as a result of conformational changes.^{46,47}

The third dataset is made of dimers of 1,3,5-trithia-2,4,6-triazapentalenyl (TTTA see Fig. 2c). In the solid state, TTTA molecules experimentally arrange forming labile 1D π -stacks⁴⁸

(see Fig. 2d). At low temperatures, these remain as alternated dimers but, at higher temperatures, a Pair-Exchange Dynamics (PED) process is triggered, in which the inter-planar distance between a given pair of TTTA molecules oscillates between *ca.* 3.25 Å and *ca.* 4.5 Å.⁴⁹ Such subtle structural changes lead to massive time- and temperature-dependent variations in the J_{AB} interaction between pairs of π -stacked TTTA units,⁵⁰ ranging from weakly ferromagnetic (FM, *ca.* 50 cm^{-1}) to strongly anti-ferromagnetic (AFM, *ca.* -5000 cm^{-1}), as shown in Fig. 2e.

Following the approach of previous works for sample mining,^{7,30,32,51} we obtained TTTA dimers from *ab initio* molecular dynamics (AIMD) simulations of the high-temperature (HT) phase, carried out at two different temperature conditions, 300 K (HT-300K) and 250 K (HT-250K), in order to explore different regions of the thermally available configurational space. Each AIMD simulation had an independent 10 ps run, yielding a variety of structures. The data from the HT-300K phase, detailed in a previous publication⁴⁹ by some of the present authors, included 30 000 configurations selected from 3 different dimers of the crystal structure (10 000 of each one). These configurations have been used to train and test the ML models, offering a diverse training set that encapsulates a wide range of structural variations. Conversely, the HT-250K phase data is a new addition, generated specifically for this study. We

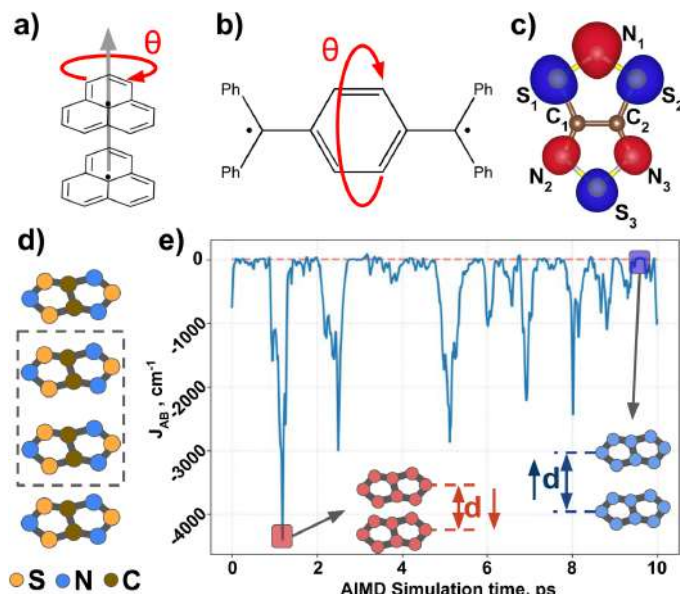


Fig. 2 (a) π -Stacked pair of phenalenyl radicals (PHYL), (b) Thiele's diradical (THIL) and (c) chemical representation overlapped with the SONO of TTTA. Red arrows indicate the rotational internal coordinate (θ) explored to obtain all geometries of PHYL and THIL. (d) π -Stack of TTTA radicals, with a dimer highlighted. (e) Time resolved evolution of J_{AB} of two TTTA radicals forming a dimer along a time interval of 10 ps of AIMD simulation. Insets highlight two regions with extremely different J_{AB} values: -4300 cm^{-1} (red) and +20 cm^{-1} (blue).



Digital Discovery

extracted 20 000 structures from this phase, following the same selection strategy. The HT-250K structures have served as a separate dataset for evaluating the extrapolation capabilities of the models, ensuring that our ML models can make accurate predictions on unseen data spanning different regions of the configurational space. A detailed analysis of the TTTA dataset can be found in Section 3 of the ESI†

Results and discussion

The discussion of the results is organized into five subsections. In the first part, we evaluate the capabilities of CIR and QIR descriptors in unsupervised learning tasks involving J_{AB} predictions by means of Agglomerative Clustering (AC). The second section demonstrates the benefits of separating intra- from intermolecular degrees of freedom in descriptors when predicting intermolecular J_{AB} values for both CIR and QIR-type representations. In the third section, we highlight the key qualitative differences between CIR and QIR-type descriptors using the TTTA dataset. Complementing the former, the fourth section validates quantitatively the observations discussed in the previous section using variance-covariance and mutual information analyses. Lastly, in the fifth section, we provide additional evidence supporting MODA's superior performance in TTTA dataset by presenting our findings from KRR predictions. Throughout the discussion, we compare structure-average SOAP and MODA as the main representative examples of CIR and QIR-class descriptors, respectively, that enables the decoupling of intra- and intermolecular components (see similar analyses for BoB in Sections 4 and 6 of the ESI†). As mentioned in previous sections, MODA describes a specific subset of molecular orbitals. In this case, the MODA representation of the systems under study is constructed using their respective SONOs, since these MOs play a crucial role in determining the J_{AB} values⁵² (see more information in Section 5 of the ESI†).

CIR vs. QIR performance on *ad hoc* datasets

PHYL and THIL datasets focus on an intermolecular J_{AB} with fixed intra-molecular components, and a fully intramolecular J_{AB} , respectively. As introduced in the dataset section, both datasets have been specially designed to simplify the structural variability to a single internal coordinate (a rotational angle, θ). DFT calculations have been performed along this coordinate to retrieve θ vs. J_{AB} curves, and the clusters resulting from the AC model are projected onto these curves (see Fig. 3 for PHYL and Section 4 of the ESI† for THIL).

The J_{AB} curve of PHYL spans a wide range of values from strongly AFM couplings in eclipsed and staggered conformations (with D_{3h} and D_{3d} point group symmetries, respectively) to slightly FM couplings at $\theta = 30^\circ$ (S_6 point group), and displays symmetry around $\theta = 30^\circ$ (see Fig. 3a). The mirror-symmetric profile of J_{AB} around $\theta = 30^\circ$ (see Fig. 3a) is at odds with the point group symmetry associated to the PHYL's conformers along θ ($D_{3h} \rightarrow S_6 \rightarrow D_{3d}$). The reason for such discrepancy can be grasped by the electronic structure of the SONOs of the PHYL

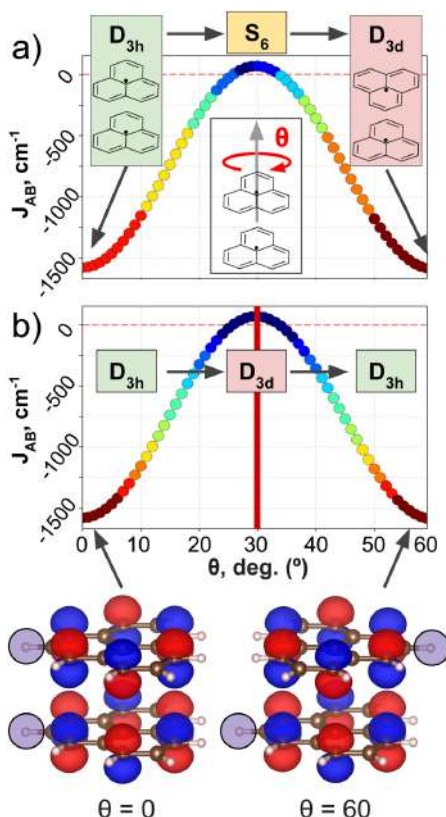


Fig. 3 J_{AB} profile of PHYL conformers from eclipsed ($\theta = 0^\circ$) to staggered ($\theta = 60^\circ$) geometries. The color code indicates the clustering produced by (a) SOAP and (b) MODA descriptors. The insets of (a) indicate the evolution of the point group symmetry at specific values of θ associated with the geometry of the dimer. Analogously, the insets in (b) indicate the point group symmetry associated with the SONOs. The bottom part of the figure illustrates the SONOs at $\theta = 0^\circ$ and $\theta = 60^\circ$, in which the highlighted hydrogen shows that the structure is different, while the SONO remains unaltered.

