J Mol Model (2014) 20:2271
DOI 10.1007/s00894-014-2271-z

ORIGINAL PAPER

Analyses of cobalt–ligand and potassium–ligand bond lengths in metalloproteins: trends and patterns

Natércia F. Brás · António J. M. Ribeiro · Marina Oliveira · Nathália M. Paixão · Juan A. Tamames · Pedro A. Fernandes · Maria J. Ramos

Received: 10 February 2014 / Accepted: 23 April 2014 / Published online: 22 May 2014

Abstract Cobalt and potassium are biologically important metal elements that are present in a large array of proteins. Cobalt is mostly found in vivo associated with a corrin ring, which represents the core of the vitamin B_{12} molecule. Potassium is the most abundant metal in the cytosol, and it plays a crucial role in maintaining membrane potential as well as correct protein function. Here, we report a thorough analysis of the geometric properties of cobalt and potassium coordination spheres that was performed with high resolution on a representative set of structures from the Protein Data Bank and complemented by quantum mechanical calculations realized at the DFT level of theory (B3LYP/ SDD) on mononuclear model systems. The results allowed us to draw interesting conclusions on the structural characteristics of both Co and K centers, and to evaluate the importance of effects such as their association energies and intrinsic thermodynamic stabilities. Overall, the results obtained provide useful data for enhancing the atomic models normally applied in theoretical and computational studies of Co or K proteins performed at the quantum mechanical level, and for developing molecular mechanical parameters for treating Co or K coordination spheres in molecular mechanics or molecular dynamics studies.

Keywords Cobalt enzymes · Potassium enzymes · Metalloproteins · PDB · DFT · B3LYP

Natércia F. Brás and António J. M. Ribeiro contributed equally to this work.

This paper belongs to Topical Collection QUITEL 2013.

Electronic supplementary material The online version of this article (doi:10.1007/s00894-014-2271-z) contains supplementary material, which is available to authorized users.

N. F. Brás · A. J. M. Ribeiro · M. Oliveira · J. A. Tamames · P. A. Fernandes · M. J. Ramos (✉)
REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal
e-mail: mjramos@fc.up.pt

N. M. Paixão
NEQC, Núcleo de Estudos em Química Computacional, Departamento de Química, Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora, Juiz de Fora, MG 36036-330, Brazil

Introduction

More than one-third of all proteins contain one of the following ten metal ions in catalytic or structural roles: Na, Mg, K, Ca, Mn, Fe, Co, Ni, Cu, Zn [1]. Characterizing the metal ion and its coordination sphere in these proteins is of the upmost importance for understanding the overall function of the protein. There are a wide variety of metal coordination spheres in proteins, and they are not usually specific to a certain metal. The same types of proteins can bind different metals, depending on the organism or cellular compartment considered [2].

In the past, several studies have focused on investigating these metal centers in proteins or inorganic molecules [1, 3–11] using the Protein Data Bank [12] or the Cambridge Structural Database [13]. Most of these studies utilize statistical characterization of the metal coordination spheres; for example: identify the most common atoms and residue ligands, their average distances to the metal, and the coordination number in relation to the oxidation state of the metal. There are also a couple of internet databases containing organized and easily accessible information about metal coordination spheres [14–16]. The extended environment of the metallic center may also be of interest [3, 17].

For the study reported in the present paper, we chose two metals with different coordination types that can be contrasted. On the one hand, we chose cobalt, a typical transition metal with an average position in the Irving and

Williams series for complex stability [18], which forms bonds with significant covalent character. On the other hand, we chose potassium as a typical example of an alkali metal. Potassium ions have a single positive charge, and present interactions that are mostly electrostatic in nature.

Cobalt is mostly found in vivo associated with a corrin ring, which represents the core of the vitamin B_{12} molecule (also known as cobalamin) [19]. There are different chemical forms of cobalamin that vary in the additional ligand bonded to the cobalt ion. Methylcobalamin (MeCbl), adenosylcobalamin (AdoCb), and cyanocobalamin have a methyl, an adenosyl, and a cyano group bonded to the cobalt, respectively. All cobalamins have a benzimidazole group bonded to the cobalt. Upon protein binding, the benzimidazole is substituted by an exogenous ligand, usually a histidine [20].

