Ab Initio Quantum Molecular Similarity Measures on Metal-Substituted Carbonic Anhydrase (M^{II}CA, M = Be, Mg, Mn, Co, Ni, Cu, Zn, and Cd)

Miquel Solà, Jordi Mestres, Miquel Duran, and Ramon Carbó[*]

Institute of Computational Chemistry and the Department of Chemistry, Universitat de Girona, Plaça de l'Hospital 6, 17071 Girona, Catalonia, Spain

Quantum molecular similarity measures are used to analyze the variation of geometric and electronic structure of the first shell in a simplified model of the carbonic anhydrase active site with metal substitution. The experimental fact showing that the native zinc enzyme and the cobalt-substituted enzyme are comparably active is well reproduced by our calculations. This result encourages use of quantum molecular similarity measures as a powerful tool in molecular engineering.

1. INTRODUCTION

Carbonic anhydrase (CA) is a zinc metalloenzyme that catalyzes the reversible hydration of CO_2 to yield the bicarbonate anion with extreme efficiency.[1-6] X-ray data[5,7-11] show that the zinc atom placed at the active site of CA is bound to three imidazole groups coming from His94, His96, and His119, whereas a water molecule completes a nearly symmetrical tetrahedral coordination geometry. It is almost universally accepted that proton transfer from this water molecule to His64 generates the zinc-bound hydroxide ion, which is the nucleophile in this catalytic reaction. It is found that this intramolecular proton transfer is the rate-limiting process at high buffer concentrations.[1,2,12-15]

Substitution of the native zinc in CA by another metal has been used to shed light on the nature of the mechanism acting in this catalytic process. To gain more insight into the structure of CA, the naturally occurring zinc has been replaced by high-spin Co(II) (CoCA),[16-20] which is recognized as a powerful spectroscopic probe in biological systems.[21] In this case, it is assumed that the structural properties of CoCA deduced from spectra can be transferred to the native enzyme. Likewise, CoCA[22] and the manganese-substituted CA[23] (MnCA) have been used as models for studying the kinetic properties of native CA. Furthermore, in order to obtain information on the CA mechanism, ^{13}C NMR T_1 data have been recorded in the presence of the inactive copper- and manganese-substituted CA.[24-26] CoCA and CuCA complexes have also been used together with the electron spin resonance technique to study selected aspects of the catalytic cycle of CA.[19] In fact, zinc can be replaced in CA by man other metal dications such as cobalt, copper, nickel, manganese, cadmium, and mercury.[1,27-30] Nevertheless, it has been found that only Co(II) and Zn(II) confer appreciable activity to the enzyme.[1,27] As a matter of fact, Tu and Silverman[22] found that the rate constants for the rate-limiting step in CoCA and ZnCA are of comparable magnitude, being 7×10^5 s^{-1} for the Co(II)-substituted enzyme and 2×10^6 s^{-1} for the native enzyme. Interestingly, the coordination flexibility of Zn(II) and Co(II) with respect to other dications has been considered determinative of the activity of CA with Zn(II) and Co(II) dications.[1] The same behavior with respect to metal substitution has been observed in other zinc enzymes.[31,32]

To the best of our knowledge and from a theoretical point of view (see ref 6 for a recent review on theoretical studies of

the CA enzyme), there are only three papers addressing the effect of metal substitution at the CA active site. In the first one, Vedani and Huhta[33] performed molecular mechanics calculations to predict structures from Zn(II)- and Co(II)-substituted human carbonic anhydrase I (HCAI). Their main conclusion was that the ZnHCAI and CoHCAI enzymes show no principal structural differences, thus explaining experimental facts. In the second paper, Garmer and Krauss[34] analyzed the variation with metal substitution of both the coordination number at the active site of CA and the proton affinity of OH$^-$ bound to the metal cation. These authors found a different behavior for the coordination number of the first shell in the high-pH active form when one compares ZnCA and CoCA with respect to CdCA and MnCA. In the last paper,[35] the same authors discussed the effect of the first-shell geometry and coordination number on the experimentally observed visible spectra.

Although the aforementioned studies lead to comparisons between metal-substituted CA enzymes, a quantitative global comparison is still missing. Such a comparison, which is the ultimate purpose of this paper, will determine whether the mere structural and electronic properties of a first-shell model of the CA active site can be used to explain the different behavior of the metal-substituted CA. One particular aim of this paper is to carry out quantum molecular similarity measures in order to classify a series of eight dications which have been used in metal-substituted experimental and theoretical models of CA.

The aforementioned quantum molecular similarity measures (QMSM)[36-49] have been defined very recently as a new branch of computer-aided molecular design methodologies. Topological approaches,[36-39] graph theory,[40] and quantum mechanics[41-49] have been used to quantify molecular similarity, whose most widespread definition is that reported first by Carbó *et al.*,[47] based on the assumption that similar molecules must have similar electron distributions. These authors propose that if the molecules under comparison, A and B, have electron densities D_A and D_B, respectively, then their similarity is given by the following expression:[48]

$$Z_{A,B}(\Theta) = \int\int D_A(\mathbf{r}_1) \, \Theta(\mathbf{r}_1,\mathbf{r}_2) \, D_B(\mathbf{r}_2) \, d\mathbf{r}_1 d\mathbf{r}_2 \qquad (1)$$

where $\Theta(\mathbf{r}_1,\mathbf{r}_2)$ is a positive definite operator depending on two-electron coordinates. In the particular case that $\Theta(\mathbf{r}_1,\mathbf{r}_2)$ is the Dirac delta function $\delta(\mathbf{r}_1-\mathbf{r}_2)$, substitution of this function in eq 1 yields the formula to calculate the overlap-like similarity. Likewise, when $\Theta(\mathbf{r}_1,\mathbf{r}_2)$ is substituted by $1/r_{12}$ in

eq 1, the expression to compute Coulomb-like similarities is obtained. Once the similarity measure has been defined, a similarity index (SI)[47] can be calculated as a generalized cosine of the angle between functions D_A and D_B in eq 1:

$$I_{A,B} = Z_{A,B}(Z_{A,A}Z_{B,B})^{-(1/2)} \qquad (2)$$

For any pair of compared systems, this cosine-liked index has a value ranging from 0 (total dissimilarity) through 1 (complete similarity), depending on the similarity associated with the two molecules. Finally, the Euclidean distance between the molecular electronic distributions of A and B can be defined as follows:[47]

$$d_{A,B} = [Z_{A,A} + Z_{B,B} - 2Z_{A,B}]^{1/2} \qquad (3)$$

It is generally found that cosine-like similarity indices and distances vary inversely one to another,[47] in such a way that two objects completely similar with a similarity index 1 have a distance of 0 between them.

