PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Physical Chemistry Chemical Physics
CREATED USING THE RSC LaTeX PCCP ARTICLE TEMPLATE – SEE www.rsc.org/electronicfiles FOR DETAILS
ARTICLE TYPE www. rsc.org/xxxxxx | XXXXXXXX

Principal component analysis of Mn(salen) catalysts†

Filipe Teixeira,*[a] Ricardo A. Mosquera,[b] André Melo,[a] Cristina Freire,[a] and

M. Natália D. S. Cordeiro*[a]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/b000000x

The theoretical study of Mn(salen) catalysts has been traditionally performed under the assumption that the Mn(acacen') (acacen' = 3,3'-(ethane-1,2-diylbis(azanylylidene))bis(prop-1-en-olate)) is an appropriate surrogate for the larger Mn(salen) complexes. In this work, the geometry and the electronic structure of several Mn(salen) and Mn(acacen') model complexes were studied using Density Functional Theory (DFT) at diverse levels of approximation, with the aim of understanding the effects of truncation, metal oxidation, axial coordination, substitution on the aromatic rings of the salen ligand and chirality of the diimine bridge, as well as the choice of the density functional and basis set. To achieve this goal, geometric and structural data, obtained from these calculations, were subjected to Principal Component Analysis (PCA) and PCA with orthogonal rotation of the components (rPCA). The results show that the choice of basis set to be of paramount importance, accounting for up to 30% of the variance in the data, while the differences between salen and acacen' complexes account for about 9% of the variance in the data, and are mostly related to the conformation of the salen/acacen' ligand around the metal centre. Variations in the spin state and oxidation state of the metal centre also account for large fractions of the total variance (up to 10% and 9%, respectively). Other effects, such as the nature of the diimine bridge or the presence of alkyl substituent in the 3,3 and 5,5 positions of the aldehyde moiety, were found to be less important in terms of explaining the variance within the data set. A matrix of discriminants was compiled using the loadings of the principal and rotated components that best performed in the classification of the entries in the data. The scores obtained from its application to the data set were used as independent variables for devising linear models of different properties, with satisfactory prediction capabilities.

1 Introduction

Manganese-salen complexes, hereafter Mn(salen), are well known catalysts in the epoxidation of alkenes and other olefinic materials. Their use in industrial applications has been prompted by their low cost, easy preparation and high activity[8]. The enantioselectivity of Mn(salen) complexes relies not only in the potentially stereogenic sp^3 hybridised carbon atoms of the diimine bridge, but also on both steric and electronic properties of the salen complex[39]. The Jacobsen's catalyst serves as an example of the high levels of catalytic efficiency and stereoselectivity achieved by this class of transition metal complexes. This has triggered further developments in asymmetric catalysis, including the development of novel chiral Mn(salen) complexes[39], the immobilisation of such complexes onto different supporting materials[21,22,28,37] and the rationalisation of the catalytic cycle of this family of compounds using both theoretical and experimental methods[14,32,35,43].

In the presence of a terminal oxidant, Mn(salen) complexes are typically oxidised to their oxo-Mn(salen) analogues. These oxo-complexes are the active species responsible for the epoxidation of the substrate[20,39]. In both types of complexes, the salen ligand adopts a nearly planar conformation. In oxo-Mn(salen) complexes, one of the axial positions is occupied by the oxo- ligand, leaving at all times one vacant axial position in the coordination sphere of the manganese atom. The vacant axial position may be occupied by a counter ion such as chloride, or a donor ligand. This possibility provides a valuable route for the anchoring of Mn(salen) complexes onto a supporting material, using a linker molecule. In recent years, this strategy allowed the development of novel materials with interesting catalytic properties[21,22].

The importance of this catalytic system and the controversies surrounding its catalytic cycle have triggered several theoretical studies. These models rely on Density Functional Theory (DFT) to provide an adequate

† Electronic Supplementary Information (ESI) available: optimised geometries and results from multivariate data analisys. See DOI: 10.1039/b000000x/

[a] *REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal; E-mail: filipe.teixeira@fc.up.pt; ncordeir@fc.up.pt*
[b] *Departamento de Química Física, Facultade de Química, Universidade de Vigo, 36310 Vigo, Galicia, Spain.*

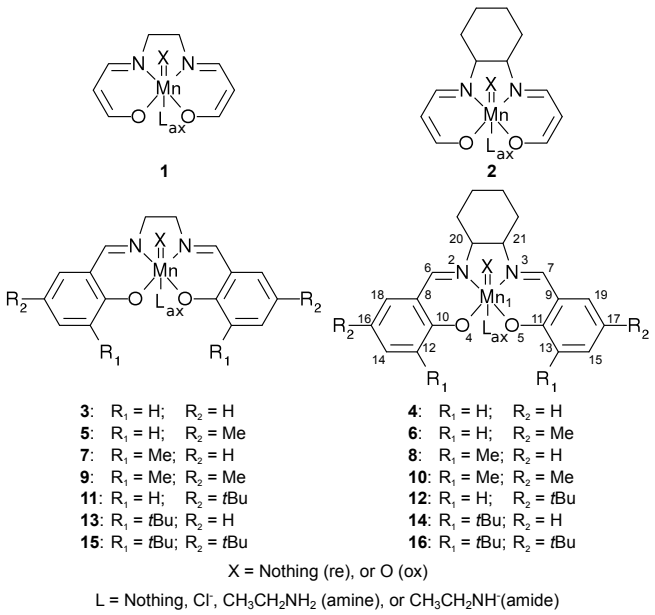

Fig. 1 Manganese-acacen' (**1** and **2**) and manganese-salen (**3** to **16**) complexes used in this study. The diagram for compounds **4**-**16** further illustrated the numbering scheme used in this work. For each compound, X may either be nothing (re) or O (ox); also, L_{ax} may either be nothing, Cl^- (cl), $CH_3CH_2NH_2$ (amine), or $CH_3CH_2NH^-$ (amide).

description of electron correlation at a reasonable cost. The quest for further savings in computational cost has lead to using Mn(acacen') complexes (where acacen' stands for 3,3'-(ethane-1,2-diylbis(azanylylidene))bis(prop-1-en-olate), $^-O(CH)_3N-C_2H_4-N(CH)_3O^-$) such as **1** (Figure 1) as surrogates for the the larger Mn(salen) compounds[14,26,39]. More rigorous methods, such as Coupled Cluster (CC) or Complete Active Space Self-Consistent Field (CASSCF), have been sparsely applied to this problem for being too computer-demanding. These techniques usually required also further truncation of the molecular models[4,50,53]. Moreover, the choice of basis sets in first principles calculations may be considered as an additional truncation on the system. Historically, most theoretical studies had resorted to Pople's split valence 6-31G(d) as the standard for these calculations, usually using a triple-ζ quality basis set only at the metal centre[1–4,10–14,25–27,34], with the exception of the latter work by Morokuma and co-workers[32], in which a triple-ζ quality basis set is used in conjunction with a proper Mn(salen) complex.

With the development of computational technologies and the increased interest in Mn(salen) catalysed epoxidations, an assessment on how the properties of Mn(acacen') may be translated to those of Mn(salen) complexes is mandatory. The question of whether the conclusions withdrawn from such studies are applicable to the larger Mn(salen) complexes, such as the Jacobsen's catalyst (Figure 1, **16**), has been partially addressed by Jacobsen and Cavallo in 2001[25]. However, this study only accounted for the steric effects due to the limitations of the approach assumed. Further insights about this question can only be provided by high-level calculations carried out on the larger Mn(salen) complexes. Despite the initial claims of Jacobsen and Cavallo, some latter studies using different models for Mn(salen) arrived at different conclusions[11,32,33], and the use of truncated models has been pointed as one of the possible causes for these disagreements[38]. Despite this, Mn(acacen') models still enjoy large popularity as surrogates for Mn(salen), making the question of equivalence between the two systems evermore important[58].

An in depth survey of the properties of different Mn(salen) complexes requires the use of a data reduction technique, such as Principal Component Analysis (PCA). This technique allows the determination of linear combinations of numerical variables, the principal components (PCs), that explain the maximum amount of variance in the data not already explained by a previous PC, thus simplifying the data analysis[30]. At the end of the PCA procedure, one obtains a matrix with the coefficients that define each component (the loadings, $\mathbf{M}_L$) and the percentage of variance associated with each PC. The application of the $\mathbf{M}_L$ to the original data matrix corresponds to a change in coordinates, yielding the scores matrix, $\mathbf{M}_S$. Each column of $\mathbf{M}_S$ contains the values that would be observed if the corresponding PC were measured directly. An analysis of the distribution of the scores for each PC allows the association of each PC to categorical variables that are associated with the observations in the original data, this procedure being usually called the supervision step[30,57]. Throughout this work, the term data set will be used to describe the collection of data available on the systems under study, while the term data matrix will be reserved for the subset of the data set that contains only the numerical variables to be subjected to multivariate analysis. Within the realm of chemistry, PCA has enjoyed large success as a tool for processing spectral information[15,48]. Over the years, some efforts have been made towards the use of PCA as a tool in Quantitative Structure-Activity Relationship (QSAR) research for deriving molecular descriptors from *ab initio* calculations[5,24,31,42], its application has also meet some criticism from the mainstream QSAR community, mainly because of the difficulties that may rize from the interpretation of the principal components[46].