MeCbl is the cofactor of several methyltransferases, and AdoCbl is the cofactor of several eliminases—isomerases—of class II ribonucleotide reductase [20]. Additionally, the potential use of B_{12} in the coupled delivery of therapeutic drugs and imaging agents to proliferating cells as well as in the oral delivery of peptides and protein molecules has been widely explored [21].

Non-corrin cobalt sites are less numerous but are also important. They are found in isoforms of sugar isomerases and nitrile hydratase, for example [10, 22]. Non-native non-corrin cobalt-binding sites are relevant too. At high concentrations, cobalt competes with their native divalent metals to bind to certain protein sites, which can cause enzymatic inhibition and irreversible structural changes, ultimately leading to cobalt poisoning [23, 24]. Cobalt and its coordination sphere can be studied by EPR (electron paramagnetic resonance), NMR (nuclear magnetic resonance), and circular dichroism techniques. Substitution by cobalt can be used to study binding sites that are otherwise silent to these methods [25–27]. On the other hand, due to its toxicity, non-corrin cobalt is usually replaced with less toxic divalent metals that have similar activities in industry [28].

Potassium is the most abundant metal in the cytosol, and its crucial role in maintaining membrane potential (together with Na^+) is widely known. Its importance for correct protein function is also considerable [29, 30].

K^+ and other monovalent ions are not catalytic per se due to their low charge density, especially in comparison with divalent ions, which have more charge and are generally smaller [31]. K^+ contributes to enzyme activation through two other mechanisms [29, 30]. In K^+-activated type I enzymes, K^+ functions as a cofactor to aid the binding of a substrate to the active center, as is the case for diol dehydratases and certain phosphoryl transfer enzymes [32, 33]. In K^+-activated type I enzymes, K^+ does not bind the active center directly, but the binding of the metal to the enzyme leads to its allosteric activation, as in ribokinase [34]. Furthermore, the binding of K^+ ions on the surfaces of proteins and nucleic acids may have a simple structural role that does not influence catalysis, or it may be an artifact of crystallization [35].

Apart from the statistical studies mentioned above, computational studies of metallic centers in macromolecules have also proven to be very useful [8, 9, 11, 36, 37]. Aside from structural studies, computational studies have consistently been employed to establish the catalytic pathways of metalloenzymatic reactions [38–44].

In this work, we start by presenting a statistical analysis of a sample of the coordination spheres of Co and K in proteins. We focus on describing the most common ligands, atoms, and coordination numbers of each metal, as well as other statistical data. Subsequently, we report the results from quantum mechanical (QM) calculations of these coordination spheres that were performed to elucidate the association energies between the metals and their ligands, as well as to find the equilibrium metal–ligand distances.

Computational methods

PDB data sets and statistical analysis

The Protein Data Bank (PDB) was used to search all proteins containing cobalt (Co) and potassium (K) atoms. All X-ray crystallographic structures of proteins with at least one Co or K ion as cofactor, with a similarity set of less than 90 %, without residue mutations, and with a resolution of <2.0 Å were considered for this study, amounting to a total of 45 and 79 Co-protein and K-protein structures, respectively. A similarity of 90 % was used to obtain a final set of structures that were considered to be representative of these specific metalloproteins. The high resolution was employed to obtain meaningful results and increase the statistical significance of these analyses. Cutoffs for the metal–atom distance of 3.0 Å and 3.5 Å were used to identify Co-binding and K-binding ligands, respectively. The raw data obtained were refined and then analyzed further. For the statistical analysis, we considered the nature of the metal-coordinating residues and metal-coordinating atoms, the number of metal atoms by protein, the coordination number for each metal coordination sphere, the corresponding set of coordination numbers by metalloprotein, and the set of ligands in each coordination sphere by coordination number and enzyme class.