In this work, eqs 1, 2, and 3 are evaluated for eight representative dications in order to predict theoretically their activity in metal-substituted CA enzymes. In particular, the Be, Mg, Mn, Co, Ni, Cu, Zn, and Cd dications have been chosen to carry out our study. In the following section we sketch the methodology needed to perform calculations in practice, whereas in the third section we analyze the results encountered and, when possible, compare them to previous experimental and theoretical studies.

2. METHODOLOGY

It is well-known that the main drawback of exact QMSM is the need to evaluate costly four-centered integrals, as found in eq 1. In order to circumvent this problem we use a fitted density expanded into a set of N_{func} Gaussian functions instead of the exact density in eq 1. If N_{basis} is taken as the number of basis functions, then the similarity evaluation using fitted densities results in an N^2_{func}-dependent process, as opposed to the N^2_{basis}-dependent process when exact densities are used. It has been shown[50] that this methodology gives results with a mean deviation error smaller than 0.5% and with important savings in CPU computing time.

The methodology used to fit the exact density in order to compute QMSM has been already reported.[50,51] In summary, an approximate density ($\rho(\mathbf{r})$) is least-squares fitted to the exact charge density ($D(\mathbf{r})$) by minimizing the integral

$$\Delta(\Theta) = \int (D(\mathbf{r}_1) - \rho(\mathbf{r}_1)) \, \Theta(\mathbf{r}_1, \mathbf{r}_2) \, (D(\mathbf{r}_2) - \rho(\mathbf{r}_2)) \, d\mathbf{r}_1 \, d\mathbf{r}_2 \qquad (4)$$

where $\Theta(\mathbf{r}_1, \mathbf{r}_2)$ is a positive definite operator depending on two-electron coordinates. In the present study $\Theta(\mathbf{r}_1, \mathbf{r}_2)$ in eq 4 has been chosen to be the Dirac delta function $\delta(\mathbf{r}_1 - \mathbf{r}_2)$. Earlier studies have shown[51] that fits of the exact density using the Dirac delta function operator or the Coulomb operator lead only to very small differences in calculated QMSM. During the fitting procedure, minimization of the integral in eq 4 is subject to the following constraint, necessary to guarantee charge conservation.

$$\int \rho(\mathbf{r}) \, d\mathbf{r} = N_e \qquad (5)$$

N_e being the total number of electrons and $\rho(\mathbf{r})$ being expanded into a set of Gaussian functions as

$$D_I(\mathbf{r}) \simeq \rho_I(\mathbf{r}) = \sum_{i \in I}^{N_{func}} a_i g_i(\mathbf{r}) \qquad (6)$$

The fitting procedure employed in this work has used the so-called BSA method,[50] where the set of $N_{func}\{g_i\}$ functions has been chosen to be the same as the squared s-type renormalized basis functions and where all integrations are performed analytically over all Cartesian space.

Throughout this paper the calculated QMSM are Coulomb-like similarities, *i.e.*, $\Theta(\mathbf{r}_1, \mathbf{r}_2) = 1/r_{12}$ in eq 1. This kind of QMSM has the enormous advantage of reducing the core dependence, which is one of the main drawbacks of the overlap-like similarities. It can be shown[51] that once the mutual orientation of the two systems being compared has been optimized to maximize the QMSM, very small displacements of one system bring about radical changes in the value of the overalp-like similarity, making the optimization process fairly complicated. Fortunately, this problem disappears upon use of the Coulomb operator in eq 1.

Owing to computational limitations, the model used to represent the first shell of CA is the $(NH_3)_3M^{II}(H_2O)$ system. This complex has been widely used in *ab initio* calculations of different mechanistic aspects of the CA enzyme.[52–57] Furthermore, the validity of this model, where the three imidazole ligands are substituted by three ammonia groups, has been previously assessed.[34,56] In the present paper, full geometry optimizations of all eight minima with no symmetry constraints were performed using Schlegel's method.[58] Wave functions using pseudopotentials to account for core electrons have been preferred to full-electron calculations, because valence similarity can be more easily related to reactivity than all-electron similarity.[44,51] Therefore, Dunning' valence double-ζ quality basis set[59,60] together with Hay and Wadt ECP's[61,62] to represent the core electrons (basis set labeled *lanl1dz* in the Gaussian-90 program) was used in all calculations.

Limitations in the model used are necessary in order to keep the process of optimizing the geometry within a reasonable computational cost. One must bear in mind that QMSM calculations require far less computing time than the geometry optimizations of the eight complexes studied. These optimizations *must be necessarily* performed; otherwise if one equates the geometry of the eight complexes studied to the geometry of the optimized $(NH_3)_3Zn^{II}(H_2O)$ system, and then one performs QMSM calculations, the relationship between all complexes turns out to be the same as that given by the periodic table. For instance, in this case, the native enzyme becomes closer to Cu(II)-substituted CA than to Co(II)-substituted CA, in clear contradiction with the experimental results.[2,19]

All geometry optimizations and calculations of the SCF wave functions from which the electron density is fitted were carried out with the help of the Gaussian-90 system of programs.[63] Calculations of QMSM using fitted densities were done by means of our own program,[64] which allows the optimization of the mutual orientation of the two systems studied in order to maximize their similarity by any of the commonplace steepest-descent, Newton, and quasi-Newton algorithms.[65] Such a program is fully interfaced with the Gaussian-90[63] suite of programs.