The primary aim of this work is to assess the implications of using acacen' as a simplified model for salen in manganese complexes. As a secondary objective, the effects of axial coordination are surveyed by comparing models without axial coordination with models coordinated with chloride, which is the traditional counter ion for these complexes. In addition to this, the axial coordination to $CH_3CH_2NH_2$ and $CH_3CH_2NH^-$

was also studied, with the purpose of surveying the effect of anchoring these complexes onto a supporting material using 3-aminopropyltriethoxysilane (APTES) as a linker[16,36]. For this purpose the geometry and the electronic structure of such type of complexes were analysed using DFT methods. This study included 112 Mn(salen) and 16 Mn(acacen') molecular models. The variants were generated, by changing the axial ligands and/or the substituents of the aldehyde moiety for the Mn(salen) complexes, according to the notation presented in Figure 1. The geometries, for the different molecular models under study, were then optimised using DFT calculations. Two additional sources of variance introduced by the choice of the level of theory at which the DFT calculations are done (density functional and basis set) are also probed. The properties of the different models were surveyed using PCA in order to better access the impact these variations on the Mn(salen) scaffold. In particular, the results presented here aim at understanding the effect truncation of Mn(salen) complexes to Mn(acacen') may have on the properties extracted from first principles calculations.

2 Methodology

*2.0.0.1 DFT Calculations*All DFT calculations were performed using version 3.0.1 of the Orca program package[41]. Natural Population Analysis (NPA) was performed using version 5.9 of the NBO program from the University of Wisconsin[23]. Two density functionals representing different approximations to the Kohn-Shan (KS) formalism were selected: the hybrid-GGA functional X3LYP by Xu and Goddard III[59,60], and the meta-GGA functional TPSS by Perdew and coworkers[54]. In addition, three basis sets were used in conjunction with each functional: the split-valence def2-SV(P) and the triple-ζ def2-TZVP (both derived from Ahlrichs and coworkers[18,19,49,56]) as well as the DefBas-3 basis set, which is defined in the software. This latter basis set corresponds to a triple-ζ valence basis set with contraction pattern $\{311\}$ for hydrogen, $\{62111/411/1\}$ for other main group elements and $\{842111/6311/411/1\}$ for manganese.

Throughout this work, we have chosen to follow the nomenclature used by Cavallo and Jacobsen[26] in which the relevant electronic states are identified by a notation indicating the respective total spin density $S = |\rho_\alpha(\mathbf{r}) - \rho_\beta(\mathbf{r})|$. The symbols *S0*, *S2* and *S4* thus refer to total spin densities of zero (singlet), two (triplet) and four (quintet) unpaired electrons, respectively. Additionally, each compound will be named after the categorical variables that describe it, in the form *Model-OxSt-AxLig-SpinSt-Functional-BasisSet*, where *Model* is the model number according to Figure 1; *OxSt* may take the values "re" or "ox" for the reduced (X=none) and oxidised (X=O) form, respectively; *AxLig* may be "null", "cl", "amine" or "amide" for L= nothing, Cl^-, $CH_3CH_2NH_2$, or $CH_3CH_2NH^-$, respectively; *SpinSt* is either *S0*, *S2* or *S4*, while *Functional* and *BasisSet* correspond to the level of theory in question. When dealing with groups of compounds, only the categorical variables common to all elements of the group are given.

For the *S2* and *S4* states, all calculations were performed under an unrestricted Kohn-Shan formalism, with independent treatment of the electron densities $\rho_\alpha(\mathbf{r})$ and $\rho_\beta(\mathbf{r})$. For *S0* systems, the restricted Kohn-Shan formalism was applied in order to lessen the computational effort of such calculations. Preliminary unrestricted and broken symmetry calculations using model **3** have shown that this would be a valid approach.

Geometry optimisation on all models was carried out using redundant coordinates. These models were constructed by applying truncations and/or substitutions of the (S,S)-Jacobsen's catalyst in its oxidised form and with the $CH_3CH_2NH^-$ axial ligand (**16-ox-amide**). For each geometry, ligand folding was evaluated following the method established by Cavallo and Jacobsen[12] by means of the ligand bent angles, ϕ_e and ϕ_u, as well as the pyramidalisation distance (d_P). A bent angle, ϕ, is defined as $\phi = 180^\circ - \angle(X_{NN} - Mn - X_{NO})$ where X_{NN} and X_{NO} refer to the midpoint between the two nitrogen atoms and the midpoint between adjacent N,O atoms, respectively. Following the numbering scheme shown in Figure 1, the adjacent N,O atoms considered in ϕ_e are N_2 and O_4, whereas ϕ_u refers to atoms N_3 and O_5. Using the same notation, d_P is defined as the separation between the manganese atom and the $O_{(4)} - X_{N,N} - O_{(5)}$ plane[12]. Information on the chirality content of each structure was provided using Avnir's Continuum Chirality Measurement (CCM)[6,44,61,62]. This measurement was obtained using software developed in-house, based on the methods published by Avnir and Alvarez[6,52]. For the purpose of this work, the CCM measurements were taken either on atoms 1 to 21 (*CCM.salen*, which is only defined for models **3** to **16**) or with exclusion of atoms 12 to 19 (*CCM.acacen*, defined for all models).

At the end of the DFT and NBO calculations, a dataset (DB0) was produced. This set consisted of 10 categorical variables and 124 numerical variables. The categorical variables describe the model identification, *Model*; axial ligand, *AxLig*; spin state, *SpinSt*; metal oxidation state, *OxSt*; structure of the diimine bridge (etildiimine, EDA, or cycloexildiimine, DEDA), *Bridge*; the substituent at R_1, R_1; the substituent at R_2, R_2; the structure of the equatorial ligand (salen or acacen'), *EqLig* and the level of theory at which the computations were performed (*Functional* and *BasisSet*). The numerical variables include: a) bond lengths between the atoms labelled in structure **4** (*see* Figure 1), each length being generically referred to as $D.A_1.A_2$, were A_1 and A_2 are the symbols and numbers of the two atoms involved; b) bond angles between triads of the mentioned atoms, each angle being referenced by $A.A_1.A_2.A_3$; c) natural charges ($Q.A_1$) and spin den-

sities ($S.A_1$) for atoms 1 to 22, the axial oxo- ligand ($Q.O_{ax}$ and $S.O_{ax}$, when present), the total natural charge and spin density of the equatorial ligand ($Q.EqLig$ and $S.EqLig$, respectively) and of the axial ligand ($Q.L_{ax}$ and $S.L_{ax}$, when present); d) global geometric parameters (d_P, ϕ_e, ϕ_u, $CCM.acacen$ and $CCM.salen$); and e) the relative energies of *S2* and *S4* with respect to *S0* and also that of *S4* relative to *S2* ($\Delta E_{0,2}$, $\Delta E_{0,4}$ and $\Delta E_{2,4}$, respectively), as well as the energetic balance of the oxidation of the reduced species with H_2O_2 and of the epoxidation of propene, assuming spin conservation (ΔE_{ox} and ΔE_{epox}). All distances were expressed in angstroms, charges and spin densities were expressed in atomic units, angles in degrees and energy differences in $kJ \cdot mol^{-1}$.

*2.0.0.2 Multivariate Analysis*On a first approach, DB0 was pruned out of the following variables: $\Delta E_{0,2}$, $\Delta E_{0,4}$, $\Delta E_{2,4}$,ΔE_{ox} and ΔE_{epox} as well as all variables referring to either axial ligands (O_{ax} and L_{ax}). This subset of DB0 will be referred to as DB1. The analysis of DB1 allowed the comparison between Mn(salen) and Mn(acacen') complexes, as well as many other important factors. However, variables involving atoms 12 to 19 are undefinable for the acacen' models (*see* Figure 1) and were defined *ad hoc* as zero. Another subset of DB0 (labelled DB2) was defined using the same criteria used in DB1 and the additional exclusion of any variable referencing atoms 12 to 19 (including $CCM.salen$).

Multivariate data analysis was carried out using the R language[45] with the Psych package[47] implementation of Principal Component Analysis (PCA) and associated techniques. In order to minimise subsequent problems of chance correlation, the protocol for analysing DB1 and DB2 started by pruning all descriptors for which the coefficient of variance was smaller than 1%. Also, because the numerical values of the remaining descriptors vary significantly, all descriptors were mean centred and scaled to unit variance. Horn's parallel analysis, as implement in the Psych package, was used to determine the number of principal components (PC) or rotated components (RC) considered relevant in PCA, which was performed both without rotation and using the varimax rotation (this latter analysis will henceforth be denoted as rPCA). Because of the large cardinality of DB1 and DB2, the analysis of the scores from PCA and rPCA was performed using density plots (filtered by category), in a manner similar to that presented by H. Attias[7]. This allowed the association of each categorical variable to one or more PCs or RCs in both data sets.

For the development of the linear models, the best performing PCs or RCs for each of the categorical variables were collected into a matrix of discriminants, or a composite loadings matrix, $\mathbf{M}_{CL}$. Numerical descriptors associated to each categorical variable were obtained by application of $\mathbf{M}_{CL}$ to the data matrix in DB0 (once standardised). These descriptors were then used as independent variables in the development of the linear models for$\Delta E_{0,2}$, $\Delta E_{0,4}$, ΔE_{ox}, ΔE_{epox}, $Q.Mn$, $S.Mn$, $Q.EqLig$, $S.EqLig$, $Q.O_{ax}$ and $S.O_{ax}$. A diagram containing the most relevant steps of this procedure is given in Figure 2.

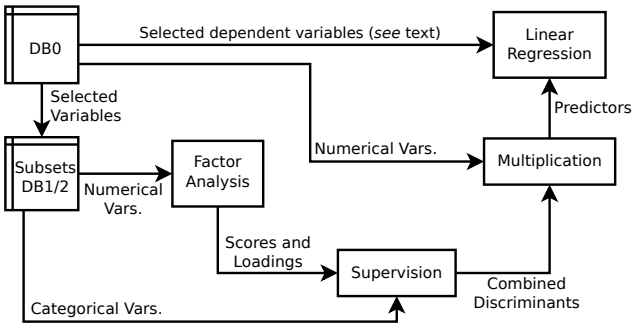

Fig. 2 Flowchart of the most relevant steps in the multivariate analysis and creation of the linear models.