dimer. Both SONOs at the eclipsed and staggered conformations belong to D_{3h} point group, while the $\theta = 30^\circ$ conformer belongs to the D_{3d} point group in this case (see SONOs in Fig. 3b). Thus, the SONOs of the PHYL conformers from $\theta = 0^\circ$ to 30° ($D_{3h} \rightarrow D_{3d}$) are mirror-symmetric to the conformers from $\theta = 30^\circ$ to 60° ($D_{3d} \rightarrow D_{3h}$), just as the J_{AB} curve along θ . That is, the relevant symmetry to interpret the J_{AB} evolution is not the symmetry associated to the geometry, but the one emerging from the electronic structure.⁵³ Consistent with this observation, the clustering patterns of SOAP (which is



exclusively based on structural information) fail to capture the symmetry around $\theta = 30^\circ$, and instead follow the clustering pattern arising from structural symmetries (Fig. 3a). In contrast, MODA, being a QIR descriptor, produces a mirror-symmetric clustering pattern around $\theta = 30^\circ$, in agreement with the molecular orbital symmetry associated to J_{AB} (Fig. 3b).

It is worth noting that the use of structure-average SOAP vectors might not fully leverage the potential of the descriptor. This is particularly relevant given that the performance of clustering techniques can be significantly impacted by minor distortions in the distance between samples. To substantiate our comparison between SOAP and MODA further, we expanded our investigation to include the use of a more general local SOAP representation, employing the Regularized Entropy Match³⁴ (REMatch) together with RBF kernel to evaluate AC performance from the local SOAP representation. This exploration consistently validated our initial observations regarding the structure-average version of the descriptor, as detailed in Section 4.2 of the ESI†

To further extend the analysis of the capabilities of SOAP and MODA, we assessed their performance using the THIL dataset, which features a molecular diradical instead of a pair of radicals. Here, J_{AB} is geometrically controlled by the through-bond conjugation of both $-\text{CPh}_2$ groups to the central phenyl group guided by the angle: J_{AB} is largely AFM when the central phenyl group remains in-plane ($\theta = 0^\circ$ and $\theta = 180^\circ$), and it drops to zero as the phenyl moiety approaches an orthogonal conformation ($\theta = 90^\circ$). Upon examining the electronic structure of THIL no special symmetry arising from the SONOs can be anticipated and, thus, the atomic disposition and electronic structure of the SONOs are bijectively related (see section 5.3 in the ESI†), which suggests that CIR descriptors can well describe J_{AB} in the THIL dataset. Agglomerative clustering results further confirm this observation, showing small variation among SOAP and MODA descriptors with no significant impact on the AC performance (elaborated in detail in Section 4 of the ESI†). This scenario arises when the system, like THIL, is well-characterized by geometry, rendering CIR and QIR-based descriptors, such as MODA, comparably efficient.

Consequently, our analyses underline MODA's superior performance over CIR methods like SOAP and BoB (as further discussed in Section 4 of the ESI†) in scenarios that present challenges to geometry-based descriptors, as evidenced in the PHYL results. This enhanced performance is due to the QIR's inclusion of wave-function information, capturing crucial, otherwise overlooked, electronic structure elements. Nonetheless, in scenarios where the system is properly defined by geometric descriptors, such as in the THIL dataset, both CIR and QIR-based descriptors can achieve similarly proficient results.

Decoupling intra- and intermolecular features in SOAP

In the previous subsection, we focused on cases with controlled *ad hoc* structural changes, in order to illustrate some fundamental differences between CIR and QIR-type descriptors. However, a more appealing challenge for a ML model is the prediction of intermolecular properties in conditions where the

intramolecular degrees of freedom of each moiety are not artificially constrained. To address this scenario, we turn our attention to the TTTA dimers dataset, where the samples originate from AIMD simulations and, thus, the motion of atoms is uncontrolled and conditioned by thermal fluctuations. As demonstrated below, this is the perfect platform to evaluate the importance of decoupling the components of descriptors. Our analysis initially focuses on structure-average SOAP, as a formerly validated choice in PHYL by Agglomerative Clustering experiments, and then extends to MODA.

The advantages of decoupling intra- and intermolecular interactions become clear when comparing the target J_{AB} with the intra- and intermolecular components of SOAP separately (see Fig. 4). Intramolecular features span a larger range of values, showing the tendency of CIR descriptors to emphasize short-range interactions, but these features show no correlation whatsoever with J_{AB} (see Fig. 4a and b). Thus, even if some degree of correlation exists between intermolecular components and J_{AB} (see Fig. 4b and c), the strong and noisy intramolecular components can make the standard (*i.e.*, not decoupled) SOAP struggle when capturing the evolution of J_{AB} values. Fine ML predictions are still possible, since the intramolecular components might still be correlated with intermolecular ones to some extent (*e.g.*, the variation of the bond lengths when forming a dimer) and hence to J_{AB} . However, such predictive models will primarily memorize data rather than generalize the trends governing the descriptor-to-target mapping, resulting in overfitted ML models that struggle to interpolate and extrapolate beyond the available data. Analogously, the result obtained for SOAP regarding the decoupling of intra/inter features is consistent with the observations for BoB and MODA (see Section 6 of the ESI†).

Performance of MODA and SOAP in the TTTA dataset

Having established (i) the importance of having a QIR descriptor for the prediction of J_{AB} values, and (ii) the advantages of the decoupling strategy, we will now discuss the role of both elements in the case of the TTTA dataset. The SONOs of TTTA have dominant S and N contributions, and negligible contribution from its C atoms (Fig. 2c). Accordingly, MODA components related to C–C interactions are negligible in comparison to the components associated with S–S or N–N interactions (see Fig. 5a). Contrarily, SOAP overemphasizes C–C interactions, thereby rendering them comparable to N–N and S–S components across most of the examined range of Fig. 4. These are the three interaction types that exhibit the highest variance, and whose curves jointly vary along the AIMD simulation time (see highlighted red-most curves in Fig. 4c).