3. RESULTS AND DISCUSSION

The fully optimized geometries of the eight $(NH_3)_3M^{II}$-(H_2O) complexes studied are depicted in Figure 1. The first
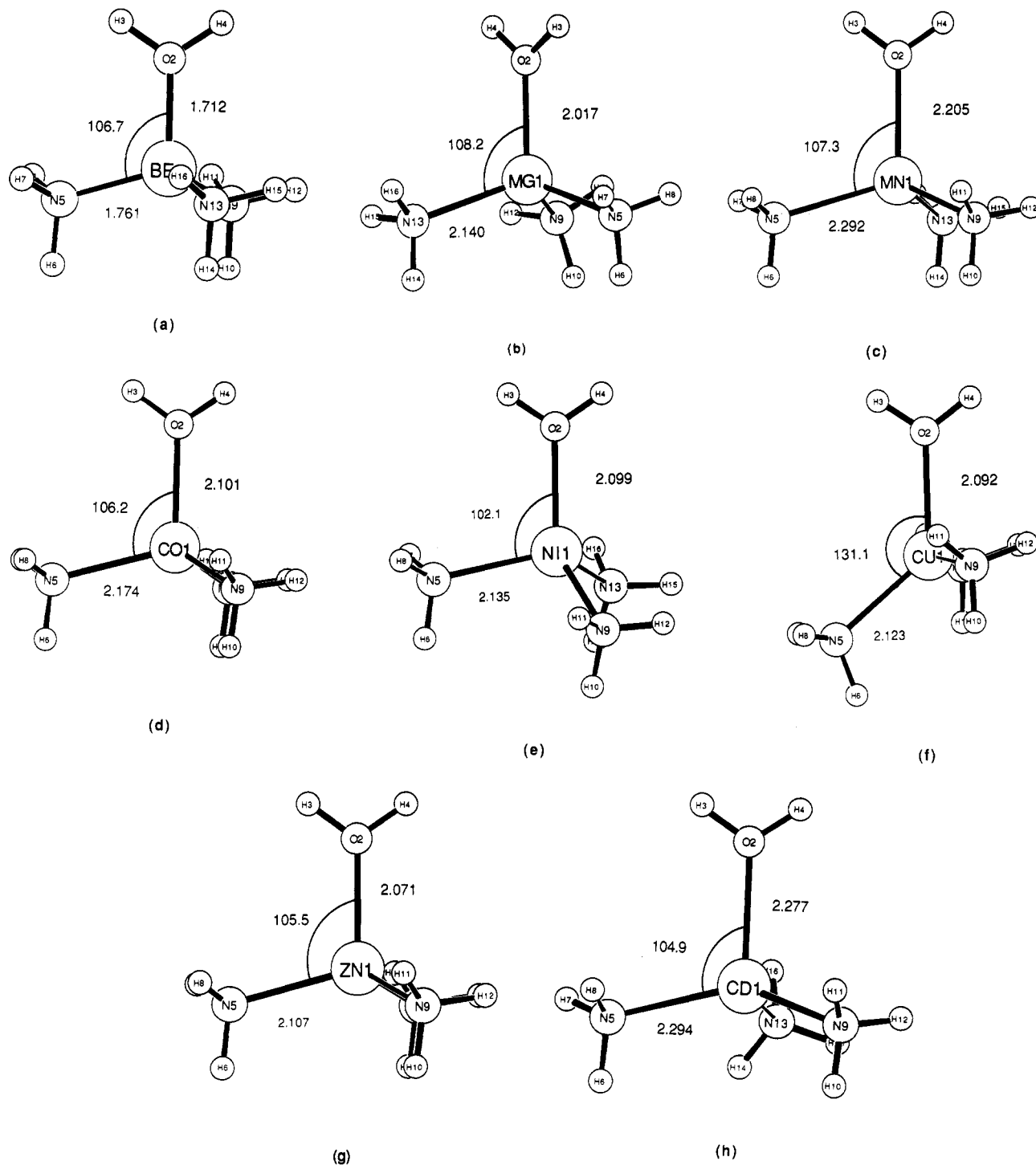
Figure 1. HF/lanl1dz-optimized geometries for the eight metal-substituted models of the CA enzyme studied: (a) Be; (b) Mg; (c) Mn; (d) Co; (e) Ni; (f) Cu; (g) Zn; and (h) Cd.

Table 1. Mulliken Charges on the Metal Atom for the Eight $(NH_3)_3M^{II}(H_2O)$ Complexes Analyzed

metal	q_M	metal	q_M
Be(II)	0.7972	Ni(II)	1.3775
Mg(II)	1.4876	Cu(II)	1.3457
Mn(II)	1.5642	Zn(II)	1.4238
Co(II)	1.4300	Cd(II)	1.3815

remarkable structural difference appears when one compares the angles of the Ni(II) and Cu(II) complexes with those of the other dications. Whereas a nearly symmetrical tetrahedral coordination is found in general, for the Cu(II) and Ni(II) complexes the tetrahedral geometry is clearly distorted. Jahn–Teller effects in tetrahedral complexes of Ni(II) and Cu(II) can account for this distortion,[66] whereas the electronic

structure of the other dications makes them unaffected by the Jahn–Teller distortion in tetrahedral structures.

Table 1 collects the Mulliken charges on the metal atom. This table shows that beryllium is far from being a dication, as opposed to the other seven metals, which are closer to the formal charge of +2. This large amount of charge transfer from ligands to beryllium causes its metal–ligand bond lengths to be shorter than in the other studied dication models. Moreover, Mn(II) (supporting the largest charge) and Cd-(II) complexes have the largest metal–ligand bond lengths.