3 Results and Discussion

3.1 DFT Calculations

Results from the DFT calculations show that the relative energies of *S0*, *S2* and *S4* depend strongly on the choice of basis set, with def2-SV(P) showing discrepancies relative to the triple-ζ sets. On average, def2-SV(P) shows a tendency to over-stabilize the higher multiplicity states, with *S2* and *S4* lying -54.0 and -70.0 $kJ \cdot mol^{-1}$relative to *S0*, respectively. On the other hand, an *S2* ground state is preferred by DefBas-3 and def3-TZVP, with a low lying *S0* state (about 6 $kJ \cdot mol^{-1}$), followed by *S4* (at about 140 $kJ \cdot mol^{-1}$). Moreover, the relative energies of these three states appear to vary with respect to the oxidation state of the metal centre. In this regard, the *S4* state of the oxo-manganese(V) complexes is pushed higher in the energy scale. The distribution of the energy gaps between states and their relation with other characteristics of the computational models will be further discussed latter.

At their equilibrium geometry, tetracoordinate complexes are arranged around Mn in a square planar geometry, while hexacoordinated present the expected octahedral geometry around the metal centre. Pentacoordinated species tend to form square pyramidal arrangements around Mn, although with some significant distortion in the case of X=none. In all cases, both salen and acacen' ligands assume a nearly planar conformation around Mn. Absolute energies for all systems, as well as relevant equilibrium geometries are given in the Supporting Information.

Although a proper timing benchmark is outside the scope of this work, some insight can be gained from the time taken from the SCF procedure. Overall, TPSS calculations took about half the time of similar X3LYP computations. Calculations using the DefBas-3 basis set were noticeably faster than

those using def2-TZVP by a factor of three, all other conditions being equal. Also, def2-SV(P) calculations ran in about 80% of the time taken by similar computations using DefBas-3. As expected, the time for each SCF iteration increased with the size of the system. More interestingly, computations using def2-TZVP show difficulties in achieving SCF convergence for the larger systems (specially when using X3LYP), probably due to linear dependencies caused by the larger basis set.

3.2 Principal Component Analysis

The association between each categorical variable (in both DB1 and DB2) and the factors devised from PCA and rPCA was done by examining the distribution of the scores for each component. Table 1 shows the links between each categorical variable and a set of principal and rotated components taken from the two subsets of DB0. Both techniques extracted principal components of DB1 that could be satisfactory associated with a categorical classification of the chemical system or level of theory applied. These principal components account for about 79% of the variance in DB1.

Some numerical variables are only definable for models **m03** to **m16**, and the missing values for **m01** and **m02** were taken as zero. This may have introduced an artificial amount of variance separating salen models from acacen' ones. As a control experiment, the DB2 subset includes only variables which are available for both salen and acacen' models, thus showing the impact of choosing acacen' as a surrogate model for salen. The impact of R1 and R2 is largely local, as it will be further discussed later in this work. Because of this, none of the extracted components of DB2 could be unequivocally associated to the substitution patterns in either R_1 or R_2, as reported in Table 1. The principal components of DB2 that could be associated with categories in this data set account for about 70% of the total variance.

Taking into account the relative variance explained by each of the relevant principal components in DB2, the variance introduced by the choice of computational model for the equatorial ligand accounts for about 9% of the variance in the data. This impact is about the same as the distinction between the reduced manganese (III) and oxidized oxo-manganese (V) species, and is only surpassed by the choice of basis set in the quantum mechanical calculations and also by the characteristics dependent on the spin state.

In order to further pursue our objectives, we shall now discuss the associations described in Table 1. In finding the best discriminant for each categorical variable, Table 1 gives either two or four candidates. The following discussion compares such candidates on the basis of the distribution of their scores. A rationalisation of the physical implications of each source of variance is performed based on the loadings of the best discriminant for each categorical variable. The values for the most prominent loadings obtained from the factor analysis are given in the Electronic Supplementary Information[†].

3.2.0.3 Level of Theory (BasisSet and Functional) Although the data presents a very diverse behaviour when considering the principal components associated with the choice of functional from those associated with the choice of basis set, both questions condense into the fundamental choice concerning the level of theory for the DFT calculations. This question is even more important as one choosing a smaller basis set accounts as another kind of truncation of the chemical model that may go unnoticed. As shown in Table 1, the choice of basis set is the largest source of variation in DB2. Figure 3 shows that these components distinguish calculations using def2-SV(P) from those using the triple-ζ basis sets. Moreover, none of the relevant principal components allowed a distinction between def2-TZVP and DefBas-3. Therefore, the two latter basis sets may be considered as virtually equivalent for the purpose at hand, despite concerns that the absence of polarization functions on hydrogen atoms, as well as the unbalanced amount of *p* and *d* functions in DefBas-3 would render it inappropriate for this type of calculations. Because DB1 is more complete than DB2 (in terms of the variables included in it), PC2 of DB1 was selected as the best discriminant for this category.

The loadings responsible for the scores displayed in Figure 3 reveal that the use of def2-SV(P) leads to larger bond lengths around the manganese atom. This may be related with the tendency shown by most density functionals to over estimate the length of bonds involving transition metal atoms[40], in which case this tendency may be over expressed upon the use of a split valence basis set, but is (at least partially) corrected when using triple-ζ quality basis sets.

On the other hand, the choice of the density functional appears to bear little impact on the variance of the data set, as shown in Table 1. Indeed, the components best associated with the categorical variable *Functional* not only account for a marginal percentage of the total variance in DB1 and DB2, they also barely distinguish between the two functionals used in this study (TPSS and X3LYP). Therefore, RC13 of DB1 was chosen as the best discriminant associated with the category *Functional*. This result is in line with the relative invariance of natural charges and spin densities in **m03-nothing** and **m03-cl** with respect to the level of theory[55]. Despite this, a closer inspection of the loadings responsible for the scores depicted in Figure 4 points towards higher natural charges and spin densities in O_4 and O_5 upon the usage of X3LYP.

3.2.0.4 Spin state (SpinSt) Such as the the choice of basis set and functional, the spin state considered in the DFT calculations is supplied as a parameter. But contrary to the former two variables, the spin state is an intrinsic property of the system. Much discussion has been held on whether Mn(salen)

Table 1 Summary of the results from the PCA and rPCA analysis performed on DB1 and DB2, showing the PCs and RCs best associated to each of the categorical variables in the data set. The percentage of variance explained by each PC or RC is given in parenthesis.

Descriptor	DB1				DB2			
	PCA		rPCA		PCA		rPCA	
EqLig	PC1	(30%)	RC1	(30%)	PC3	(9%)	RC4	(8%)
BasisSet	PC2	(23%)	RC2	(19%)	PC1	(35%)	RC1	(24%)
SpinSt	PC3	(8%)	RC3	(12%)	PC2	(11%)	RC2	(15%)
OxSt	PC4	(5%)	RC5	(4%)	PC4	(7%)	RC3	(9%)
AxLig	PC5	(5%)	RC4	(5%)	PC6	(5%)	RC5	(8%)
Bridge	PC8	(3%)	RC9	(2%)	PC7	(4%)	RC7	(5%)
R_1	PC12	(2%)	RC11	(2%)	NA[a]		NA[a]	
R_2	PC10	(2%)	RC10	(2%)	NA[a]		NA[a]	
Functional	PC14	(1%)	RC13	(2%)	PC11	(2%)	NA[a]	

[a]NA: No clear association was found for this descriptor.

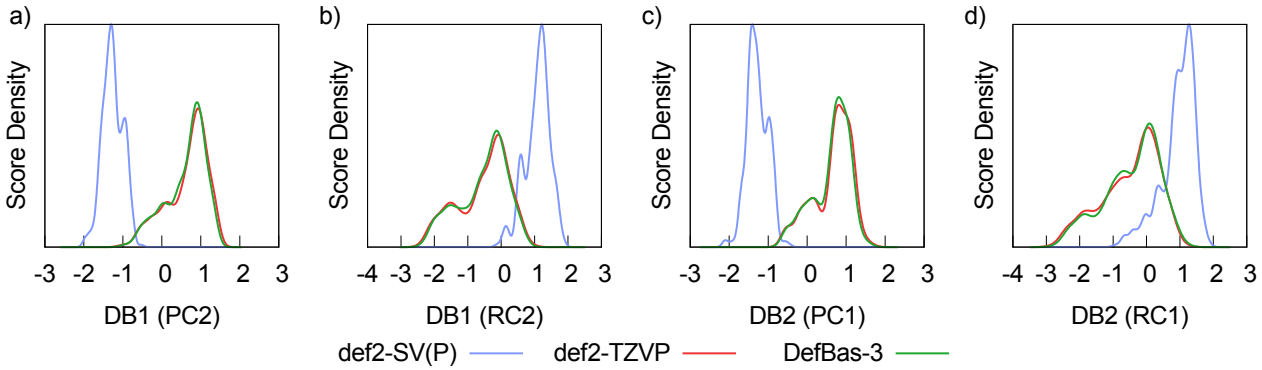

Fig. 3 Distribution of the scores of the principal and rotated components of DB1 and DB2 best associated with the choice of basis set in the DFT calculations.