Moreover, MODA's features exhibit a stronger correlation with J_{AB} compared to the intermolecular SOAP components discussed above. This correlation becomes apparent when examining the descriptor response ($\Delta\rho$) to changes in J_{AB} over the AIMD simulation time (t). For instance, when TTTA dimers undergo abrupt structural changes that shift their associated J_{AB} from paramagnetic to strongly AFM regimes (large $\Delta_{12}J$, t_1 vs. t_2 in Fig. 5b), MODA accurately captures these changes (large $\Delta_{12}\rho_{\text{MODA}}$,



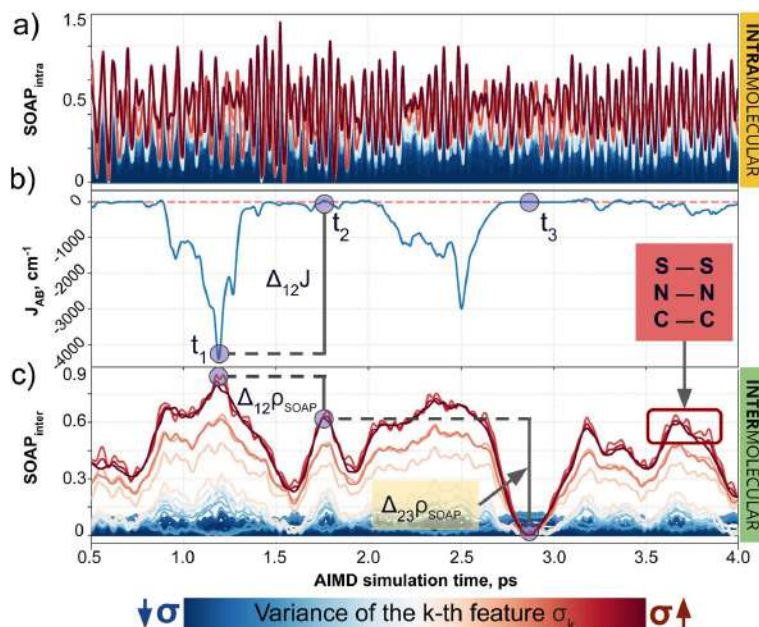


Fig. 4 Evolution of (a) intra- and (c) intermolecular SOAP features (colored by variance value) of a TTTA dimer along the structures explored during the 0.5–4.0 ps time interval of AIMD, as well as (b) the associated time-resolved evolution of J_{AB} in this interval. The blue circles in (b) and (c) indicate three regions with specific values of J_{AB} , where one can compare the change associated with J_{AB} and SOAP.

Fig. 5a). Conversely, SOAP fails to accurately represent these variations (small $\Delta_{12}\rho_{\text{SOAP}}$, Fig. 4c), causing completely opposite structures to appear incorrectly similar. In the same vein, J_{AB} values at $t = t_2$ and $t = t_3$ are almost identical in both cases. In this situation, MODA assigns a similar response to both representations, while SOAP indicates an abrupt change ($\Delta_{23}\rho_{\text{SOAP}}$), even larger than the change associated with the extremely contrasting regimes at $t = t_1$ and $t = t_2$. The primary explanation for this discrepancy is illustrated in Fig. 5d and e, which displays the structure and SONOs of the TTTA dimer at $t = t_2$ and $t = t_3$, see d_{AB} in the figure. Accordingly, the TTTA moieties at $t = t_2$ are notably closer than at $t = t_3$. Nevertheless, the overlap between the SONOs is nearly zero in both cases (compare to the shared isosurface between TTTA SONOs of $t = t_1$ in Fig. 5c), resulting in a similar J_{AB} . This situation is particularly challenging for SOAP, which fails to capture these subtleties due to its lack of electronic structure information. As a result, MODA's ability to capture electronic structure information translates into a higher correlation with J_{AB} than that exhibited by SOAP in this specially critical ill-defined situation for CIR descriptors.

Mutual information and covariance analyses

Our qualitative comparison between descriptors can be complemented using the point-wise Global Feature Reconstruction

Error, $\text{GFRE}^{(l)}(F, F')$. This method involves reconstructing the features of a given descriptor (F') using another one (F), in order to assess the presence/absence of mutual information.⁵⁵ In this case we have assessed how intermolecular components of SOAP can reconstruct MODA, and *vice versa*. For completeness, we have used two flavors of MODA: the one we have used along this section, using only the SONOs ($\text{MODA}_{\text{SONOs}}$), and a complete one in which all the occupied MOs are considered (referred to as MODA_{occ}). The large value around $t = t_3$ in both $\text{GFRE}^{(l)}$ spectra indicates that neither descriptor can reproduce the other, meaning that $\text{MODA}_{\text{SONOs}}$ and SOAP carry substantially different information in their representation (see Fig. 6a). This is expected, since SOAP contains structural information, while $\text{MODA}_{\text{SONOs}}$ contains electronic structure information specially devoted to capture J_{AB} changes. Besides $t = t_3$, we can observe that $\text{MODA}_{\text{SONOs}}$ frequently struggles to reconstruct SOAP, while the opposite rarely happens (see Fig. 6a). The overall perspective can be quantified using the global $\text{GFRE}(F, F')$ (see Fig. 6b), resulting in a reconstruction error that is larger when going from $\text{MODA}_{\text{SONOs}}$ to SOAP (0.531) than the other way around (0.029) (see light blue frames in Fig. 6b). Interestingly, when using MODA_{occ} , mutual reconstruction error with SOAP is similar and low. This suggests that MODA, in its different flavors dictated by the choice of molecular



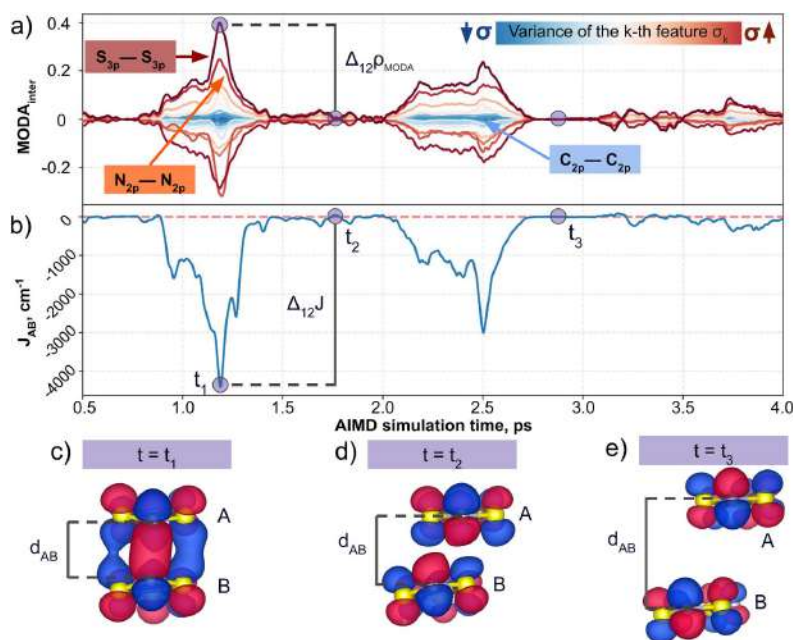


Fig. 5 Evolution of (a) intermolecular MODA features (colored by variance value) of a TTTA dimer along the structures explored during the 0.5–4.0 ps time interval of AIMD, as well as (b) the associated time-resolved evolution of J_{AB} in this interval. The blue circles in (a) and (b) allow for comparison between variations in J_{AB} and variations in the features of MODA. TTTA dimer's SONOs at three representative time steps: (c) $t = t_1$, (d) $t = t_2$, and (e) $t = t_3$.

orbitals, can range from being a specialized (e.g., $\text{MODA}_{\text{SONOs}}$) to a general (e.g., MODA_{occ}) descriptor. On a lesser note, the reconstruction error associated with BoB, irrespective of

during reconstruction or being reconstructed, is substantially higher compared to the rest of the cases (see red-framed values in Fig. 6b). However, the descriptor accomplishing this task