In Table 2 we have collected the QMSM matrix for the eight studied complexes. As expected, along a periodic table's row (consider the Co(II)–Ni(II)–Cu(II)–Zn(II) dications), it is found that the self-QMSM increases with the atomic

Table 2. Quantum Molecular Coulomb-Like Similarity Measures Matrix for the Studied Metal-Substituted Models of CA Computed from Eq 1

	Be(II)	Mg(II)	Mn(II)	Co(II)	Ni(II)	Cu(II)	Zn(II)	Cd(II)
Be(II)	605.03							
Mg(II)	486.33	510.40						
Mn(II)	512.89	523.89	599.59					
Co(II)	561.16	571.54	612.08	662.37				
Ni(II)	580.92	564.51	605.86	655.62	704.17			
Cu(II)	583.14	529.77	583.90	624.51	639.93	740.25		
Zn(II)	608.63	599.75	630.05	691.91	696.02	669.24	743.46	
Cd(II)	564.70	563.59	645.93	662.67	660.14	642.82	689.10	704.42

Table 3. Similarity Indices Matrix for the Studied Metal-Substituted Models of CA Computed from Eq 2

	Be(II)	Mg(II)	Mn(II)	Co(II)	Ni(II)	Cu(II)	Zn(II)	Cd(II)
Be(II)	1.000							
Mg(II)	0.875	1.000						
Mn(II)	0.852	0.947	1.000					
Co(II)	0.886	0.983	0.971	1.000				
Ni(II)	0.890	0.942	0.932	0.960	1.000			
Cu(II)	0.871	0.862	0.876	0.892	0.886	1.000		
Zn(II)	0.907	0.974	0.944	0.986	0.962	0.902	1.000	
Cd(II)	0.865	0.940	0.994	0.970	0.937	0.890	0.952	1.000

Table 4. Distance Matrix for the Studied Metal-Substituted Models of CA Computed from Eq 3

	Be(II)	Mg(II)	Mn(II)	Co(II)	Ni(II)	Cu(II)	Zn(II)	Cd(II)
Be(II)	0.000							
Mg(II)	11.949	0.000						
Mn(II)	13.373	7.887	0.000					
Co(II)	12.046	5.450	6.149	0.000				
Ni(II)	12.139	9.250	9.595	7.437	0.000			
Cu(II)	13.379	13.824	13.117	12.394	12.828	0.000		
Zn(II)	11.456	7.374	9.108	4.691	7.455	12.051	0.000	
Cd(II)	13.419	9.362	3.486	6.438	9.397	12.611	8.348	0.000

number. Self-QMSM has been defined as the similarity between two identical systems perfectly superimposed, resulting in a similarity index of 1 and a Euclidean distance of 0. Table 3 shows the similarity indices (SI) matrix obtained through eq 2 and from the QMSM of Table 2. Finally, Table 4 gathers the distances calculated using eq 3 between the eight metal-substituted models of CA.

Although there are no experimental data on Be(II)-substituted models of CA, complexes with Be(II) instead of Zn(II) have been used in several theoretical works to study the mechanism of action of CA, given that the small size of this dication permits calculations to be performed at higher levels of theory.[67-69] Despite the distance from ZnCA being the closest to BeCA (11.46 in Table 4), the small SI and the large distance between these two models of CA strongly recommended not to use this dication in models of CA.[70] The short metal–ligand bond lengths of BeCA accounts for the large dissimilarities found between BeCA and the remainder of the MCA complexes analyzed.

To our knowledge, no Mg(II)-substituted CA has been yet synthesized as a model of CA. The intermediate distance and SI between MgCA and CoCA or ZnCA may suggest the use of this dication in experimental models of CA. An earlier theoretical work has already pointed out the large parallelism in the energy profile of a simple model of ZnCA and MgCA;[71] furthermore, it has been shown that Zn(II) exhibits intermediate behavior between Be(II) and Mg(II), yet closer to Mg(II). This result is well reproduced by our QMSM, the distance between BeCA and ZnCA (11.46) being larger than the distance found between MgCA and ZnCA (7.37). The intermediate behavior of the Zn(II) dication is also reflected

in Table 1, where one can see that the charge transfer from ligands to the metal follows the series Be > Zn > Mg.

The largest SI and smallest distance of all the related metal-substituted models of CA are found between Mn(II) and Cd-(II). Geometrical parameters of Figure 1 show an almost identical geometrical structure for these two complexes. This large similarity must be related to the electronic structure of high-spin Mn(II) and Cd(II). Since MN(II) is a d^5 dication and Cd(II) is a d^{10} dication, both complexes exhibit neither ligand field stabilization energy nor Jahn–Teller effects in tetrahedral complexes. On the other hand, it is well-known[66] that the ionic radius diminishes when going from Mn(II) to Cu(II) and then slightly increases from Cu(II) to Zn(II). It also increases along a column in the periodic table. Therefore, it is found that the ionic radii for these three dications follows the series Cd > Mn > Zn. The same series is found for the metal–oxygen bond lengths shown in Figure 1. The large charge transfer from the ligands to the 5s and 5p orbitals of the metal in Cd(II) as compared to Mn(II) (Table 1), combined with the value of ionic radii, explains the close bond lengths and large similarity found for the Mn(II) and Cd(II) complexes. It must be pointed out that small structural differences have also been found for the same dications with the same model of CA, but using another kind of pseudo-potentials.[34]

The most interesting result turns out to be the small distance and the large SI found between CoCA and ZnCA, in agreement with the experimental fact that only Co(II)-substituted CA confers meaningfull activity to the CA enzyme. This large similarity must be attributed to structural and electronic resemblances. As has been previously reported,[33,34] the geometrical parameters of Zn(II)- and Co(II)-substituted CA are quite similar (Figure 1). Further, the electronic structure of these two cations, d17 for high-spin Co(II) and d^{10} for Zn(II), gives rise to comparable electronic behavior in tetrahedral geometries. For instance, these cations do not suffer from Jahn–Teller effects in tetrahedral complexes. Moreover, going from tetrahedral to octahedral complexes, the ligand field stabilization energy is 0 for Zn(II) and very small for Co(II) complexes. These facts result in a large coordination flexibility of Co(II) and Zn(II); given that the catalysis of CO_2 hydration by CA implies changes in the coordination number of the first shell,[56] this coordination flexibility can be determinant of the activity of CA. Another key point turns out to be the charge transfer from the ligands to the metal. The larger the charge transfer, the more difficult the deprotonation process, which is the rate-limiting reaction in the catalytic cycle.[1,2,12-15] On the other hand, the larger the charge transfer, the higher the nucleophilicity of the generated hydroxide. Since these two effects counteract,[72,73] there must be an optimal charge transfer from the ligands to the metal. Interestingly, the charge transfers in Zn(II) and Co(II) are almost identical (Table 1), indicating again a large similarity between these two dications.