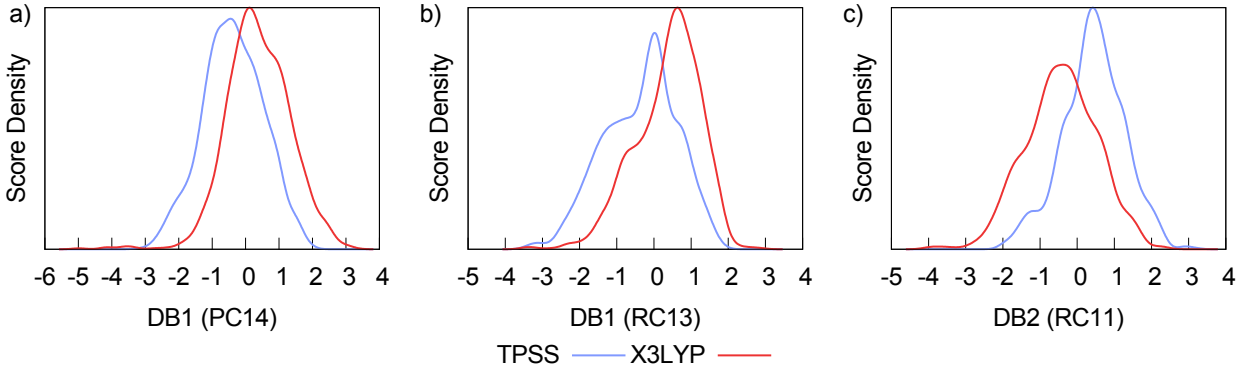

Fig. 4 Distribution of the scores of the principal and rotated components of DB1 and DB2 best associated with the choice of density functional in the DFT calculations.

complexes hold a singlet (*S0*) or a triplet (*S2*) ground state and/or if the reactions in which they are involved occur with or without crossing of spin states[2,4,11,13,26]. This historical record justifies the interest regarding how the spin state defined in DFT calculations may affect the properties of the system in terms of the resulting geometries and population analysis data.

As shown in Table 1, the choice of spin state is among the largest sources of variance in both DB1 and DB2, being the second largest source of variance in DB2. The scores of the principal and rotated components of DB1 and DB2 associated with *SpinSt* are distributed as depicted in Figure 5 and typically show an ordering of the spin states, from *S0* to *S4*, with *S0* species scoring negative values in the principal component axis, while *S4* ones occupy the positive domain. The third principal component of DB1 was chosen as the best descriptor associated with *SpinSt*.

The loadings of the components shown in Figure 5 are similar for both DB1 and DB2, with the atomic spin densities and the distances between the nitrogen atoms an their closest carbon atoms bearing the largest coefficients, as well as the natural charge of N_2, N_3, O_4 and O_5. This latter dependence on the natural charges of some atoms in the equatorial ligand is the main responsible for the partial sensitivity of these components to the oxidation state of the metal centre, giving rise to the bi-modal distributions shown for each spin state in Figure 5. On the other hand, the dominance of spin densities in the loadings associated to *SpinSt* is an expected result when taking into consideration the design of the data sets. By definition the spin density in all atoms of the *S0* species is zero. Also, in the *S2* state, most of the spin density is concentrated in the Mn atom, while the *S4* state is characterised by the promotion of at least one unpaired electron to the conjugated π system of the equatorial ligand, in accordance to previous DFT and correlated wave function calculations[50,51,53,55].

*3.2.0.5 Oxidation state of Mn (OxSt)*Changes in the oxidation state of the metal centre account for about 8% of the variance in DB2, being among the largest sources of variance. This effect is somewhat diluted when variables referring to atoms 12 to 19 (*see* Figure 1) are included in the data set, and the same effect accounts for only about 5% of the variance in DB1. Again, PCA analysis of both data matrices rendered linear combinations of variables that accomplishes a satisfactory separation between the reduced Mn(salen) complexes and their oxidised oxo-Mn(salen) analogues (Figures 6a and 6c), with PC4 of DB2 being chosen as the best discriminant for this category. This separation becomes less pronounced upon rotation of the principal components, as depicted in Figures 6b and 6d.

Inspection of the loadings associated with the principal components of both data sets reveals a predominance of the natural charges, particularly those of N_2, N_3, O_4 and O_5, as well as those of C_6 and C_7. As expected, manganese (V) complexes present higher charges in the salen ligand. This behaviour is linked to the overall tendency of salen to act as a "charge reservoir" for the manganese atom, transferring part of its charge to the metal centre upon oxidation[55].

*3.2.0.6 Salen/acacen model (EqLig)*By design, the weight of this category is over estimated in DB1. The principal component analysis of DB2 clearly shows that the substitution of salen by acacen' has a profound impact on the properties of the system, accounting for about 8% of the total variance in the data. In both data sets, the separation between salen by acacen' was remarkable, as shown in Figure 7.

Not surprisingly, the largest coefficients in PC1 and RC1 of DB1 refer to variables describing atoms 12 to 19 (*see* Figure 1), thus justifying the use of DB2 in order to attain some insight on the physical consequences of choosing between acacen' and salen models. Despite this, the excellent separation between salen and acacen' entries in DB1, justifies the choice of PC1 of DB1 as the best discriminant for this category. Among the most prominent loadings associated with the PC3 and RC4 of DB2 are the angles centred in the metal centre ($A.O_4.Mn.N_2$, $A.O_5.Mn.N_3$, $A.N_2.Mn.N_3$ and $A.O_4.Mn.O_5$), as well as the chirality measurement *CCM.acacen*, the natural charges at C_8 and C_9 and the bond lengths between these atoms and their neighbours. In summary, the truncation of Mn(salen) complexes to Mn(acacen') yields significant distortions in the chemical environment around the metal centre, the consequences of which will be explored later in this work.

*3.2.0.7 Axial ligand (AxLig)*The variability introduced by the axial ligand of the Mn atom (or by the absence of such ligand) proved to be one of the most difficult challenges for PCA and rPCA, not only because of the relatively low variance that remained in the data once the components associated with *BasisSet*, *SpinSt*, *OxSt*, and *EqLig* were extracted, but also because this categorical variable encompasses four values (instead of the two or three values seen in the other variables). As a result, poor separations between the different categories of *AxLig* were obtained, as shown in Figure 8. Although not being able to achieve complete separation between classes of *AxLig*, the PC6 of DB2 (Figure 8c) suggests that the four classes are organised as four overlapping populations around four different modal values, being the component best associated with this category.

Using the PC6 of DB2 as the best possible discriminant that can be associated to *AxLig*, models without any axial ligand (**null**) and those coordinated with $CH_3CH_2NH^-$ (**amide**) appear to be the most extreme cases. This is in accordance to what would be expected as the amide ligand is a strong donor ligand capable of transferring a significant amount of charge to the manganese atom. The **cl** and the **amide** classes lie close

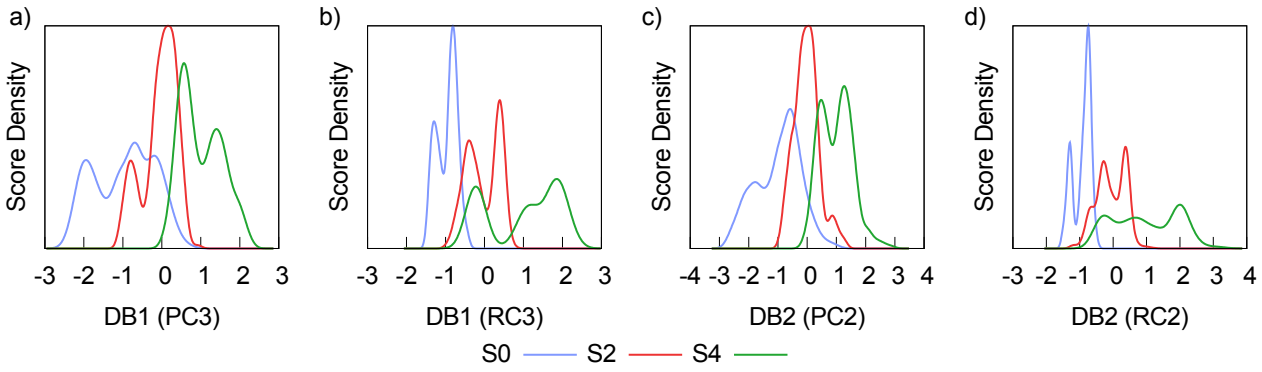

Fig. 5 Distribution of the scores of the principal and rotated components of DB1 and DB2 best associated with the choice of spin state in the DFT calculations.

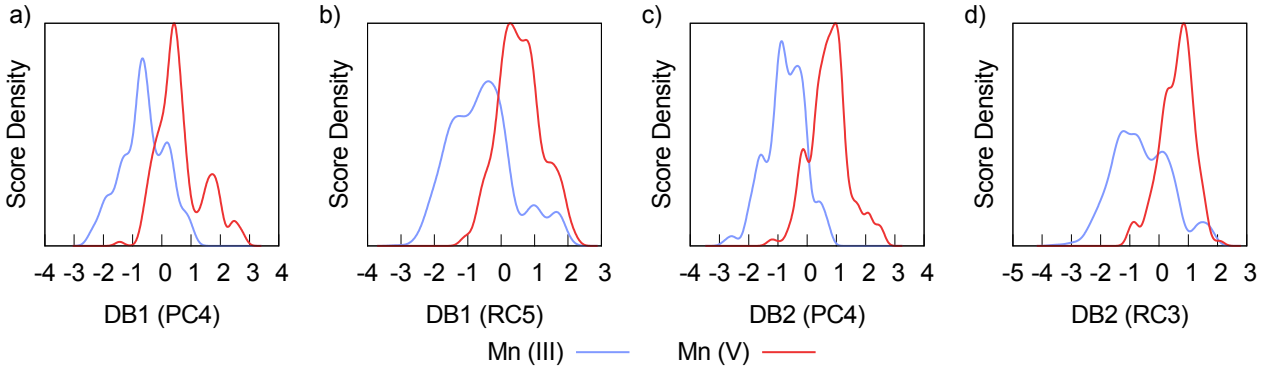

Fig. 6 Distribution of the scores of the principal and rotated components of DB1 and DB2 best associated with variations in the oxidation state of the manganese atom.