Quantum calculations

To investigate the thermodynamic stabilities of the representative coordination spheres of Co and K, the association energies of the most common ligand combinations for Co

coordination numbers of 5 and 6 and K coordination numbers of 3, 4, 5, 6, and 7 were investigated using quantum mechanics. Models of the systems were prepared from the PDB structures with the best resolution for each combination of ligands in the metal coordination sphere considered. We used the following Co-protein structures: 1CLK [45], 1EEX [46], 1QQ0 [47], 1RQB [48], 1Y80, 2F7V [49], 2H9A [50], 2PTM [51], 2VKE [52], 3BFQ [53]; and the following K-protein structures: 1AKD [54], 1BW9 [55], 1CS0 [56], 1M5H [57], 1PKX [58], 1PVN [59], 1SX3 [60], 1W19 [61], 1WXX [62], 1XKY [63], 1YNF [64], 1ZZ0 [65], 2A6V [66], 2B3H [67], 2DWU [68], 2G50 [69], 2GJ8 [70], 2H88 [71], 2IX4 [72], 2QV6 [73], 2VQM [74]. The models selected for QM calculations are presented in Figs. SI-1, SI-2, SI-3, SI-4, and SI-5 of the "Electronic supplementary material" (ESM). Only the first coordination sphere of the metal atom was included, and the side chains of the residues were modeled in the conventional way. For example, the aspartate (or glutamate), histidine, serine (or threonine), and tyrosine were modeled as acetate, imidazole, methanol, and phenol groups, respectively. A very common ligand, the carbonyl group of a peptide bond, was modeled as acetamide, in a similar manner to asparagine (and glutamine) residues. Truncated residues are generally employed in cluster model studies [75–77]. This approach accurately mimics the complete residues present in crystallographic files while reducing the computational cost of the optimization calculations. However, to validate this protocol, four K-protein systems (pdb codes 1ynf, 1pkx, 2g50, and 1w85, with a hexacoordinate K center) with complete first coordination metal spheres were created, and the association energies for these models were calculated. According to the data obtained and shown in Table SI-1 of the ESM, the results obtained and interpretation are similar to those for previous truncated models, which validates our discussion and conclusions. Each of these initial structures was geometrically optimized using DFT, with the exchange correlation functional B3LYP [78, 79] and the basis set 6-31G (d) [80] employed for all atoms except for cobalt and potassium, for which the SDD pseudopotential was employed [81]. No geometric constraints were imposed in the Co-protein optimization calculations. However, most of the K-protein systems had the α-carbon atom positions frozen during the geometrical optimization calculations.

All calculations were performed using the Gaussian 09 software package [82]. DFT calculations have been shown to give very accurate results for systems involving metals. Following geometry optimization of all the metal coordination sphere models, the polarized continuum method (PCM) with the polarizable conductor calculation model (C-PCM) [83, 84] was used to evaluate the effect of a generic enzymatic environment on the energy values. Using this method with an empirical dielectric constant (ε) of 4, the long-range and nonspecific effects of a standard protein in the metal coordination sphere were included. This value was used because previous studies on the active sites of proteins [85–87] have shown that an empirical dielectric constant of 4 generally leads to good agreement with experimental data, and accounts for the average effect of both the protein and buried water molecules. However, the influence of intermediate hydrophobic protein environments on the association energies of four K-protein systems (pdb codes 1ynf, 1pkx, 2g50 and 1w85, with a hexacoordinate K center) was also evaluated. Hence, six different empirical dielectric constant values (2, 4, 10, 20, 40, and 80) were considered. According to the results shown in Table SI-1 of the ESM, the association energies decrease with increasing hydrophilic character of the environment. Furthermore, complexes with a global charge of 0 are generally more stable than those with a global positive charge.

C-PCM energy calculations with B3LYP/6–311++G (2d,2p)|SDD and ε = 4 were performed for all minimized structures. We were not able to calculate the energy of the model built from the PDB structure 2VKE with this basis set due to the excessive computational time required. We used the 6–311+G (d,p) basis set instead.

The difference between the energy of each metal coordination sphere and the sum of the energies of its individual ligands corresponds to the association energy (assuming that the thermal contributions to both terms are similar and thus cancel each other). These energetic values were used to evaluate the thermodynamic stabilities of the several possible metal coordination spheres in a general enzymatic environment.