Ni(II) and primarily Cu(II) exhibit large dissimilarities when compared to all other dications. For instance the smallest distance from CuCA to any other metal-substituted CA is 12.05 and from NiCA is 7.44 (see Tables 3 and 4). This result can be rationalized taking into account the aforementioned Jahn–Teller distortion suffered by these two cations in tetrahedral complexes. The large Euclidean distances obtained when compared to the Zn(II) native enzyme fully support the experimental fact that nickel- and copper-substituted CA are inactive.[1,29] Moreover, and as predicted by the ligand field stabilization energy theory, tetrahedral complexes of Ni(II)
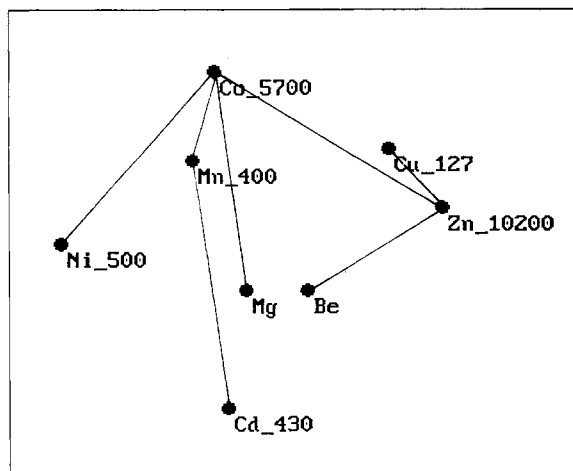
Figure 2. Kruskal tree relating the eight metal-substituted CA models of CA studied. Numbers in labels correspond to experimental activities of Table 5.

Table 5. Euclidean Distances Referred to Zinc-Enzyme and Experimental Hydration of CO_2 Activities of Metal-Substituted Enzymes

metal	d	Ac^a	metal	d	Ac^a
Zn(II)	0.00	10 200	Cd(II)	8.35	430
Co(II)	4.69	5700	Mn(II)	9.11	400
Ni(II)	7.46	500	Cu(II)	12.05	127

^{a} Activities from ref 29.

and Cu(II) are those most unstable when compared to the same octahedral complexes in the series studied.[66] Therefore, once the tetrahedral coordination in NiCA and CuCA is lost along the catalytic cycle of CA, it can be foreseen that reversion to a tetrahedral coordination must be difficult, thus preventing the reinitialization of the enzymatic catalysis. Moreover, it has been claimed that for an efficient catalysis the geometry and symmetry of the free enzyme must be close to those of the transition state (entatic[28] hypothesis), in such a way that the structure of the enzyme has to be ready yet to minimize the transition-state energy. Jahn–Teller distortions in NiCA and CuCA should alter the geometry of minima and transition states, and since in ZnCA these geometries are expected to be optimal, this would increase the energy barriers, hence resulting into the observed reduction of the biological activity of the enzyme.

The remarkable distance found between the Cd(II)-substituted model of CA and the Zn(II) model must be attributed to the different ionic radii of these dications, given that the electronic properties of these two dications are quite analogous. As above, in the light of the entatic hypothesis, it seems plausible that the larger bond lengths in CdCA as compared to ZnCA may alter the geometry of minima and transition states, hence increasing the energy barriers of the CO_2 hydration catalyzed by CA.

Another clue on the eight dications studied can still be provided. By means of a Kruskal tree[49] it is possible to connect all metal-substituted models of CA by a relation of minimum Euclidean distances. The Kruskal tree is a way to quickly visualize in a two-dimensional graph the main information gathered in the distance matrix of Table 4. Figure 2 depicts the plot obtained from the Euclidean distance matrix with the help of the Hypercloud program,[74] a set of graphic design facilities developed in our laboratory.

The above discussion has focused mainly on the reasons accounting for the differences and similarities between metal-substituted CA. However, a global comparison with experiment must still be made. For such a goal, we must refer to the paper by Coleman,[29] which carries out an experimental study on the kinetics of metal-substituted CA. Thus, in Table 5 we collect the Euclidean distances from a given metal-substituted CA to a zinc model of CA, together with the experimental activity of metal-substituted CA to hydration of CO_2. It is striking to see how well the activities are related to Euclidean quantum molecular similarity distances: the larger the distance, the smaller the experimental activity of

CA. Not only is the experimental ordering correctly reproduced but also a semiquantitative value for the activities is reflected. Moreover, Table 5 can allow the anticipation of the experimental activities of Mg(II)- and Be(II)-substituted CA. The quantum molecular similarity distances of such enzymes to the model enzyme are 7.37 and 11.46, respectively (Table 4). If one correlates the values of Table 5, one obtains $\ln Ac = 9.58 - 0.40d$, with a correlation coefficient of 0.96, which translates into a confidence of 0.1%.[75] Thus, the value of $d = 7.34$ for Mg translates into an activity of 759. Therefore, it is expected that substitution of Zn(II) by Mg(II) would not confer appreciable experimental activity to CA. Likewise, the distance from Be(II) results in an interpolated activity of 148, so Be(II) is not likely to confer activity to CA.

A last remark must be stated yet. It is noteworthy that our QMSM are almost basis set independent once the geometries of the systems studied are known. Nevertheless, the basis set used can be determinant in the process of computing and fully optimizing the geometries of the intervening species and therefore has a large influence on the final QMSM obtained. On the other hand, the model used does not include the protein and the solvent polarization of the enzyme, which have been predicted to have a large influence.[55,56,72] Obviously, both the basis set and the model used may be determinant in the final result obtained. However, the results of Table 5 highly encourage use of QMSM in order to predict biological and pharmacological activities of chemical compounds.