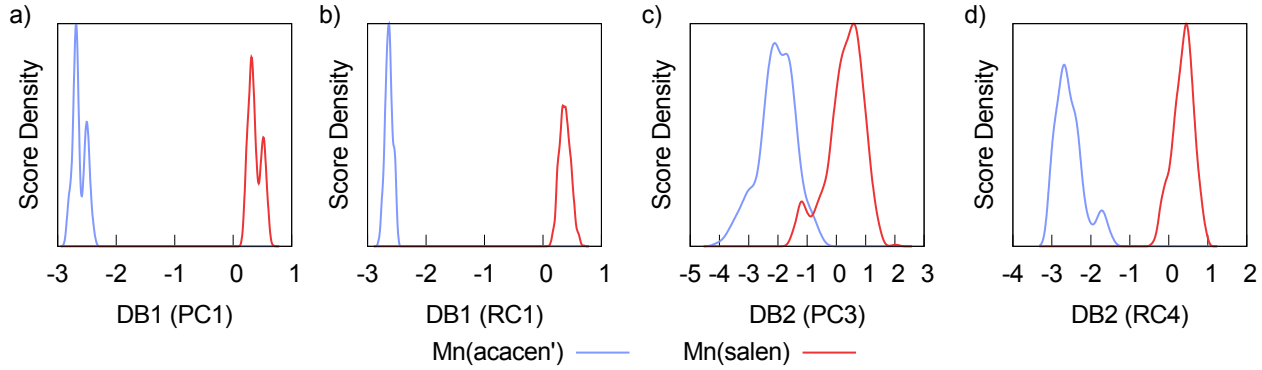

Fig. 7 Distribution of the scores of the principal and rotated components of DB1 and DB2 best associated with the variance introduced by the existence of salen and acacen' complexes in the data.

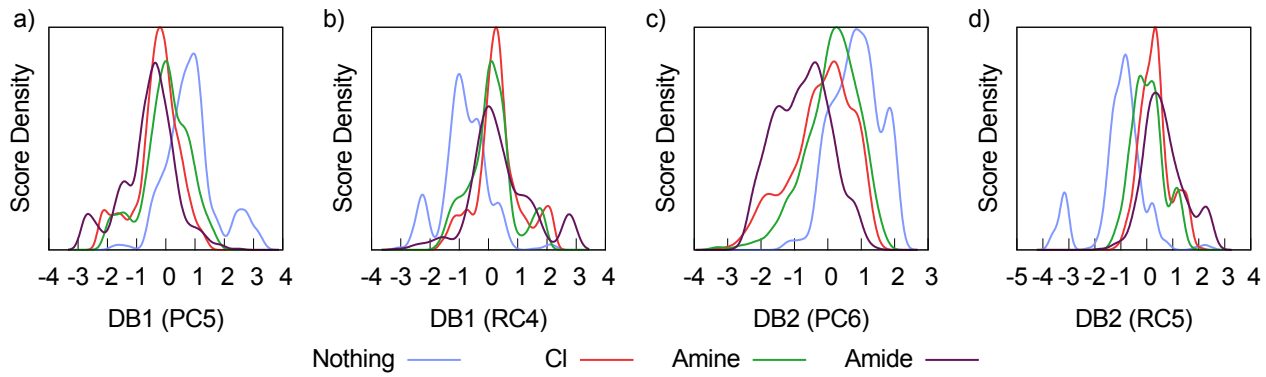

Fig. 8 Distribution of the scores of the principal and rotated components of DB1 and DB2 best associated with variations in the axial coordination site of Mn(salen) and Mn(acacen') complexes.

together, as well as those models belonging to the **amine** and **nothing** families. This results in a sorting between cationic and neutral systems.

The most prominent loadings in PC6 of DB2 are the overall charge of the equatorial ligand, as well as the natural charges of C_6 and C_7, as well as the natural spin densities in N_2 and N_3, D_P and also an assortment of angles and bond lengths among atoms Mn, N_2, N_3, C_{20} and C_{21}. The results show that coordination with a donor ligand in the axial position of Mn(salen) and Mn(acacen') complexes has a moderate effect in the geometry as well as in the charge and spin distribution around the metal centre, although the impact of such variable is dwarfed in comparison to other sources of variance which were included in this study.

*3.2.0.8 Diimine bridge (Bridge)*As shown in 1, this descriptor is best associated with PC8 and RC9 of DB1 and PC7 and RC7 of DB2, accounting for about 5% of the variance in the latter data set. As shown in 9, the effect of the diimine bridge is relatively subtle and the rotation of the principal components helped achieving a good separation between entries representing models with EDA, and models with CHDA. Both RC9 of DB1 and RC7 of DB2 allowed excellent classification of the entries in each data set, with RC9 of DB1 being chosen as the best discriminant to be associated with the descriptor *Bridge*.

An inspection of the loadings of RC9 of DB1 reveals that the charges of C_{20} and C_{21} bear the most preponderant coefficients, together with an assortment of bond lengths and angles among atoms Mn, N_2, N_3, C_{20} and C_{21}, as well as *CCM.acacen*. These results reflect the inductive effect exerted by the carbon atoms in CHDA, as well as the asymmetry imparted by the CHDA bridge in the equilibrium geometry of the equatorial ligand.

*3.2.0.9 Substitution at R_1*As show in Table 1, the substitution pattern at R_1 (the 3,3' positions in the salen ligand) accounts for only a small percentage of the variance in DB1. The removal of variables referring to atoms 12 to 19 renders the component analysis protocol useless in determining the effect of this variable on the properties of the different compounds, and no clear association was found between the principal components of DB2 and this category. Considering only the results obtained for DB1, PCA was unable to produce a meaningful separation among the different classes in this descriptor. Rotation of the principal components allowed the ordering of the different models in accordance with the substitution pattern in R_1, as shown in Figure 10. The results sugest that RC11 of DB1 would be a good descriptor for the R_1 category. An analysis of the loadings from this rotated component shows high coefficients for $Q.C_{12}$ and $Q.C_{13}$ as well as significant weights on $Q.O_4$ and $Q.O_5$, compatible with the inductive effect of the alkyl groups (Me and ^{t}Bu) in some of the salen models. Moreover, *CCM.acacen* attained significant weight in these loadings, showing the steric effect that substitutions at this position impart on the conformation of the equatorial ligand around the metal centre.

*3.2.0.10 Substitution at R_2*Finding a PC or RC that would account for the variance introduced by the substitution pattern at R_2 (the 5,5' positions in the salen ligand) was only accomplished in the analysis of DB1, as shown in Table 1. The effect of this category is subtle and PCA was unable to produce a meaningful separation among the different classes in this descriptor. Rotation of the principal components allowed the models without an alkyl group at their 5,5' position to be segregated from those with either methyl or *tert*-butyl (as well as the acacen' models), as depicted in 11b. Thus, RC10 of DB1 was chosen as the descriptor for R_1. The loadings associated with RC10 are dominated by the charges of C_{16}, C_{17} and, to a lesser extent, *CCM.acacen* and *CCM.salen*. These results correspond to a limited effect from this source of variance towards the overall geometry and electronic structure of

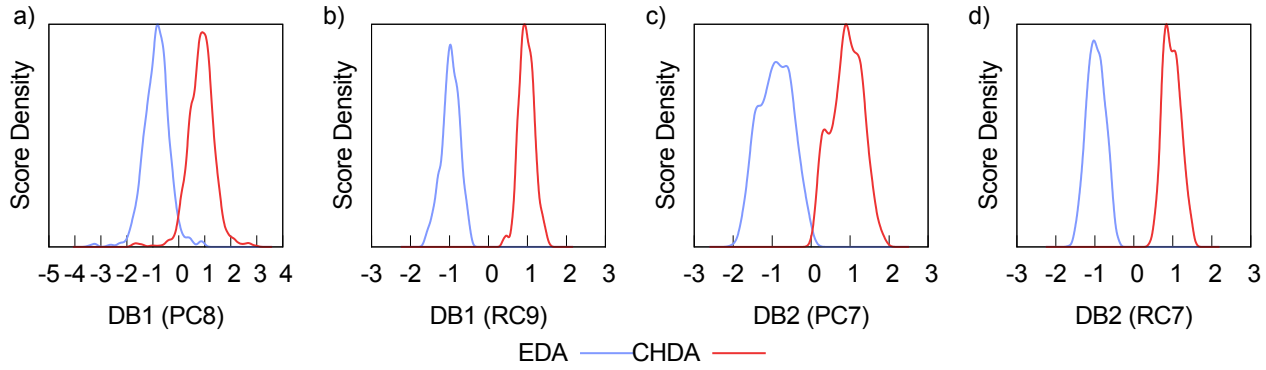

Fig. 9 Distribution of the scores of the principal and rotated components of DB1 and DB2 best associated with variations in the diimine bridge: etildiimine (EDA) *versus* cyclohexildiimine (CHDA).

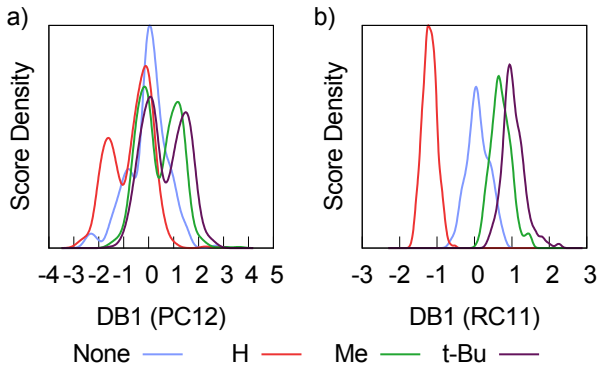

Fig. 10 Distribution of the scores of the principal and rotated components of DB1 best associated with the variance introduced by the substitution patterns in R_1.