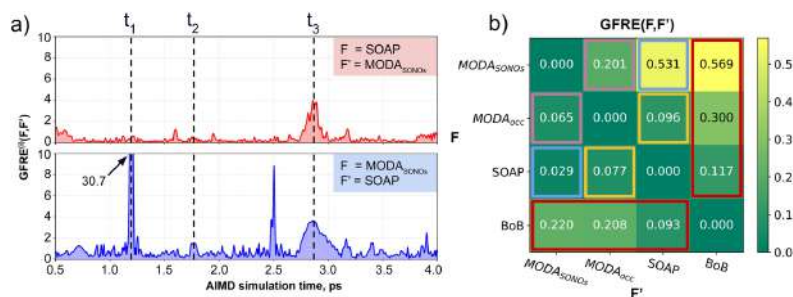


Fig. 6 (a) Point-wise Global Feature Reconstruction Error $\text{GFRE}^{(0)}(F,F')$ during the 0.5–4 ps time interval of AIMD. The upper panel (red curve) shows the error spectrum associated to the reconstruction of $\text{MODA}_{\text{SONOs}}$ (F') using SOAP (F), while the lower panel (blue curve) corresponds to the reconstruction of SOAP features (F') using $\text{MODA}_{\text{SONOs}}$ (F). Dashed black lines indicate three different time regions discussed in the main text. The matrix in (b) indicates the Global Feature Reconstruction Error $\text{GFRE}(F,F')$ using the descriptors in each row (F) to reconstruct the descriptor indicated in each column label (F'). Some matrix elements are highlighted to facilitate the discussion in the text (using color frames).



Digital Discovery

more successfully is SOAP, emphasizing the shared CIR origin of both representations.

To further support our conclusions, we have performed a variance–covariance analysis of the intermolecular features of BoB, SOAP and MODA (refer to Fig. 7). For MODA, the features bearing a larger variance (that is, those exerting the greatest influence on the overall molecular descriptor, see Section 8 in ESI† for an elaborate discussion) simultaneously possess the highest covariance with J_{AB} (Fig. 7a). In essence, the components with the highest variance are also those exhibiting the strongest linear correlation with J_{AB} . This finding stands in stark contrast to the results obtained for SOAP and BoB (Fig. 7b and c). For instance, in SOAP, the component with the largest impact (σ_1 in Fig. 7b) demonstrates a covariance with J_{AB} that is comparable to the 6th most influential feature (σ_6 in Fig. 7a) of the MODA descriptor. In addition, the covariance of the 20 top most determining features of SOAP and BoB (338.2 and 162.1, respectively) are smaller than in MODA (598.2). However, we acknowledge that the variance–covariance analysis assumes linear relationships, while the features may generally exhibit non-linear connections with target properties. Indeed, while linear relationships are favored for their simplicity and stronger generalization capabilities that illustrates significant differences of the performance of each descriptor in the TTTA dataset, we are aware that variance–covariance analysis can oversimplify the data complexity. To account for non-linearity, we have performed a Kernel Ridge Regression (KRR), further enhancing our understanding of descriptor behavior.

KRR predictions

In order to assess the effect of non-linear relations between data representation and J_{AB} , we have employed KRR together with RBF kernel. The learning curves produced by the evaluation of optimal KRR models are presented in Fig. 8, where SOAP, BoB and MODA descriptors have been used. Within the cross-validation leave-p-groups out scheme (see Section 7 of ESI† for more details), we have used different fractions of the HT-300K data for training, testing, and searching the optimal hyperparameters for γ' , γ'' , and α (the regularization parameter of

KRR method), while the model performance is evaluated with data samples from the HT-250K AIMD trajectory (validation set). As discussed in the dataset section, the HT-250K trajectory mostly explores the same potential energy surface region compared to HT-300K, but it also includes unseen regions due to the different thermally-available configurational space (see Section 3.1 of the ESI†), which enables to assess the ability of the model in both interpolation and extrapolation tasks.

For all the descriptors, the test curves (dashed lines in Fig. 8) still exhibit a reduction in MAE beyond a training size of 40%. However, the validation curves (solid lines) reach a saturation plateau at 25% regardless of the descriptor in use. This indicates that each model achieves an adequate training size at 25%, and increasing the data size beyond this point only results in KRR models with similar extrapolation and interpolation capabilities. Concerning the accuracy of the different

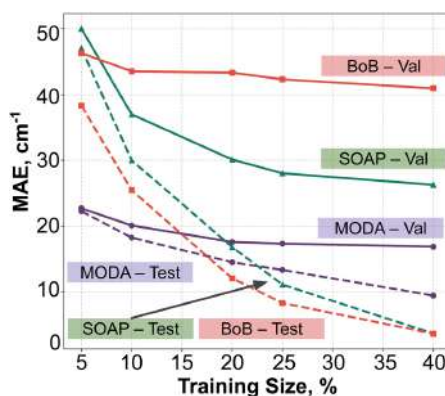


Fig. 8 KRR learning curves using BoB (red), SOAP (green) or MODA (purple), using the leave p-groups-out strategy. Solid and dashed lines indicates the data source: HT-300K (test) and HT-250K (validation), respectively.

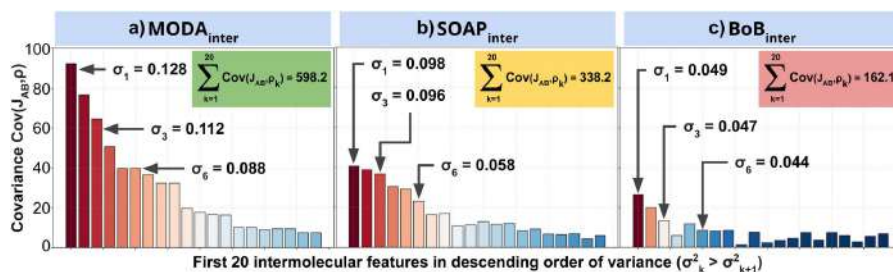


Fig. 7 Variance–covariance plots for intermolecular components of (a) MODA, (b) SOAP and (c) BoB with J_{AB} . The x-axis on each plot (and the color of each bar) corresponds to the variance of the 20 intermolecular features with highest variance, while the y-axis corresponds to the covariance of the component with J_{AB} . Insets show the cumulative covariance of all the plotted features.



descriptors, KRR models based on BoB and SOAP display MAEs around 40 cm^{-1} and 28 cm^{-1} , respectively. In contrast, the KRR model utilizing MODA representation outperforms both SOAP and BoB, registering a substantially lower MAE under 20 cm^{-1} . While the test curve for BoB and SOAP (dashed red and dashed green lines in Fig. 8) reaches the lowest MAE, their respective validation curves (solid red and solid green lines) register substantially higher MAEs in both cases. In contrast, test and validation curves employing MODA (dashed and solid purple lines in Fig. 8, respectively) show a joint evolution along the entire range, which indicates that KRR model trained upon MODA does not present as much overfitting as in BoB and SOAP. These differences reinforce the superior performance of MODA in predicting J_{AB} values in the TTTA dataset, making it a better choice for capturing the underlying relationships in the data and generating more accurate predictions of J_{AB} at a reasonable computational cost (see Section 8 in the ESI†).