4. CONCLUSIONS

It has been shown that quantum molecular similarity measures can be very useful when one tries to classify a series of metal-substituted enzymes as a function of their biological activities. In particular, for the CA enzyme, it has been possible to predict theoretically, with a simple model of CA and the use of quantum molecular similarity measures, an activity of the Co(II)-substituted CA close to that of the naturally occurring zinc enzyme. The small Euclidean distance between these two systems has been associated with their small structural and electronic differences. MnCA and CdCA have been also predicted to have akin activities. Furthermore, the large distance found between BeCA and ZnCA invalidates the use of the beryllium dication as a good model for the zinc dication.

Despite that further work needs to be performed in the future, our calculations fully support the use of quantum molecular similarity measures as an efficient molecular enginnering tool to predict biological and pharmacological activities.

ACKNOWLEDGMENT

1052 *J. Chem. Inf. Comput. Sci., Vol. 34, No. 5, 1994*

SOLÀ ET AL.

Supercomputació de Catalunya (CESCA) for providing us with computing facilities.

(1) Woolley, P. Models for metal ions function in Carbonic Anhydrase. *Nature* **1975**, *258*, 677–682.

(2) Silverman, D. N.; Lindskog, S. The Catalytic Mechanism of Carbonic Anhydrase: Implications of a Rate-Limiting Protolysis of Water. *Acc. Chem. Res.* **1988**, *21*, 30–36.

(3) Pocker, Y.; Deits, T. L. The Bicarbonate Proton in Carbonic Anhydrase Catalysis. *J. Am. Chem. Soc.* **1983**, *105*, 980–986.

(4) Sen, A. C.; Tu, C. K.; Thomas, H.; Wynns, G. C.; Silverman, D. N. Comparison of Inhibition of Native and Co(II)-Substituted Bovine Carbonic Anhydrase II and III. In *Zinc Enzymes*; Bertini, I., Luchinat, C., Maret, W., Zeppezauer, M., Eds.; Birkhäuser: Boston, 1986; p 329.

(5) Eriksson, E. A.; Jones, T. A.; Liljas, A. Crystallographic studies of Human Carbonic Anhydrase II (HCA II). In *Zinc Enzymes*; Bertini, I., Luchinat, C., Maret, W., Zeppezauer, M., Eds.; Birkhäuser: Boston, 1986; p 317.

(6) Mulholland, A. J.; Grant, G. H.; Richards, W. G. Computer modelling of enzyme catalysed reaction mechanisms. *Protein Enginnering* **1993**, *46*, 133–147.

(7) Eriksson, A. E.; Liljas, A. Refined Structure of Bovine Carbonic Anhydrase III at 2.0 Å Resolution. *Proteins* **1993**, *16* 29–42.

(8) Kannan, K. K.; Ramanadham, M.; Jones, T. A. Structure, refinement and function of Carbonic Anhydrase isoenzymes: Refinement of Human Carbonic Anhydrase I. *Ann. N.Y. Acad. Sci.* **1984**, *429*, 49–60.

(9) Eriksson, A. E.; Liljas, A. Crystallization of and preliminary X-ray data for Bovine Carbonic Anhydrase III. *J. Biol. Chem.* **1986**, *261*, 16247–16248.

(10) Eriksson, A. E.; Jones, T. A.; Liljas, A. The refined structure of Human Carbonic Anhydrase II at 2.0 Å resolution. *Proteins* **1988**, *4*, 274–282.

(11) Kannan, K. K.; Ramanadham, M. Structure, Refinement, and Function of Human Carbonic Anhydrase-B *Int. J. Quantum Chem.* **1981**, *20*, 199–209.

(12) Davis, R. P. The Kinetics of the Reaction of Human Erythrocyte Carbonic Anhydrase. II. The Effect of Sulfanilamide, Sodium Sulfide and Various Chelating Agents. *J. Am. Chem. Soc.* **1959**, *81*, 5674–5678.

(13) Pocker, Y.; Janjić, N. Molecularity of Water in Enzymic Catalysis. Application to Carbonic Anhydrase II. *J. Am. Chem. Soc.* **1989**, *111*, 731–733.

(14) Rowlett, R. S.; Silverman, D. N. Kinetics of the Protonation of Buffer and Hydration of CO$_2$ Catalyzed by Human Carbonic Anhydrase II. *J. Am. Chem. Soc.* **1982**, *104*, 6737–6741.

(15) Tu, C.; Silverman, D. N.; Forsman, C.; Jonsson, B.-H.; Lindskog, S. Role of Histidine 64 in the Catalytic Mechanism of Human Carbonic Anhydrase II Studied with a Site-Specific Mutant. *Biochemistry* **1989**, *28*, 7913–7918.

(16) Bertini, I.; Canti, G.; Luchinat, C.; Scozzafava, A. Characterization of Cobalt(II) Bovine Carbonic Anhydrase and of its Derivatives. *J. Am. Chem. Soc.* **1978**, *100*, 4873–4877.

(17) Tu, C. K.; Silverman, D. N. Observation of the Visible Absorption Spectrum of Co(II)-Carbonic Anhydrase III during Catalytic Hydration of CO$_2$. *J. Am. Chem. Soc.* **1986**, *108*, 6065–6066.

(18) Williams, T. J.; Henkens, R. W. Dynamic ^{13}C NMR Investigation of Substrate Interaction and Catalysis by Co(II) Human Carbonic Anhydrase I. *Biochemistry* **1985**, *24*, 2459–2462.

(19) Haffner, P. H.; Coleman, J. E. Structure of the Active Site of Carbonic Anhydrase as Determined by Electron Spin Resonance. *J. Biol. Chem.* **1975**, *250*, 996–1005.

(20) Pesando, J. M. Proton Magnetic Resonance Studies of Carbonic Anhydrase. II. Group Controlling Catalytic Activity. *Biochemistry* **1975**, *14*, 681–688.