In general, the present strategy represents the usual approach employed in quantum chemical computational enzymology, and it has previously been applied in the study of other metalloenzymes [9].

Results and discussion

Cobalt study

Statistical analysis

Number of Co atoms by protein Figure 1 shows the number of Co atoms in each protein, as well as the coordination number of each center in each protein. 37 % of the proteins studied have only one Co ion (e.g., 1CLK, 1QQ0, and 2VKE) and 22 % have two cobalt ions (e.g., 1CB7, 1EEX, and 1N2Z). In proteins that have two or more Co ions, two general scenarios are possible: the protein has just one metallic center containing several metal ions, or it has several centers each with only one metal ion. Based on our analysis of these cases, we concluded that centers with only one Co ion are most common. Because of this, we selected a dozen of these systems to perform a series of quantum mechanical studies (shown in the section

below). Apart from single-metal centers, there are also several examples of binuclear centers, and one example of a trinuclear center (1ZJC) [88]. Thus, proteins with more than one Co ion may have any combination of single-metal and binuclear centers.

The most common coordination numbers are 6 and 5: these occur in 52 % and 29 % of the cases, respectively. Centers with smaller coordination numbers do exist but they are usually centers with higher coordination numbers that include some transient water molecules as ligands, and those ligands do not appear on the X-rays.

The nature of Co coordination spheres

The compositions of Co coordination spheres in terms of atoms and ligands are summarized in Fig. 1. The atoms that occur far more frequently than any others are oxygen and nitrogen: they account for 55 % and 40 % of all the bonds to cobalt, respectively. This observation is in accord with the following analysis of the ligands. The most notable members of the "other atoms" category are the carbon atoms of cyanocobalamin and methylcobalamin, which are two of the few examples of metal–carbon bonds found in organic compounds.

The abundances of various ligands in Co coordination spheres are shown also in Fig. 1. The most common ligand is water, which represents 34 % of all the bonds to Co. Negatively charged amino acids account for 23 % of all the ligands, and histidines are responsible for another 21 % of the ligands. The "other ligands" category, which comprises 13 % of all the ligands, is very important in the case of cobalt because it includes corrin rings—characteristic of the cobalamin family of molecules—as well as the ammonia molecules in the hexaamminecobalt(III) ion. Binding to other amino acids accounts for the remaining 7 %. If we consider specific amino acid configurations, some occur more often among the cases studied than others, but this is more due to repetition of the same coordination sphere in the same protein structure than repetition of the same configuration in different proteins. For this reason, the charts in Fig. 2 are only indicative of the repetition of particular spheres in the same protein, not repetition across several proteins. Among the spheres with a coordination number of 5, the most common configuration is D-H-H-H-Wat, which occurs six times. There are three configurations that each occur four times: C-H-H-BB2-BB2, D-E-E-H-WAT, and E-E-H-H-Q. Finally, there are 24 configurations that occur only once.

Spheres with a coordination number of 6 are also observed to be very heterogeneous in nature. There is a sphere with 12 occurrences: D-D-D-WAT-WAT-WAT; two with four occurrences: C-C-E-H-M-WAT and D-D-H-H-WAT-WAT; four with three occurrences: B12-B12-B12-B12-B12-H, CSD-CSO-CSO-CSO-C-S-SO4, DSN-H-H-M-PLP-WAT, and H-H-WAT-WAT-WAT-WAT; four with two occurrences: CL-CNC-CNC-CNC-CNC-CNC, D-E-E-H-M1C-M1C, E-E-H-H-H-WAT, and H-H-H-WAT-WAT-WAT; and another 16 configurations with only one occurrence each.