Finally, it is pertinent to place the improvement demonstrated in MODA in a proper context. While the overall range of J_{AB} extends approximately from -4000 cm^{-1} to $+100\text{ cm}^{-1}$, a detailed examination reveals that the data points predominantly cluster around 0 cm^{-1} (as detailed in Fig. 3 of the ESI†). The distribution is significantly skewed, with the mean J_{AB} value around 300 cm^{-1} , and a median, which is a more appropriate measure of central tendency for this skewed distribution, approximately at 200 cm^{-1} .

From this standpoint, the MAE of 20 cm^{-1} recorded for MODA corresponds to a 10% error margin with respect to the median, a substantial enhancement compared to the 15% and 21% errors represented by MAEs of 30 cm^{-1} and 42 cm^{-1} for SOAP and BoB, respectively. This improvement is particularly impactful in practical applications. For instance, in molecular magnetism applications, the accurate J_{AB} predictions are crucial, as they serve as intermediate steps in calculating thermodynamic properties like magnetic susceptibility. It is well-established that the prediction of such macroscopic observables from first-principles J_{AB} calculations is highly sensitive to even small variations in J_{AB} , especially in regions nearing 0 cm^{-1} .⁵⁶ Therefore, the enhanced precision of J_{AB} prediction offered by MODA is often of critical importance.

Conclusions

We report the development, implementation and validation of the Molecular Orbital Decomposition and Aggregation (MODA) approach, a novel QIR-class descriptor that encodes the electronic structure of molecules. MODA uses the density matrix of a subset of molecular orbitals to euclidate the contributions of atom pairs as bond orders, and groups these interactions by atom types and classes of atomic orbitals, leading to an invariant representation for molecules. MODA is designed to work in multi-moieties regularization schemes, where the kernel is split into intra- and intermolecular components. We demonstrate that this particular setup results in a better performance in unsupervised and supervised learning tasks targeting intermolecular interactions.

Our study offers a detailed analysis comparing the performance of MODA to some separable CIR descriptors. The examination of pairwise sample-sample maps, along with the application of unsupervised Agglomerative Clustering on *ad hoc* PHYL and THIL datasets, emphasize the importance of capturing the key electronic structure information for an accurate representation of properties rooted in quantum mechanics. Notably, our results show that MODA effectively captures the molecular orbital symmetry, which is crucial for the classification of J_{AB} values, as opposed to other CIR-type representations. The performance of MODA has been further assessed with TTTA dataset, which reinforces the importance of this separation in both CIR and QIR-based representations. Particularly, the variance-covariance analysis and KRR predictions provide additional evidence of MODA's superior performance in capturing relationships with J_{AB} compared to other separable CIR representations.

Overall, MODA, as the first separable QIR-type descriptor, shows potential in enhancing the prediction of molecular properties rooted in quantum chemistry. Based on the specific evaluations presented in this study for J_{AB} values, we believe that MODA will be also applicable to other properties such as HOMO energies or transfer integral predictions.

Computational details

Data mining and J_{AB} evaluation

The AIMD simulation of the HT phase of TTTA at 250 K and 300 K used the same computational protocol employed in earlier works describing this molecule,^{49,50} and other organic radicals.^{57,58} A monoclinic supercell was prepared containing 32 TTTA molecules arranged in 8 stacks of radicals, each of them containing 4 radicals. The AIMD simulation was run for *ca.* 10 ps and a time step of 4 a.u. Vanderbilt ultrasoft pseudopotentials⁵⁹ were employed, together with the PBE functional⁶⁰ within the spin unrestricted formalism and the Grimme D_2 correction,⁶¹ and a Γ -point sampling of the Brillouin zone. The Car-Parrinello⁶² propagation scheme was employed as implemented in CPMD,⁶³ using a fictitious mass for the orbitals of 400 a.u. The simulations were performed in the canonical (or NVT) ensemble using Nosé-Hoover chain thermostats.⁶⁴ All J_{AB} values were evaluated as the difference between the energy of the Broken Symmetry⁶⁵ (E_{BS}) and Triplet (E_T) electronic states assuming the Heisenberg hamiltonian, $\hat{H} = -2J_{AB}\hat{S}_A\hat{S}_B$.⁶⁶ E_{BS} and E_T were computed at the UB3LYP⁶⁷⁻⁶⁹/6-31+G^{70,71} level as implemented in Gaussian09.⁷² We have employed the two-sample Kolmogorov-Smirnov test⁷³ (as implemented in the *scipy* module in Python) on the statistical distributions of both the centroid-centroid distances between TTTA units, and the J_{AB} values associated with structures extracted from the HT-300K and HT-250 K trajectories, to guarantee that the configurational space sampled at 300 K and 250 K is different (see more details in Section 3.1 of the ESI†). Moreover, PHYL and THIL conformations have been obtained *via* a rigid rotational scan of the angle θ , as described in the dataset section, after a geometry relaxation in the triplet state. We have then calculated J_{AB} values of every configuration following the same level of theory



Digital Discovery

described above for TTTA, and without imposing structural or orbital symmetry (“NoSymm” keyword).

Descriptors and molecular representations

The decoupled version of BoB, SOAP and MODA descriptors have been obtained through our code implemented in Python and collected in the MLcool module (<https://github.com/GEM2-UB/MODA>). The decoupling scheme applied to BoB (refer to Section 2.1 in the ESI†) is similar to MODA, as both are built from 2-body interactions. However, SOAP, as implemented in many packages, is a 3-body descriptor,⁷⁴ and necessitates a different approach. Therefore, we adopted the approach proposed by Cersonsky *et al.* for SOAP decoupling,⁷⁵ detailed in Section 2.2 in the ESI†. In particular, BoB has been implemented from scratch, while we use the DSCribe⁷⁶ module for local and structure-average versions of SOAP are used in the backend. The version of MODA employed in the manuscript computes the natural orbitals (NOs) and occupation numbers by means of the following two steps: (1) the density matrix is obtained by means of the SAD guess (with guess = “atom”) as implemented in pySCF.⁷⁷ (2) A Lowdin’s symmetric diagonalization⁷⁸ is performed to solve the generalized eigenvalues equation⁷⁹ ($DSC = Cn$, see details in Section 5.1 of the ESI†). It is important to note that all the molecules considered in this work are radicals, thus, the SAD guess has been computed in the open-shell spin state. Specifically, for THIL, we have computed the density matrix in the triplet state. In contrast, for PHYL and TTTA (both being pairs of radicals), the density matrix was computed for each monomer separately and subsequently, the dimer electronic structure was approximated as a sum of doublets (see more details in section 5.1.1 of the ESI†). SOAP partial power spectrum vector has been calculated on the atomic positions both locally (average = “off” in DSCribe) and in its structure-average version (average = “outer”). We used a Gaussian-type orbital radical decay function and set the n_{\max} , l_{\max} and r_{cut} hyperparameters to 6, 4, and 7, respectively. Moreover, σ has been varied to specific values (0.1, 1, 2, 3 and 10), with no extra radial scaling weighting of the atomic density to let intermolecular components dominate the SOAP spectrum. MODA representations have been computed using different basis sets (STO-6G, 6-31G, 6-31G*, 6-31+G* and aug-cc-pvdz). We found a significant effect of the basis set on the performance of the model (see Section 1 of the ESI†). Only the results of 6-31+G* basis set are reported in the main text, as it has been found to be the best balance for accuracy and computational cost.