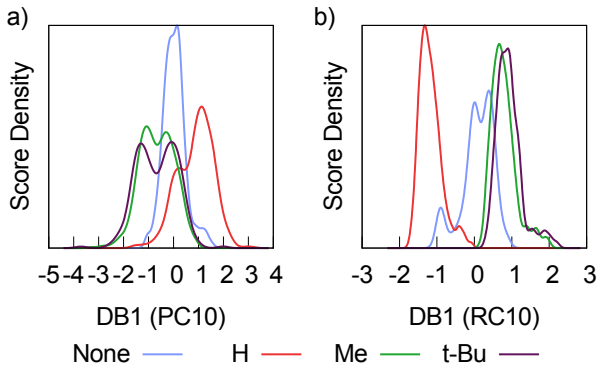

Fig. 11 Distribution of the scores of the principal and rotated components of DB1 best associated with the variance introduced by the substitution patterns in R_2.

the Mn(salen) complexes, which can be mostly explained by the inductive effect of the alkyl groups.

In the face of these results, it would seem reasonable to use either model **3** or model **4** as models for the reactivity of the larger Mn(salen) complexes. It is therefore plausible to hypothesise that the effect of these bulky substituents may be properly treated using a QM/MM strategy, although not with the same truncation scheme used by Cavallo and Jacobsen[25]. The present discussion, however, refers only to the effect of alkyl groups in the periphery of the salen scaffold and further work must be developed in order to access the effects of groups capable of greater interaction with the aromatic system of the salen ligand, such as OH or NO_2.

3.3 Linear Models

Having been able to provide some answers as to the importance and consequences of the use of truncated models in the study of the structure and reactivity of Mn(salen) catalysts, we tested whether the information collected from the factor analysis procedure is capable of shedding some light on the way these factors influence several properties of Mn(salen) complexes. For this purpose, a composite discriminant matrix $\mathbf{M}_{CD}$ was created, having the loadings associated with he best discriminants for each categorical variable (chosen earlier in this work) as columns. Application of $\mathbf{M}_{CD}$ to the data matrix in DB0 determined the discriminant scores, a_X, where X represents each of the categorical variables mentioned earlier.

These scores were used as independent variables in the development of linear models for $\Delta E_{0,4}$, ΔE_{ox} and $S.O_{\mathrm{ax}}$. For each linear model the significance of its coefficients was asserted using a bootstrap technique, implemented in the boot package for R[9,17]. In this procedure, an empirical distribution of the coefficient associated to each a_X was obtained from repeated random sampling with replacement from the data which was supplied to the linear regression routine, with 10^4 bootstrap replicates. A stepwise backward regression protocol was used in order to retain only the most significant coeffi-

cients: in the first iteration, the dependent variable was fitted against all independent variables. In the following steps, the least significant independent variable was removed from the data. This evaluation was made from the standardized error for the coefficients associated with each a_X, and the procedure stopped when only independent variables with a significance level above 99% were observed. Moreover, only the data regarding oxo-complexes were considered in devising linear models for properties regarding O_{ax}. In the same manner, the creation of linear models concerning natural spin densities only accounted the complexes in the *S2* and *S4* states. The relative weights of the coefficients obtained at the end of this procedure were evaluated using the relative weights method by Johnson[29].

Linear models for *S.Mn*, *Q.EqLig*, *S.EqLig*, $\Delta E_{0,4}$, ΔE_{ox} and ΔE_{epox} were developed, using the protocol described above. The expressions for each of these models can be constructed from the data in Table 2, which lists the coefficients obtained in each linear model, as well as their relative weights. The results from the fitting procedure can also be appreciated in Figure 12. In general, the linear models gave fair to good predictions. Linear models for *Q.Mn*, $Q.O_{ax}$, $S.O_{ax}$ and $\Delta E_{0,2}$ were also attempted, but none achieved an adjusted-r^2 ($r^2_{adj.}$) higher than 0.70.

Because *S.Mn*, *Q.EqLig* and *S.EqLig* were also accounted in the PCA procedure, the quality of their linear models provides a diagnose on how complete the variance in each of these three variables was scattered among the principal components that compose $\mathbf{M}_{CD}$. The results show that most of the variance in these variables was transferred to the different discriminants. On the other hand, the failed attempt to model *Q.Mn* suggest that this variable is not well represented in $\mathbf{M}_{CD}$. This may be partially caused by the low variance of *Q.Mn*, specially when considering only DefBas-3 and def2-TZVP entries in DB0. A similar situation is observed for $Q.O_{ax}$ and $S.O_{ax}$, for which most of the variance is due to def2-SV(P) calculations.

All models displayed in Table 2 and Figure 12 show a clear dependence on $a_{BasisSet}$, which again empathises the importance of choosing a proper basis set in this type of calculations. More interestingly, a_{AxLig} and a_{OxSt} carry significant weight to *Q.EqLig*, which reflects the model according to which the salen ligand acts as a "charge reservoir", donating electron density to the metal centre upon changes in the coordination sphere and/or oxidation state of the latter[55].

According to Table 2, the most influential contributions to ΔE_{ox} (apart from $a_{BasisSet}$) are a_{SpinSt}, and a_{AxLig}. The relatively large weight of a_{SpinSt} is likely due to the fact that this reaction in the *S4* state is much less exoenergetic than at either *S2* or *S0* (considering only the DefBas-3 and def2-TZVP data). At the same time, the dependence on a_{AxLig} suggests that a strong donor group at the axial position of Mn reduces the magnitude of ΔE_{ox} (for the triple-ζ basis sets, this quantity is always negative, ranging from -600 to -300 kJ.mol^{-1}). Conversely a strong donor axial group would make the epoxidation energetically more favourable, which is in accordance with the known experimental evidence[27]. As expected, the epoxidation of propene completes the catalytic cycle of these complexes and the linear model for ΔE_{epox} is virtually symmetric to that of ΔE_{ox}.

The linear model for $\Delta E_{0,4}$ is dominated by a_{OxSt} and a_{AxLig}. Since the scores a_{OxSt} are distributed in the same manner as that depicted in 6a, the results suggest that the energy gap between *S0* and *S4* is greater in the case of oxo-complexes. This is in excellent agreement with the basic electronic structure of these complexes. Reduced Mn(salen) complexes have a d^4 configuration in the manganese atom, allowing the existence of *S0*, *S2* and *S4* states within the space of the 3*d* orbitals. On the other hand, the manganese atom in the oxo-Mn(salen) complexes has a d^2 configuration, and an *S4* state is only attainable by excitation of an inner electron of the metal atom, or unpairing of two electrons from the salen valence shell. Nevertheless, a closer look at Figure 12f shows negative values for $\Delta E_{0,4}$, all of which come from calculations using the def2-SV(P). The model further predicts large contributions from a_{AxLig}, suggesting that a strong donor ligand may stabilise the *S4* state relative to *S0*. Unfortunately the relatively poor discrimination between the different axial ligands in the data set does not allow further exploration of this effect. Further attempts to model $\Delta E_{0,2}$ were made, although unsuccessfully. The near-degeneracy of the *S0* and *S2* states may lead to an inaccurate estimation of $\Delta E_{0,2}$ within the framework of DFT, thus invalidating any model that could be drawn from these results.

Finally, Table 2 shows that a_{EqLig} has surprisingly low weights in all models. This is in apparent contradiction with the results from PCA, in which *EqLig* is one of the major sources of variance. Such a result allows us to enlighten the results of Cavallo and Jacobsen, who concluded the general equivalence between salen and acacen' complexes[11,25], to the apparent contradictions found in latter works which were attributed to the use of different models[33,38]. Most of the validation of acacen' as a surrogate of the salen ligand was based on the ordering of the energies of the different spin states. The framework of DFT is possibly inaccurate for the estimation of $\Delta E_{0,2}$ (*S0* and *S2* being quasi-degenerate), and the results shown earlier show that $\Delta E_{0,4}$ is plausibly independent of the combination of atomic and molecular properties that best describe the differences between salen and acacen' complexes (a_{EqLig}). The same could be said for the energies involved either in the formation of the active oxo-complexes (ΔE_{ox}), or for the energy involved in the oxygen transfer from the oxo-complex to the substrate. It is therefore plausible that comparisons of Mn(salen) and Mn(acacen') complexes (and their re-

Table 2 Coefficients for the linear models describing *S.Mn*, *Q.EqLig*, *S.EqLig*, $\Delta E_{0,4}$, ΔE_{ox} and ΔE_{epox} , as well as their relative weights.

	S.Mn		*Q.EqLig*		*S.EqLig*		ΔE_{ox}		ΔE_{epox}		$\Delta E_{0,4}$	
(Intercept)	1.808		−0.774		1.012		−355.74		143.93		67.13	
$a_{BasisSet}$	−0.019	(37%)	−0.011	(27%)	0.021	(33%)	−6.33	(52%)	6.29	(52%)	3.41	(39%)
$a_{Functional}$	0.066	(8%)	−0.018	(9%)	0.064	(6%)	5.17	(7%)	−4.94	(7.5%)	—	
a_{SpinSt}	0.051	(7%)	−0.008	(2%)	0.084	(32%)	9.79	(20%)	−9.76	(20%)	−0.74	(¡1%)
a_{OxSt}	−0.103	(43%)	0.027	(11%)	0.071	(19%)	−5.12	(2%)	5.10	(2%)	17.00	(37%)
a_{EqLig}	0.005	(2%)	0.003	(1%)	—		1.26	(1%)	−1.25	(1%)	−1.72	(4%)
a_{AxLig}	—		0.086	(45%)	−0.021	(1%)	−17.5	(7%)	17.53	(7%)	13.75	(12%)
a_{Bridge}	0.020	(2%)	−0.022	(2%)	0.031	(1%)	8.18	(2%)	−8.22	(2%)	−4.35	(2%)
a_{R1}	—		0.028	(1%)	0.018	(3%)	2.92	(3%)	−2.84	(3%)	−4.93	(2%)
a_{R2}	—		0.037	(2%)	0.011	(5%)	4.19	(4%)	−4.14	(4%)	−6.91	(4%)
$r^2_{adj.}$	0.7896		0.7597		0.9712		0.8401		0.8381		0.7423	