Enzyme classification

Among the 45 proteins we focused on in this study, 29 are enzymes (62 %). Figure 3 shows the number of Co enzymes in each enzymatic class. The hydrolase class is the most well represented, with 12 occurrences (41 %), followed by the lyase and transferase classes, both with five occurrences. Additionally, there are four isomerases, two oxidoreductases, and one ligase. Co does not play the same role in all of these enzymes, but some trends are evident. In hydrolases, for example, the role of the cobalt is to stabilize the negatively charged hydroxide molecule that will act as a nucleophile [22]. In corrin lyase, transferase, and oxidoreductase enzymes, Co acts in conjunction with the corrin ring to facilitate the formation of radical species [20]. In these enzymes, another crucial aspect of corrin Co activity is its ability to bind methyl groups—a unique characteristic among metals in biological systems.

Quantum mechanical calculations

Following the statistical analysis, we performed a series of quantum mechanical calculations with a handful of models to evaluate the relative stabilities of K and Co coordination spheres.

Since the abundance of a particular sphere was only related to the presence of several configurations in the same protein, not in several proteins, we did not use that criterion to select the spheres to use in the calculations. Instead, we selected six examples of spheres with Co(II) and another six examples with Co(III). We selected Co(II) spheres with different charges (from −1 up to 2), and coordination numbers (5 and 6, the most common). For spheres that contain Co(III), we chose five cobalamin systems and the hexaamminecobalt ion. We modeled each of these systems as described in the "Methods" section. The models used in the QM calculations are shown in Figs. SI-1 and SI-2 of the ESM.

The results for the QM calculations are shown in Fig. 4. We should first comment on the D-D-D-H configuration, which has five metal bonds in the X-ray configuration (D-D-D-D-H) but has only four metal bonds in the optimized structure. During the optimization, the bond of an oxygen atom of one aspartate to the metal is broken. This aspartate remains bound to the metal, but only through one of its oxygen atoms. This