(21) Bertini, I.; Luchinat, C. Cobalt(II) as a Probe of the Structure and Function of Carbonic Anhydrase. *Acc. Chem. Res.* **1983**, *16*, 272–279.

(22) Tu, C. K.; Silverman, D. N. Catalysis by Cobalt(II)-Substituted Carbonic Anhydrase II of the Exchange of Oxygen-18 between CO$_2$ and H$_2$O. *Biochemistry* **1985**, *24*, 5881–5887.

(23) Wilkins, R. G.; Williams, K. R. Kinetics of Formation and Dissociation of Manganese-Bovine Carbonic Anhydrase B. *J. Am. Chem. Soc.* **1974**, *96*, 2241–2245.

(24) Bertini, I.; Luchinat, C.; Monnani, R.; Roelens, S.; Moratal, J. M. Interaction of CO$_2$ and Copper(II) Carbonic Anhydrase. *J. Am. Chem. Soc.* **1987**, *109*, 7855–7856.

(25) Lanir, A.; Navon, G. Nuclear Magnetic Resonance Studies of Carbonic Anhydrase. Binding of Sulfacetamide to the Manganese Enzyme. *Biochemistry* **1972**, *11*, 3536–3544.

(26) Led, J. J.; Neesgaard, E. Paramagnetic Carbon-13 NMR Relaxation Studies on the Kinetics and Mechanism of the HCO$_3^-$/CO$_2$ Exchange Catalyzed by Manganese(II) Human Carbonic Anhydrase I. *Biochemistry* **1987**, *26*, 183–192.

(27) Coleman, J. E. Mechanism of Action of Carbonic Anhydrase. *J. Biol. Chem.* **1967**, *242*, 5212–5219.

(28) Vallee, B. L.; Gades, A. The metallobiochemistry of zinc enzymes. *Adv. Enzymol.* **1984**, *56*, 283–430.

(29) Coleman, J. E. Metal Ion Dependent Binding of Sulphonamide to Carbonic Anhydrase. *Nature* **1967**, *214*, 193–194.

(30) Lipscomb, W. N. Structure and Catalysis of Enzymes. *Ann. Rev. Biochem.* **1983**, *52*, 17–34.

(31) Stöcker, W.; Wolz, R. L.; Zwilling, R.; Strydom, D. J.; Auld, D. S. Ascatus Protease, a Zinc Metalloenzyme. *Biochemistry* **1988**, *27*, 5026–5032.

(32) Bertini, I.; Donaire, A.; Messori, L.; Moratal, J. M. Interaction of Phosphate and Pyrophosphate with Co(II)-Carboxypeptidase A. *Inorg. Chem.* **1990**, *29*, 202–205.

(33) Vedani, A.; Huhta, D. W. A New Force for Modeling Metalloprotein. *J. Am. Chem. Soc.* **1990**, *112*, 4759–4767.

(34) Garmer, D. R.; Krauss, M. Metal Substitution and the Active Site of Carbonic Anhydrase. *J. Am. Chem. Soc.* **1992**, *114*, 6487–6493.

(35) Garmer, D. R.; Krauss, M. Cobalt Spectroscopy and the Structure of the Active Site of Carbonic Anhydrase. *In. J. Quantum Chem.* **1992**, *42*, 1469–1477.

(36) Mezey, P. G. *Shape in Chemistry: An introduction to molecular shape and topology*; VCH Publishers: New York, 1993.

(37) Mezy, P. G. Three-Dimensional Topological Aspects of Molecular Similarity. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990; p 321.

(38) Mezey, P. G. New Symmetry Theorems and Similarity Rules for Transition Structures. In *The Role of Computational Models and Theories in Biotechnology*; Bertrán, J., Ed.; Kluwer Academic Publishers: Dordrecht, 1992; p 93.

(39) Mezey, P. G. Similarity Analysis in two and three dimensions using lattice animals and polycubes. *J. Math. Chem.* **1992**, *11*, 27–45.

(40) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Ratón, FL, 1993.

(41) Cioslowski, J.; Fleischmann, E. D. Assessing Molecular Similarity from Results of ab Initio Electronic Structure Calculations. *J. Am. Chem. Soc.* **1991**, *113*, 64–67.

(42) Cioslowski, J.; Challacombe, M. Maximum Similarity Orbitals for Analysis of the Electronic Excited States. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1991**, *25*, 81–93.

(43) Cioslowski, J. Quantifying the Hammond Postulate: Intramolecular Proton Transfer in Substituted Hydrogen Catecholate Anions. *J. Am. Chem. Soc.* **1991**, *113*, 6756–6760.

(44) Richards, W. G.; Hodgkin, E. E. Molecular Similarity. *Chem. Br.* **1988**, 1141–1144.

(45) Good, A. C.; Hodgkin, E. E.; Richards, W. G. Utilization of Gaussian Functions for the Rapid Evaluation of Molecular Similarity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 188–191.

(46) Good, A. C.; Richards, W. G. Rapid Evaluation of Molecular Similarity using Gaussian Functions. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 112–116.

(47) Carbó, R.; Arnau, M.; Leyda, L. How Similar is a Molecule to Another? An Electron Density Measure of Similarity between Two Molecular Structures. *Int. J. Quantum Chem.* **1980**, *17*, 1185–1189.

(48) Carbó, R.; Calabuig, B. Molecular Quantum Similarity Measures and N-Dimensional Representation of Quantum Objects. I. Theoretical Foundations. *Int. J. Quantum Chem.* **1992**, *42*, 1681–1693.

(49) Carbó, R.; domingo, Ll. LCAO-MO Similarity Measures and Taxonomy. *Int. J. Quantum Chem.* **1987**, *32*, 517–545.

(50) Mestres, J.; Solá, M.; Duran, M.; Carbó, R. On the Calculation of Ab Initio Quantum Molecular Similarities for Large Systems: Fitting the Electron Density. Submitted for publication.

(51) Mestres, J.; Solá, M.; Duran, M.; Carbó, R. Fast calculation of Quantum Molecular Similarities from ab initio wave functions. In *Proceedings of the First Girona Seminar on Molecular Similarity*; Mezey, P. G.; Carbó, R., Eds.; submitted for publication.