Supervised and unsupervised learning

AC and KRR models have been used as implemented in scikit-learn Python’s package.⁸⁰ In both cases the kernel = “pre-computed” option was used to provide our implementation of intra/intermolecular separable RBF kernel version (as described in the methodology section and ESI†) transpiled from Fortran95 and capable of parallel computing. For the local SOAP representation we have employed the Regularized Entropy Match kernel⁸⁴ (REMatch) to evaluate global similarities from local representations. In this case we have experimented with different values of the entropic penalty parameter: $\alpha = 0.01$ (best match),

$\alpha = 1.0$ (intermediate) and $\alpha = 10$ (average-like regime), which constitutes a reasonable choice to sweep the spectrum of regimes that this kernel can work in ref. 81 (see Section 4.1 of the ESI†). Before the RBF kernel is computed, the intra- and intermolecular features have been separately normalized to max = 1 in order to guarantee reasonable optimal values of the hyperparameters. In AC experiments, the whole PHYL and THIL datasets have been used with linkage = “complete”,³⁹ which minimizes the maximum distance between observations of pairs of clusters. A varying number of clusters has been used, ranging from 2 to 10 (the latter is shown in the main text and the rest in Section 4 of the ESI†). We have fitted the KRR model according to a custom implementation of a grid search cross-validation (CV) with leave p-groups-out strategy over γ^l and γ^i hyperparameters of the decoupled RBF kernel, and the ridge regularization term (“alpha” in sklearn implementation). We applied the CV search to TTTA structures of the HT-300K dataset computing R^2 , MAE and RMSE mean values across all the $N!/p!(N-p)!$ CV splits (where N and p stand for total number of groups and number of groups out, respectively), and then used data from the HT-250K phase as our validation set. The grid search for the ridge regularization parameters explored 50 equidistant values in a logarithmic scale ranging from 10^{-11} to 10^{-1} . Similarly, 100 values have been tested for γ^l and γ^i in the interval (10^{-100} , 10^3). The learning curves have been produced by selecting a specific number of total groups and groups out N/p , to get 5% (20/1), 10% (10/1), 20% (5/1), 25% (4/1) and 40% (5/2). GFRE(F,F) and GFRE^(l)(F,F) feature reconstruction measures have been calculated by means of the scikit-matter⁸² Python module, using only intermolecular features of the descriptors. The regression weights for GFRE^(l) were calculated from the samples in the time interval $t \in (4, 10)$ ps and evaluated at $t \in (0.5, 4)$ ps, while a strict 50/50 split have been used in GFRE before random shuffling to decorrelate time-evolution samples.

Data availability

The code and tutorial for the MLcool package, including examples for using BoB, SOAP, and MODA, are available at the GitHub repository: <https://github.com/GEM2-UB/MODA>. A tutorial showing the data analysis of this paper is available in the interactive notebook “MODA_tutorial.ipynb” at <https://github.com/GEM2-UB/MODA>. The complete dataset used for this study, encompassing THIL, PHYL, and TTTA datasets, is accessible at Zenodo: <https://doi.org/10.5281/zenodo.8032717>. Other algorithms and analyses have been uploaded as a part of the ESI†. Both the code and dataset are required to be formally cited in the reference section of this article.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors are thankful for the financial support of the Spanish Ministerio de Economía y Competitividad (Project:



View Article Online

Paper

Digital Discovery

CTQ2017-87773-P/AEI/FEDER), the Spanish Ministerio de Ciencia e Innovación (Projects PID2020-117803GB-I00 and CEX2021-001202-M). The authors also acknowledge support from project grant 2021SGR00354 funded by the Generalitat de Catalunya. R. S. acknowledges a predoctoral FPI grant from MINECO under grant agreement PRE2018-084053. S. V. acknowledges Generalitat de Catalunya for a Beatriu de Pinós grant (2020 BP 00043).

References

- J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K. R. Müller and A. Tkatchenko, *Chem. Rev.*, 2021, **121**, 9816–9872.
- B. Huang and O. A. von Lilienfeld, *Chem. Rev.*, 2021, **121**, 10001–10036.
- A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé and A. Laio, *Chem. Rev.*, 2021, **121**, 9722–9758.
- A. Nandy, C. Duan, M. G. Taylor, F. Liu, A. H. Steeves and H. J. Kulik, *Chem. Rev.*, 2021, **121**, 9927–10000.
- H. J. Kulik, T. Hammerschmidt, J. Schmidt, S. Botti, M. A. Marques, B. Boley, M. Scheffler, M. Todorović, P. Rinke, C. Oses, A. Smolyanyuk, S. Curtarolo, A. Tkatchenko, A. P. Bartók, S. Manzhos, M. Ihara, T. Carrington, J. Behler, O. Isayev, M. Veit, A. Grisafi, J. Nigam, M. Ceriotti, K. T. Schütt, J. Westermayr, M. Gastegger, R. J. Maurer, B. Kalita, K. Burke, R. Nagai, R. Akashi, O. Sugino, J. Hermann, F. Noé, S. Pilati, C. Draxl, M. Kuban, S. Rigamonti, M. Scheidgen, M. Esters, D. Hicks, C. Toher, P. V. Balachandran, I. Tamblyn, S. Whitelam, C. Bellinger and L. M. Ghiringhelli, *Electron. Struct.*, 2022, **4**(2), 023004.
- N. Fedik, R. Zubatyuk, M. Kulichenko, N. Lubbers, J. S. Smith, B. Nebgen, R. Messerly, Y. W. Li, A. I. Boldyrev, K. Barros, O. Isayev and S. Tretiak, *Nat. Rev. Chem.*, 2022, **6**, 653–672.
- M. Deffner, M. P. Weise, H. Zhang, M. Mücke, J. Proppe, I. Franco and C. Herrmann, *J. Chem. Theory Comput.*, 2023, **19**, 992–1002.
- K. P. Greenman, W. H. Green and R. Gómez-Bombarelli, *Chem. Sci.*, 2022, **13**, 1152–1162.
- J. Westermayr and P. Marquetand, *Chem. Rev.*, 2021, **121**, 9873–9926.
- P. O. Dral and M. Barbatti, *Nat. Rev. Chem.*, 2021, **5**, 388–405.
- P. Marquetand, *Chem. Rev.*, 2022, **122**, 15996–15997.
- H. Türk, E. Landini, C. Kunkel, J. T. Margraf and K. Reuter, *Chem. Mater.*, 2022, **34**, 9455–9467.
- S. Wengert, G. Csányi, K. Reuter and J. T. Margraf, *J. Chem. Theory Comput.*, 2022, **18**, 4586–4593.
- J. T. Margraf, H. Jung, C. Scheurer and K. Reuter, *Nat. Catal.*, 2023, **6**, 112–121.
- M. F. Langer, A. Goßmann and M. Rupp, *npj Comput. Mater.*, 2022, **8**, 41.
- D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
- S. R. Heller, A. McNaught, I. Pletnev, S. Stein and D. Tchekhovskoi, *J. Cheminf.*, 2015, **7**, 23.
- H. L. Morgan, *J. Chem. Doc.*, 1965, **5**, 107–113.
- A. Capecchi, D. Probst and J. L. Reymond, *J. Cheminf.*, 2020, **12**, 43.
- A. P. Bartók, R. Kondor and G. Csányi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 184115.
- B. Huang and O. A. von Lilienfeld, *Nat. Chem.*, 2020, **12**, 945–951.
- W. Pronobis, A. Tkatchenko and K.-R. Müller, *J. Chem. Theory Comput.*, 2018, **14**, 2991–3003.
- H. Huo and M. Rupp, *Mach. Learn.: Sci. Technol.*, 2022, **3**, 045017.
- K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller and A. Tkatchenko, *J. Phys. Chem. Lett.*, 2015, **6**, 2326–2331.
- J. Behler, *J. Chem. Phys.*, 2011, **134**, 074106.
- A. Fabrizio, K. R. Briling and C. Corminboeuf, *Digital Discovery*, 2022, 286–294.
- K. Karandashev and O. A. von Lilienfeld, *J. Chem. Phys.*, 2022, **156**, 114101.
- S. Llenga and G. Gryn'ova, *ChemRxiv*, 2023, preprint, DOI: [10.26434/chemrxiv-2023-9pggs](https://doi.org/10.26434/chemrxiv-2023-9pggs).
- M. Rinderle, W. Kaiser, A. Mattoni and A. Gagliardi, *J. Phys. Chem. C*, 2020, **124**, 17733–17743.
- C.-I. Wang, I. Joanito, C.-F. Lan and C.-P. Hsu, *J. Chem. Phys.*, 2020, **153**, 214113.
- A. Grisafi and M. Ceriotti, *J. Chem. Phys.*, 2019, **151**, 204105.
- C.-I. Wang, M. K. E. Braza, G. C. Claudio, R. B. Nellas and C.-P. Hsu, *J. Phys. Chem. A*, 2019, **123**, 7792–7802.
- S. Lehtola, *J. Chem. Theory Comput.*, 2019, **15**, 1593–1604.
- J. H. V. Lenthe, R. Zwaans, H. J. J. V. Dam and M. F. Guest, *J. Comput. Chem.*, 2006, **27**, 926–932.
- S. Lehtola, *Phys. Rev. A*, 2020, **101**, 012516.
- R. Hoffmann, *J. Chem. Phys.*, 1963, **39**, 1397–1412.
- M. P. Bahlke, N. Mogos, J. Proppe and C. Herrmann, *J. Phys. Chem. A*, 2020, **124**, 8708–8723.
- T. Hastie, J. Friedman and R. Tibshirani, *The Elements of Statistical Learning*, Springer, New York, 2001.
- D. Müllner, *Modern hierarchical, agglomerative clustering algorithms*, 2011.
- A. J. Bridgeman, G. Cavigliasso, L. R. Ireland and J. Rothery, *Dalton Trans.*, 2001, 2095–2108.
- T. Hofmann, B. Schölkopf and A. J. Smola, *Ann. Stat.*, 2008, **36**, 1171–1220.
- A. E. Hoerl and R. W. Kennard, *Technometrics*, 1970, **12**, 55–67.
- O. Sato, *Nat. Chem.*, 2016, **8**, 644–656.
- Z. H. Cui, H. Lischka, H. Z. Benerberu and M. Kertesz, *J. Am. Chem. Soc.*, 2014, **136**, 5539–5542.
- J. Thiele and H. Balhorn, *Ber. Dtsch. Chem. Ges.*, 1904, **37**, 1463–1470.
- C. Wenstrup, M. J. Regimbald-Krnel, D. Müller and P. Comba, *Angew. Chem., Int. Ed.*, 2016, **55**, 14600–14605.
- Y. Su, X. Wang, L. Wang, Z. Zhang, X. Wang, Y. Song and P. P. Power, *Chem. Sci.*, 2016, **7**, 6514–6518.
- W. Fujita and K. Awaga, *Science*, 1999, **286**, 261–262.
- S. Vela, F. Mota, M. Deumal, R. Suizu, Y. Shuku, A. Mizuno, K. Awaga, M. Shiga, J. J. Novoa and J. Ribas-Arino, *Nat. Commun.*, 2014, **5**, 4411.