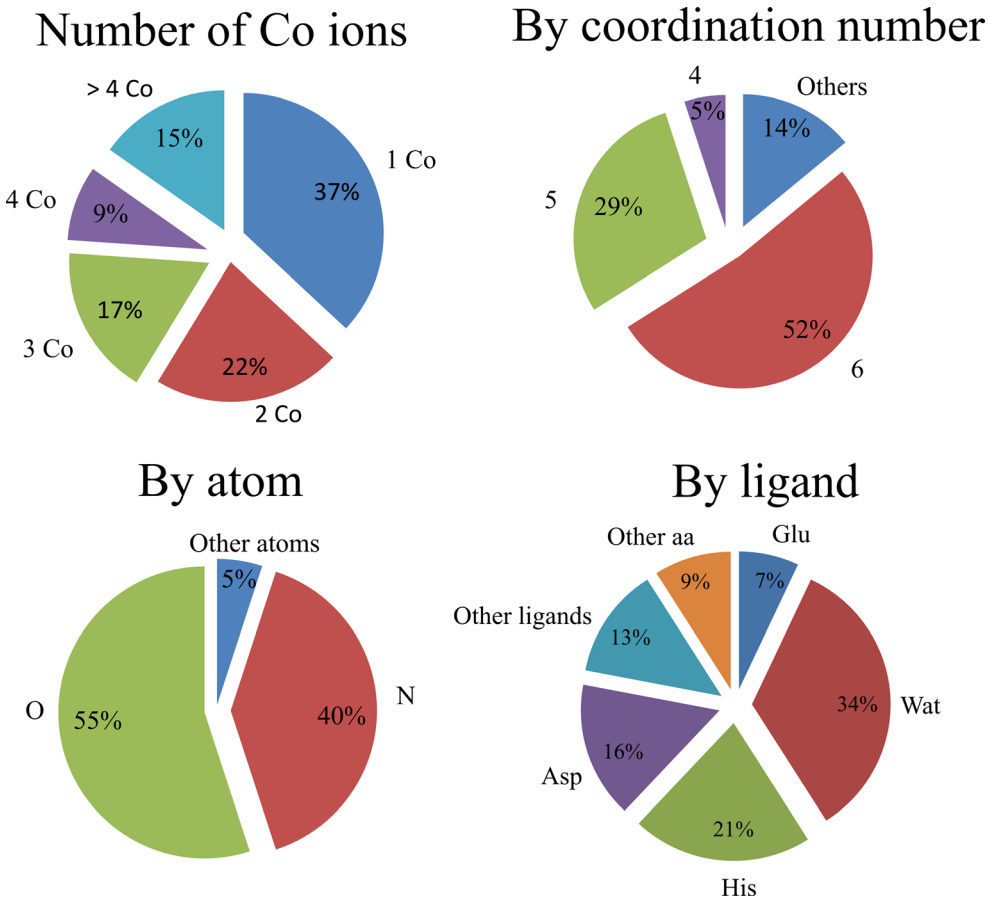

Fig. 1 Percentage Co ions in each protein and their respective coordination numbers, as well as the composition of each Co coordination sphere by atom and ligand

could mean that the correct configuration is tetrahedral, but that it was incorrectly assigned in the crystal. No trend in the number of ligands is seen for the coordination spheres of Co(II). Instead, there is a strong correlation between the overall charge of the system and the association energy. The D-D-D-H model, with a charge of −1, has an association energy of −365 kcal mol^{-1}. The two neutral models, KCX-KCX-H-H-D-WAT and D-D-D-H-WAT, have weaker association energies: −330 and −300 kcal mol^{-1}, respectively. Models with positive charges have even lower association energies. The model with a charge of +1, H-TAC-TAC-WAT-WAT-WAT, has an energy of −265 kcal mol^{-1}. The two models with a charge of +2, H-WAT-WAT-WAT-WAT and H-H-WAT-WAT-WAT, have energies of −192 and −188 kcal mol^{-1}, respectively. This pattern of association energies can be understood on the grounds of the electrostatic stabilization provided by the ligands to the metal, and vice versa. Models with an overall charge of +2 have ligands that do not have any charge, while models with an overall charge of zero have ligands with of a charge of −2. It is expected that the latter are much more stable when brought in contact with the positively charged metal ion. This effect is even more pronounced for the D-D-D-H system, which has an overall charge of −1. However, this trend should not continue for more negatively charged systems, as the repulsion between ligands becomes significant.

The results for models with Co(III) are also shown in Fig. 4. These systems have much stronger association energies than Co(II) models—particularly the corrin systems. Since corrin systems are all very similar, and they all have the same charge, it is possible to study the effect of changes to the fifth and sixth ligands on the association energies. The two models with stronger association energies are both corrin systems with an aliphatic carbon directly bonded to the cobalt ion. The 1EEX model has a longer aliphatic chain, while the 1CB7 model (methylcobalamin) has only a methyl group. The longer aliphatic chain seems to provide more stabilization (−899 kcal mol^{-1}) than the methyl group does (−858 kcal mol^{-1}). The fact that the bottom ligand differs for these systems (histidine vs. a similar benzimidazole) should not be significant, as will be explained below. After the aliphatic carbon, the preferred ligand for the corrin-bound cobalt is a hydroxide ion: the 1Y80 model (hydroxocobalamin in the base-off form) has an association energy of −836 kcal mol^{-1}. In order to test the effect of changing the fifth ligand on the association energy, we created a model of hydroxocobalamin in the base-off form, with a benzimidazole group as fifth ligand (OHCbl in Fig. SI-1). This OHCbl model