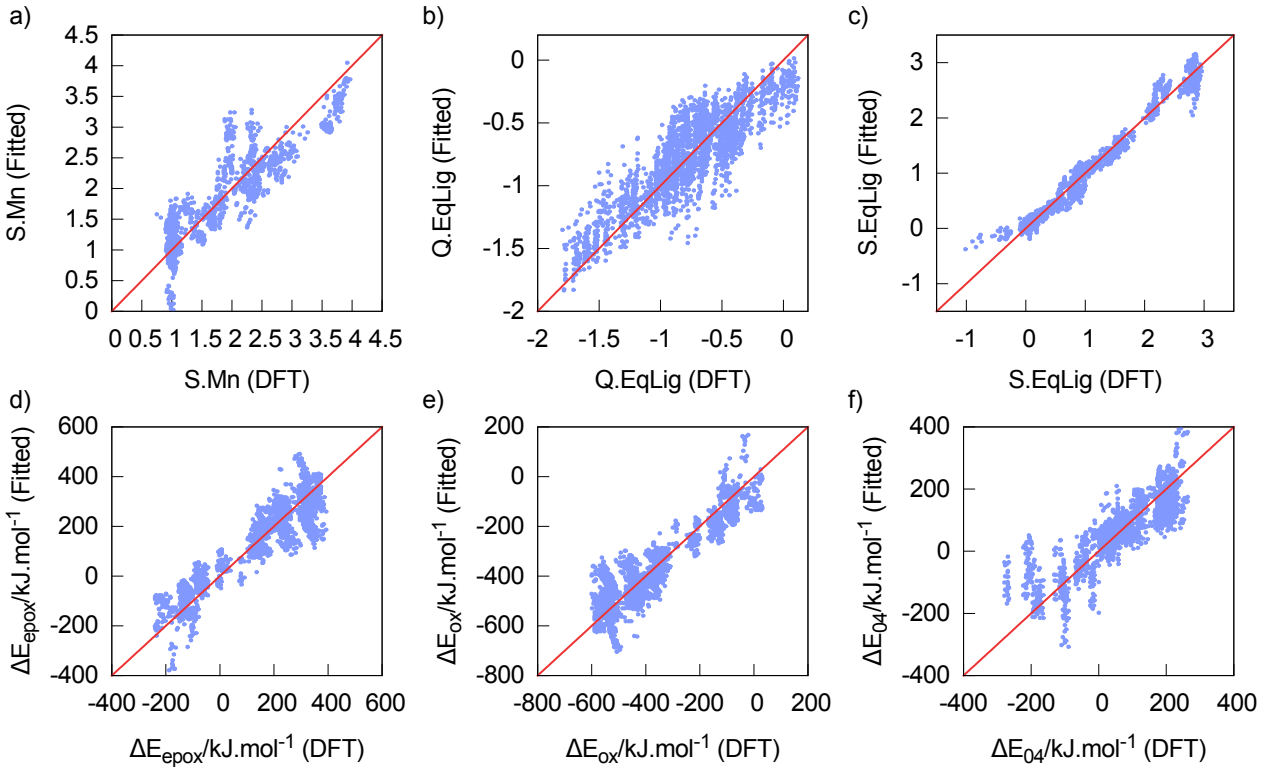

Fig. 12 Plots of the computed (DFT) versus fitted values for the linear regression models of *S.Mn* (a), *Q.EqLig* (b), *S.EqLig* (c), $\Delta E_{0,4}$ (d), ΔE_{ox} (e), and ΔE_{epox} (f). In each plot the identity function is displayed for comparison.

spective oxo- derivatives) based on energy differences would not account for any significant difference between these two families of complexes. Our results also show that other characteristics of both Mn(salen) and Mn(acacen') complexes may be relatively independent of the choice of the equatorial ligand, at their equilibrium geometries. Nonetheless, *EqLig* is still one of the major contributors to the variance in the data considered and its contribution is likely to be noticed when accounting for the interaction between the oxo-complexes and their substrates, or away from their equilibrium geometries. This conjecture may justify the discrepancies between the study of the catalytic cycle of Mn(salen) species by Cavallo and Jacobsen (using Mn(acacen') models)[12,14,26,27] and the latter work by Morokuma and co-workers, who used **m03-cl** as their model compound, as well as a more flexible basis set[32].

4 Conclusions

In this work, PCA was used to successfully associate structural features of different Mn(salen) and Mn(acacen') compounds to the major sources of variation in the data describing their geometrical features, as well as the charge and spin distribution across the metal-salen or metal-acacen' system, using different basis sets and density functions.

The results show that the choice of a triple-ζ basis set is paramount for a correct description of this type of systems. In a cost *versus* accuracy comparison, the use of a triple-ζ basis set with omission of polarization functions on the hydrogen atoms (DefBas-3) offers an accuracy comparable to that of a properly balanced triple-ζ basis set (def2-TZVP) with only small computational cost over the use of a properly balanced split valence one (def2-SV(P)).

With respect to the adequacy of using acacen' as a truncated model for salen in the study of Mn(salen) complexes, the results from factor analysis point that this variation in the data set may account for about 9% of the total variance, being surpassed by the choice of basis set and spin state of the complexes (*S0*, *S2* or *S4*). The results also show that the use of acacen' as a surrogate model for salen bears a strong effect on the charge and spin distribution among the molecule, as well as in a possible misrepresentation of the local geometry around the metal centre. Given the evolution of information technology, and the development of linear scaling DFT techniques, the use of Mn(acacen') as a surrogate model for the larger Mn(salen) complexes is discouraged, as well as the use of split valence basis sets.

The results also suggest that the nature of the diimine bridge and the presence of methyl or *tert*-butyl groups at the 3,3' and 5,5' positions of the salen scaffold, despite being important for the reactivity and enantioselectivity of the larger Mn(salen) catalysts, bear only some limited influence on the electronic structure of the salen ligand.

The results further allowed to shed some light on the effect of using APTES as the axial ligand when anchoring Mn(salen) complexes onto a supporting material. The structural features of compounds coordinated to a model for the neutral form of APTES are similar to those of the chloride-coordinated compounds usually found in homogeneous media, bearing no clear distinction under factor analysis. On the other hand, coordination to a model for the anionic form of APTES yields some alterations in the overall structure of the metal-salen system, specially with respect to the charge distribution and chiral content of the ligand. Furthermore, the results suggest that this feature strongly affects the energetic balance of the oxidation of Mn(salen) and Mn(acacen) compounds to their oxo- derivatives, thus deserving further study.

5 Acknowledgements

This work has been supported by Fundao para a Ciłncia e Tecnologia (FCT) through Grant No. PEst-C/EQB/LA0006/2011 and Project NORTE-07-0124-FEDER-000067 - NANOCHEMISTRY. FT further acknowledges FCT for a doctoral Grant (SFRH/BD/64314/2009).

Additional Supporting Information may be found in the online version of this article.

References

1 Y. G. Abashkin and S. K. Burt. (salen)mn-catalyzed epoxidation of alkenes: A two-zone process with different spin-state channels as suggested by dft study. *Org. Lett.*, 6:59–62, 2004.

2 Y. G. Abashkin and S. K. Burt. (salen)mn(iii) compound as a nonpeptidyl mimic of catalase: Dft study of the metal oxidation by a peroxide molecule. *J. Phys. Chem. B*, 108:2708–2711, 2004.

3 Y. G. Abashkin and S. K. Burt. (salen)mniii compounds as nonpeptidyl mimics of catalase. mechanism-based tuning of catalase activity: A theoretical study. *Inorg. Chem.*, 44:1425–1432, 2005.

4 Y. G. Abashkin, J. R. Collins, and S. K. Burt. (salen)mn(iii)-catalyzed epoxidation reaction as a multichannel process with different spin states. electronic tuning of asymmetric catalysis: A theoretical study. *Inorg. Chem.*, 40:4040–4048, 2001.

5 B. K. Alsberg, V. R. Jensen, and K. J. Borve. Use of multivariate methods in the analysis of calculated reaction pathways. *J. Comput. Chem.*, 17(10):1197–1216, July 1996.

6 S. Alvarez, P. Alemany, and D. Avnir. Continuous chirality

measures in transition metal chemistry. *Chem. Soc. Rev.*, 34(4):313–326, 2005.

7 H. Attias. Independent factor analysis. *Neural Comput.*, 11:803–851, 1999.

8 Carlos Baleizão and Hermenegildo Garcia. Chiral salen complexes: an overview to recoverable and reusable homogeneous and heterogeneous catalysts. *Chem. Rev.*, 106(9):3987–4043, 2006. PMID: 16967927.

9 Angelo Canty and B. D. Ripley. *boot: Bootstrap R (S-Plus) Functions*, 2014. R package version 1.3-11.

10 L. Cavallo and H. Jacobsen. Radical intermediates in the jacobsen – katsuki epoxidation. *Angew. Chem. Int. Ed.*, 39(3):589–592, February 2000.

11 L. Cavallo and H. Jacobsen. Electronic effects in (salen)mn-based epoxidation catalysts. *J. Org. Chem.*, 68(16):6202–6207, July 2003.

12 L. Cavallo and H. Jacobsen. Manganese-salen complexes as oxygen-transfer agents in catalytic epoxidations - a density functional study of mechanistic aspects. *Eur. J. Inorg. Chem.*, 2003:892–902, 2003.

13 L. Cavallo and H. Jacobsen. Transition metal mediated epoxidation as test case for the performance of different density functionals: A computational study. *J. Phys. Chem. A*, 107:5466–5471, 2003.

14 L. Cavallo and H. Jacobsen. Toward a catalytic cycle for the mn—salen mediated alkene epoxidation: A computational approach. *Inorg. Chem.*, 43:2175–2182, 2004.

15 R. Climaco-Pinto, A.S. Barros, N. Locquet, L. Schmidtke, and D.N. Rutledge. Improving the detection of significant factors using anova-pca by selective reduction of residual variability. *Anal. Chim. Acta*, 653(2):131 – 142, 2009.