- 50 S. Vela, M. Deumal, M. Shiga, J. J. Novoa and J. Ribas-Arino, *Chem. Sci.*, 2015, **6**, 2371–2381.
- 51 O. Çaylak, A. Yaman and B. Baumeier, *J. Chem. Theory Comput.*, 2019, **15**, 1777–1784.
- 52 J. J. Girerd, Y. Journaux and O. Kahn, *Chem. Phys. Lett.*, 1981, **82**, 534–538.
- 53 Despite the MO symmetry is usually given in terms of an irreducible representation of the point group to which the molecular structure belongs, in this case we are treating the molecular orbitals as the object to which the point group is assigned to.
- 54 S. De, A. P. Bartók, G. Csányi and M. Ceriotti, *Phys. Chem. Chem. Phys.*, 2016, **18**, 13754–13769.
- 55 A. Goscinski, G. Fraux, G. Imbalzano and M. Ceriotti, *Mach. Learn. Sci. Technol.*, 2021, **2**, 025028.
- 56 J. J. Novoa, M. Deumal and J. Jornet-Somoza, *Chem. Soc. Rev.*, 2011, **40**, 3182.
- 57 A. Paul, R. Nasani, A. Mondal, S. Roy, S. Vela and S. Konar, *Cryst. Growth Des.*, 2020, **20**, 6296–6301.
- 58 S. Vela, M. B. Reardon, C. E. Jakobsche, M. M. Turnbull, J. Ribas-Arino and J. J. Novoa, *Chem.–Eur. J.*, 2017, **23**, 3479–3489.
- 59 D. Vanderbilt, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1990, **41**, 7892–7895.
- 60 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- 61 S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799.
- 62 R. Car and M. Parrinello, *Phys. Rev. Lett.*, 1985, **55**, 2471–2474.
- 63 R. Car and M. Parrinello, *Phys. Rev. Lett.*, 1985, **55**, 2471–2474.
- 64 G. J. Martyna, M. L. Klein and M. Tuckerman, *J. Chem. Phys.*, 1992, **97**, 2635–2643.
- 65 L. Noodleman, *J. Chem. Phys.*, 1981, **74**, 5737–5743.
- 66 R. Caballol, O. Castell, F. Illas, I. d. P. R. Moreira and J. P. Malrieu, *J. Phys. Chem. A*, 1997, **101**, 7860–7866.
- 67 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.
- 68 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 69 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.
- 70 P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, 1973, **28**, 213–222.
- 71 M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees and J. A. Pople, *J. Chem. Phys.*, 1982, **77**, 3654–3665.
- 72 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Revision A.1*, Gaussian Inc., Wallingford CT, 2009.
- 73 F. James, *Statistical Methods in Experimental Physics*, WORLD SCIENTIFIC, 2006.
- 74 V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti and G. Csányi, *Chem. Rev.*, 2021, **121**, 10073–10141.
- 75 R. K. Cersonsky, M. Pakhnova, E. A. Engel and M. Ceriotti, *Chem. Sci.*, 2023, **14**, 1272–1285.
- 76 L. Himanen, M. O. J. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke and A. S. Foster, *Comput. Phys. Commun.*, 2020, **247**, 106949.
- 77 Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters and G. K.-L. Chan, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2017, **8**, e1340.
- 78 P.-O. Löwdin, *J. Chem. Phys.*, 1950, **18**, 365–375.
- 79 P. Pulay and T. P. Hamilton, *J. Chem. Phys.*, 1988, **88**, 4926–4933.
- 80 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
- 81 F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi and M. Ceriotti, *Chem. Rev.*, 2021, **121**, 9759–9815.
- 82 A. Goscinski, V. P. Principe, G. Fraux, S. Kliavinek, B. A. Helfrecht, P. Loche, M. Ceriotti and R. K. Cersonsky, *Open Res. Europe*, 2023, **3**, 81.