(52) Jacob, O.; Cardenas, R.; Tapia, O. An ab Initio Study of Transition Structures and Associated Products in [ZnOHCO$_2$]$^+$, [ZnHCO$_3$H$_2$O]$^+$, [Zn(NH$_3$)$_3$HCO$_3$]$^+$ Hypersurfaces. On the Role of Zinc in the Catalytic Mechanism of Carbonic Anhydrase. *J. Am. Chem. 4Soc.* **1990**, *112*, 8692–9705.

(53) Bertini, I.; Luchinat, C.; Rosi, H.; Sgamellotti, A.; Tarantelli, F. pK_a of Zinc-Bound Water and Nucleophilicity of Hydroxo-Containing Species. Ab Initio Calculations on Models for Zinc Enzymes. *Inorg. Chem.* **1990**, *29*, 1460–1463.

(54) Zheng, Y.-J.; Merz, K. M., Jr. Mechanism of the Human Carbonic Anhydrase II Catalyzed Hydration of Carbon Dioxide. *J. Am. Chem. Soc.* **1992**, *114*, 10498–10507.

(55) Solá, M.; Lledós, A.; Duran, M.; Bertrán, J. Ab Initio Study of the Hydration of CO$_2$ by Carbonic Anhydrase. A Comparison between the Lipscomb and Lindskog Mechanisms. *J. Am. Chem. Soc.* **1992**, *114*, 869–877.

(56) Solá, M.; Lledós, A.; Duran, M.; Bertrán, J. Theoretical Study of the Catalyzed Hydration of CO$_2$ by Carbonic Anhydrase: A Brief Overview. In *Molecular Aspects of Biotechnology: Computation Models and Theories*; Bertrán, J., Ed.; Kluwer Academic Publishers: The Netherlands, 1992; p 263.

(57) Krauss, M.; Garmer, D. R. Active Site Ionicity and the Mechanism of Carbonic Anhydrase. *J. Am. Chem. Soc.* **1991**, *113*, 6426–6435.

(58) Schlegel, H. B. Optimization of Equilibrium Geometries and Transition Structures. *J. Comput. Chem.* **1982**, *3*, 214–218.

(59) Dunning, T. H., Jr. Gaussian Basis Functions for use in Molecular Calculations. I. Contraction of (9s5p) Atomic Basis Sets for the First-Row Atoms. *J. Chem. Phys.* **1970**, *53*, 2823–2833.

METAL-SUBSTITUTED CARBONIC ANHYDRASE

J. Chem. Inf. Comput. Sci., Vol. 34, No. 5, 1994 **1053**

(60) Dunning, T. H., Jr. Gaussian Basis Functions for use in Molecular Calculations. IV. Representation of Polarization Functions for the First Row Atoms and Hydrogen. *J. Chem. Phys.* **1971**, *55*, 3958–3966.

(61) Wadt, W. R.; Hay, P. J. Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi. *J. Chem. Phys.* **1985**, 482, 284–298.

(62) Hay, P. J.; Wadt, W. R. Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.* **1985**, *82*, 284–298.

(63) Frisch, M. J.; Head-Gordon, M.; Trucks, G. W.; Foresman, J. B.; Schlegel, H. B.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Martin, R. L.; Kahn, L. R.; Stewart, J. J. P.; Topiol, S.; Pople, J. A. *GAUSSIAN 90*, Revision H; GAUSSIAN Inc.: Pittsburgh, PA, 1990.

(64) *MESSEM*, version 1.0; Mestres, J.; Solá, M.; Duran, M.; Carbó, R.; Girona CAT, 1993.

(65) Scales, L. E. *Introduction to Non-Linear Optimization*; Springer-Verlag: New York, 1985.

(66) Cotton, F. A.; Wilkinson, G. *Advanced Inorganic Chemistry*; John Willey & Sons, Inc.: New York, 1966.

(67) Liang, J.-Y.; Lipscomb, W. N. Hydration of Carbon Dioxide by Carbonic Anhydrase: Internal Proton Transfer of Zn^{+2}-bound $HCO_3 1-$. *Biochemistry* **1987**, *26*, 5293–5301.

(68) Liang, J.-Y.; Lipscomb, W. N. Binding of Sulfonamide and Acetamide to the Active Site Zn^{+2} in Carbonic Anhydrase: A theoretical study. *Biochemistry* **1989**, *28*, 9724–9733.

(69) Liang, J.-Y.; Lipscomb, W. N. Theoretical Study of Carbonic Anhydrase-Catalyzed Hydration of CO_2: A Brief Review. *Int. J. Quantum Chem.* **1989**, *36*, 299–312.

(70) Osman, R.; Weinstein, H.; Topiol, S. Models for Active Sites of Metalloenzymes. II. Interactions with a model substrate. *Ann. N.Y. Acad. Sci.* **1981**, *367*, 356–369.

(71) Solá, M.; Lledós, A.; Duran, M.; Bertrán, J. Proton transfer in the water dimer catalyzed by doubly charged cations (Zn^{+2}, Be^{+2}, and Mg^{+2}). An ab initio study. *Theor. Chim. Acta* **1992**, *81*, 303–318.

(72) Solá, M.; Lledós, A.; Duran, M.; Bertrán, J. Anion binding and Pentacoordination in Zinc(II) Complexes. *Inorg. chem.* **1991**, *30*, 2523–2527.

(73) Kimura, E.; Koike, T.; Toriumi, K. A. Trigonal-Bipyramidal Zinc(II) Complex of a Phenol-Pendant Macrocyclic Triamine. *Inorg. Chem.* **1988**, *27*, 3678–3688.

(74) Carbó, R.; Calabuig, B. Molsimil-88: Molecular Similarity Calculations using a CNDO-like Approxiation. *Comput. Phys. Commun.* **1989**, *55*, 117–126.

(75) Sachs, L. *Applied Statistics. A Handbook of Techniques*; Springer-Verlag: New York, 1982.