16 P. Das, A. R. Silva, A. P. Carvalho, J. Pires, and C. Freire. Enantioselective epoxidation of alkenes by jacobsen catalyst anchored onto aminopropyl-functionalised laponite, mcm-41 and fsm-16. *Catal. Lett.*, 129:367–375, 2009.

17 A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge, 1997. ISBN 0-521-57391-2.

18 K. Eichkorn, O. Treutler, H. Ohm, M. Haser, and R. Ahlrichs. Auxiliary basis sets to approximate coulomb potentials. *Chem. Phys. Lett.*, 240(4):283–290, June 1995.

19 K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs. Auxiliary basis sets for main row atoms and transition metals and their use to approximate coulomb potentials. *Theor. Chem. Acc.*, 97(1-4):119–124, 1997.

20 D. Feichtinger and D. A. Plattner. Direct proof for o=mn(salen) complexes. *Angew. Chem. Int. Ed. Engl.*, 36(16):1718–1719, September 1997.

21 Cristina Freire, Clara Pereira, and Susana Rebelo. Green oxidation catalysis with metal complexes: from bulk to nano recyclable hybrid catalysts. In *Catalysis: Volume 24*, volume 24, pages 116–203. The Royal Society of Chemistry, 2012.

22 H. Gaspar, M. Andrade, C. Pereira, A. M. Pereira, S. L. H. Rebelo, J. P. Araujo, J. Pires, A. P. Carvalho, and C. Freire. Alkene epoxidation by manganese(iii) complexes immobilized onto nanostructured carbon cmk-3. *Catal. Today*, 203:103–110, March 2013.

23 E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, and F. Weinhold. Nbo 5.9. Theoretical Chemistry Institute, University of Wisconsin, Madison, WI, 2009; http://www.chem.wisc.edu/ nbo5.

24 K. Heberger and A. Lopata. Assessment of nucleophilicity and electrophilicity of radicals, and of polar and enthalpy effects on radical addition reactions. *J. Org. Chem.*, 63(24):8646–8653, November 1998.

25 H. Jacobsen and L. Cavallo. A possible mechanism for enantioselectivity in the chiral epoxidation of olefins with [mn(salen)] catalysts. *Chem. Eur. J.*, 8(4):800–807, February 2001.

26 H. Jacobsen and L. Cavallo. Re-evaluation of the mn(salen) mediated epoxidation of alkenes by means of the b3lyp* density functional. *Phys. Chem. Chem. Phys.*, 6:3747–3753, 2004.

27 H. Jacobsen and L. Cavallo. Donor-ligand effect on the product distribution in the manganese-catalyzed epoxidation of olefins: A computational assessment. *Organometalics*, 25:177–183, 2006.

28 Z. Jia, X. Fu, Y. Lu, H. Zhang, X. Huang, and H. Wu. Chiral salen mn(iii) complex axial coordination immobilized on diamine modified alspp and highly efficient large-scale epoxidation of olefins as a recoverable catalyst. *J. Inorg. Organomet. Polym.*, 22:415–422, 2012.

29 Jeff W. Johnson. A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivar. Behav. Res.*, 35(1):1–19, 2000.

30 I.T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, New York, 2nd edition, 2002.

31 M. Karelson, V. S. Lobanov, and A. R. Katritzky. Quantum-chemical descriptors in qsar/qspr studies. *Chem. Rev.*, 96(3):1027–1044, May 1996.

32 I. V. Khavrutskii, D. G. Musaev, and K. Morokuma. Epoxidation of unfunctionalized olefins by mn(salen) catalyst using organic peracids as oxygen source: A theoretical study. *Proc. Natl. Acad. Sci. U.S.A.*, 101(16):5743–5748, April 2004.

33 L. Kürti, M. M. Blewett, and E. J. Corey. Origin of enantioselectivity in the jacobsen epoxidation of olefins. *Org. Lett.*, 11:4592–4595, 2009.

34 C. Linde, B. Akermark, P. Norrby, and M. Svensson. Timing is critical: Effect of spin changes on the diastereose-

lectivity in mn(salen)-catalyzed epoxidation. *J. Am. Chem. Soc.*, 121:5083–5084, 1999.

35 C. Linde, N. Koliaï, P. Norrby, and B. Akermark. Experimental evidence for multiple oxidation pathways in the (salen)mn-catalyzed epoxidation of alkenes. *Chem. Eur. J.*, 8(11):2568–2573, June 2002.

36 L. Lou, S. Jiang, K. Yu, Z. Gu, R. Ji, Y. Dong, and S. Liu. Mesoporous silicas functionalized with aminopropyl via co-condensation: Effective supports for chiral mn(iii) salen complex. *Microporous Mesoporous Mater.*, 142:214–220, 2011.

37 L. Ma, F. Su, W. Guo, S. Zhang, Y. Guo, and J. Hu. Epoxidation of styrene catalyzed by mesoporous propylthiol group-functionalized silica supported manganese(iii) salen complexes with different pore morphologies. *Microporous Mesoporous Mater.*, 169:16–24, March 2013.

38 K. Malek, A. P. J. Jansen, C. Li, and R. A. van Santen. Enantioselectivity of immobilized mn-salen complexes: A computational study. *J. Catal.*, 246:127–135, 2007.

39 E. M. McGarrigle and D. G. Gilheany. Chromium and manganese-salen promoted epoxidation of alkenes. *Chem. Rev.*, 105(5):1563–1602, April 2005.

40 Y. Minenkov, A. Singstad, G. Occhipinti, and V. R. Jensen. The accuracy of dft-optimized geometries of functional transition metal compounds: a validation study of catalysts for olefin metathesis and other reactions in the homogeneous phase. *Dalton Trans.*, 41:5526–5541, 2012.

41 F. Neese. The orca program system. *WIREs Comput. Mol. Sci.*, 2(1):73–78, 2012.

42 G. Occhipinti, H. Bjorsvik, and V. R. Jensen. Quantitative structure—activity relationships of ruthenium catalysts for olefin metathesis. *J. Am. Chem. Soc.*, 128(21):6952–6964, 2006.

43 G. A. E. Oxford, D. Dubbeldam, L. J. Broadbelt, and R.Q. Snurr. Elucidating steric effects on enantioselective epoxidation catalyzed by (salen)mn in metal-organic frameworks. *J. Mol. Catal. A: Chem.*, 334:89–97, 2011.

44 M. Pinsky and D. Avnir. Continuous symmetry measures. 5. the classical polyhedra. *Inorg. Chem.*, 37(21):5575–5582, 1998.

45 R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

46 M. Randi. On characterization of chemical structure. *J. Chem. Inf. Comput. Sci.*, 37:672–687, 1997.

47 William Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois, 2013. R package version 1.3.10.

48 Joo A. Rodrigues, Antnio S. Barros, Beatriz Carvalho, Tiago Brando, Ana M. Gil, and Antnio C. Silva Ferreira. Chapter 81 - evaluation of beer deterioration by gas chromatography-mass spectrometry/multivariate analysis: A rapid tool for assessing beer composition. In Vicente FerreiraRicardo Lopez, editor, *Flavour Science*, pages 435 – 440. Academic Press, San Diego, 2014.

49 A. Schaefer, H. Horn, and R. Ahlrichs. Fully optimized contracted gaussian basis sets for atoms li to kr. *J. Chem. Phys.*, 97(4):2571–2577, 1992.

50 J. S. Sears and C. D. Sherrill. The electronic structure of oxo-mn(salen): Single-reference and multireference approaches. *J. Chem. Phys.*, 124(14):144314, 2006.

51 J. S. Sears and C. D. Sherrill. Assessing the performance of density functional theory for the electronic structure of metal-salens: The d2-metals. *J. Phys. Chem. A*, 112:6741–6752, 2008.

52 Y. Solomon and D. Avnir. Continuous symmetry measures: Finding the closest c2-symmetric object or closest reflection-symmetric object using unit quaternions. *J. Comput. Chem.*, 20(8):772–780, June 1999.

53 T. Takatani, J. S. Sears, and C. D. Sherrill. Assessing the performance of density functional theory for the electronic structure of metal-salens: The m06 suite of functionals and the d4-metals. *J. Phys. Chem. A*, 114:11714–11718, 2010.

54 J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria. Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.*, 91(14):146401–146405, 2003.

55 F. Teixeira, R. A. Mosquera, A. Melo, C. Freire, and M. N. D. S. Cordeiro. Charge distribution in mn(salen) complexes. *Int. J. Quantum Chem.*, 114(8):525–533, Apr. 2014.

56 F. Weigend and R. Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7(18):3297–3305, 2005.

57 S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometr. Intell. Lab.*, 2(1-3):37–52, August 1987.

58 Sebastian Wouters, Thomas Bogaerts, Pascal Van Der Voort, Veronique Van Speybroeck, and Dimitri Van Neck. Communication: Dmrg-scf study of the singlet, triplet, and quintet states of oxo-mn(salen). *J. Chem. Phys.*, 140(24):241103, 2014.

59 X. Xu and W. A. Goddard III. The x3lyp extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proc. Natl. Acad. Sci. U.S.A.*, 101:2673–2677, 2004.

60 X. Xu, Q. Zhang, R. P. Muller, and W. A. Goddard III. An extended hybrid density functional (x3lyp) with improved descriptions of nonbond interactions and thermo-

dynamic properties of molecular systems. *J. Chem. Phys.*, 122:014105, 2005.

61 H. Zabrodsky and D. Avnir. Chirality continuous symmetry measures. 4. chirality. *J. Am. Chem. Soc.*, 117(1):462–473, 1995.

62 H. Zabrodsky, S. Peleg, and D. Avnir. Continuous symmetry measures. *J. Am. Chem. Soc.*, 114(20):7843–7851, September 1992.