Chapter 6
Conclusions

The present PhD thesis has covered three distinct, yet interconnected, topics that contribute to the body of knowledge related to open-shell molecules and materials. Specifically, it has focused on the design, evaluation and enhancement of their associated magnetic properties, together with the acceleration of their computation. As the detailed conclusions of each study have been outlined in their respective chapters, here only the the most relevant results are summarized. Additionally, a brief discussion about the future perspectives related to the potential synergies emerging from the insights gained in each study is conducted at the end of this section.

In **Chapter 3**, we proved that the application of both isotropic out-of-plane compression and anisotropic in-plane tensile strain in triarylmethyl-based 2D Covalent Organic Radical Frameworks (TAM-based 2D-CORFs) are promising mechanisms to enhance the relative stability of the semimetallic (SM) and closed-shell quinoidal (CSQ) states with respect to the antiferromagnetic insulator (AFMI) ground state. In both cases, the relative difference in energy has been proven to be mostly driven by two factors: the instrumental rotation of the aryl rings with respect to the material plane and the consequential bond stretching. Specifically, our results unveil that isotropic out-of-plane compression induces a flattening of the aryl rings, thus approaching a graphene-like SM state. In turn, anisotropic in-plane strain leads to an uneven aryl twist through the material, allowing access to the CSQ state. Importantly, the stabilization of the aforementioned states only takes place at low levels of compressive and tensile stress, compatible with experimentally realizable conditions.

In **Chapter 4**, we explored the design of pentalene-based diradicals, specifically focusing on substituted dicyclopenta- $[n]$ acene series (**R₂-DPA $[n]$**) and

dibenzopentalene derivatives (**R₂-DBP**). Our investigation was grounded on the combination of quantum mechanical insights and qualitative models, resulting in the discovery and characterization of diradicals with significantly positive singlet-triplet gaps (ΔE_{ST}), i.e., displaying ferromagnetic (FM) couplings. We demonstrated that **R₂-DPA[n]** compounds with $n < 2$ predominantly exhibit FM interactions between the unpaired electrons, while larger n values lead to an open-shell singlet ground state. In line with the observations related to the xylylene diradicals, we assessed the critical role of the topological connectivity of the spin-bearing unit (**R**) to the coupler for both **R₂-DPA[n]** and **R₂-DBP** sets of diradicals. The magnetic interactions have been found to be drastically affected by this feature. Specifically we found that **2,5-R₂-DPA[0]** and **2,7-R₂-DBP** are the most promising carbon connectivities of the whole set to bolster FM interactions. Additionally, the synergistic stabilization effect between the antiaromatic character of the couplers and the spin-bearing units was a notable finding that complement the related experimental evidences found for antiaromatic-based mono-radicals. In this regard, our results suggest that precursors displaying the highest antiaromatic character display the largest dampening of their antiaromatic character upon the diradical formation. This is further validated by a more favorable Radical Stabilization Energy.

In **Chapter 5**, we introduced a new Machine Learning (ML) descriptor for predicting complex intermolecular magnetic exchange couplings, J_{AB} , referred to as MODA (i.e., Molecular Orbital Decomposition and Aggregation). This descriptor represents a novel confluence of Quantum Chemistry and ML, designed to capture the intricacies of electronic structures and being the first of its kind capable of differentiating intramolecular and intermolecular interactions.

We demonstrated the superior performance of MODA compared to traditional classical-informed representations (CIR), like Bag of Bonds (BoB) and Smooth Overlap of Atomic Positions (SOAP). Our findings were assessed by the application of BoB, SOAP and MODA together with Kernel Ridge Regression on the prediction of J_{AB} for distinct families of open-shell organic molecular materials. Overall, the predictive accuracy of MODA may position it as a transformative tool in the ML-mediated prediction of molecular properties, also potentially accelerating the advancements in the discovery of new molecules and materials.

Notably, a significant synergy emerges when integrating the insights from our studies on TAM-based 2D-CORFs with the properties of pentalene couplers. In this regard, potential pentalene-based 2D-CORFs are expected to display a honeycomb lattice akin to TAM-based 2D-CORFs. However, while TAM-based 2D-CORFs are proven to predominantly exhibit AFM interactions, the pentalene-based 2D-CORFs could display FM interactions. This shift in their magnetic interactions could lead to spin-polarized band structures, which are highly sought after in the design of half-metals or spin filters. The intersection of these two areas of study not only demonstrates the complementary nature of our research but also allows innovative proposals beyond the scope of this PhD thesis.

Another synergy is identified in the potential extension of the MODA descriptor to solid-state systems like 2D-CORFs. The foundational principle of MODA is its reliance on a "guess" electronic structure, characterized by the Superposition of Atomic Densities (SAD). This approach, while formulated for molecular systems, is inherently compatible with solid-state electronic structures. Thus, by adapting MODA for solid-state systems, we can significantly broaden its range

of applicability, potentially making it a versatile tool capable of addressing both molecular and extended material systems.

Overall, this PhD thesis presents advancements in understanding and manipulating the magnetic properties of open-shell molecules and materials, paving the way for innovative applications in technology and materials science. Through theoretical and computational insights, together with the development of a novel Machine Learning tool, it contributes to the future perspectives in the field of organic magnetism.

Chapter 7

List of publications

Publication #1: Santiago, R., Alcón, I., Ribas-Arino, J., Deumal, M., de P. R. Moreira, I., & Bromley, S. T. (2020). *2D Hexagonal Covalent Organic Radical Frameworks as Tunable Correlated Electron Systems*. *Advanced Functional Materials* (Vol. 31, Issue 6).

Contribution: *The candidate performed all the DFT calculations in solid state, developed the software to calculate the Young's modulus, and extract t , U and U/t . The candidate also calculated the band structure and the magnetic exchange couplings. Finally, the candidate participated in the discussion and preparation of the manuscript.*

Publication #2: Alcón, I., Santiago, R., Ribas-Arino, J., Deumal, M., Moreira, I. de P. R., & Bromley, S. T. (2021). *Controlling pairing of π -conjugated electrons in 2D covalent organic radical frameworks via in-plane strain*. *Nature Communications* (Vol. 12, Issue 1).

Contribution: *The candidate performed some of the DFT calculations, specifically in the highest ranges of strain and the whole range of the TPM/PTM material. The candidate prepared the software to extract the Young's modulus and participated in the discussion and preparation of the manuscript.*

Publication #3 (draft): Santiago, R., Carvajal, M. A., Poater, J., Moreira, I. de P. R., Bromley, S. T., Deumal, M. & Ribas-Ariño, J. (2024) *Rational design of organic diradicals with robust high-spin ground state based on antiaromatic building blocks*. (To be submitted)

Contribution: *The candidate had the original idea of exploring antiaromatic moieties as couplers to enhance ferromagnetic interactions. The candidate performed the DFT calculations and the analyses associated to those (NICS, BLA , J_{AB}). The candidate participated in the discussion and preparation of the manuscript.*

Publication #4: Santiago, R., Vela, S., Deumal, M., & Ribas-Arino, J. (2024). *Unlocking the predictive power of quantum-inspired representations for intermolecular properties in machine learning*. Digital Discovery (unissued).

Contribution: *The candidate performed some of the calculations of the 250K dataset of TTTA and all the calculations of THIL and PHYL datasets. Developed the software to decouple BoB and SOAP descriptors. The candidate proposed, developed and programmed MODA and participated in the discussion and manuscript preparation.*