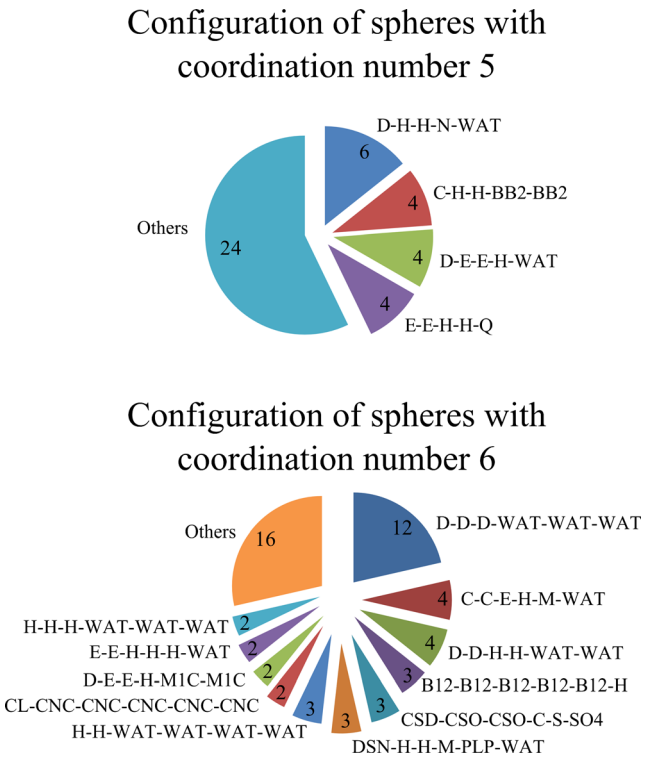

Fig. 2 Ligand configurations for Co coordination spheres with coordination numbers of 5 and 6. Three-letter codes corresponding to their PDB names are given for non-amino-acid ligands: *WAT* water, *BB2* actinonin, *B12* cobalamin, *CSD* 3-sulfinoalanine, *CSO* *S*-hydroxycysteine, *SO4* sulfate, *DSN* D-serine, *PLP* pyridoxal-5′-phosphate, *CNC* cyanocobalamin, *M1C* (3*S*)-3-amino-1-(cyclopropylamino)-2,2-heptanediol. *Single letters* refer to the one-letter codes of amino acids

has an association energy of −832 kcal mol^{-1}, showing that the change in the fifth ligand is negligible. The last corrin model with six ligands is the 1N2Z model (cyanocobalamin), with an energy of −816 kcal mol^{-1}. The five-ligand corrin model (2H9A), with a hydroxide at the sixth position, has an association energy of −783 kcal mol^{-1}. By comparison with the 1Y80 model, it can be concluded that the contribution of the fifth ligand alone amounts to ca. 50 kcal mol^{-1}. Finally, the

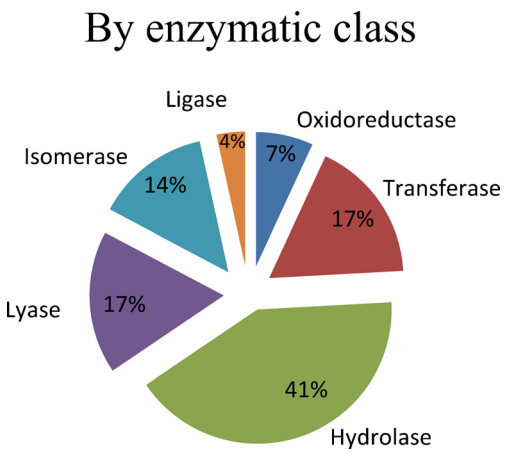

Fig. 3 Proportion the enzymes in each enzyme class that are Co enzymes

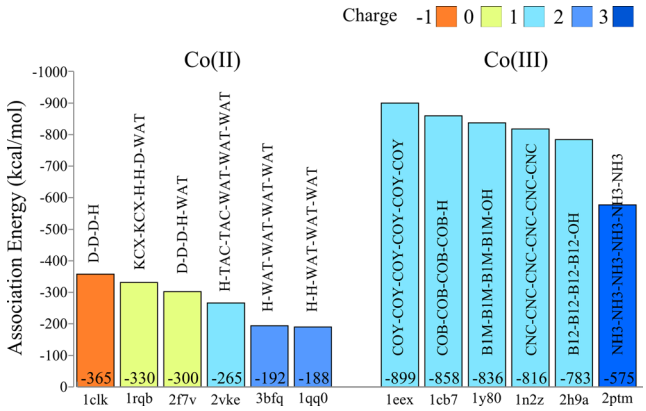

Fig. 4 Association energies of several Co coordination spheres, as calculated with QM methods. Three-letter codes corresponding to their PDB names are given for non-amino acid ligands: *KCX* lysine, *NZ* carboxylic acid, *WAT* water, *TAC* tetracycline, *COY* co-(adenin-9-yl-pentyl)-cobalamin, *COB* co-methylcobalamin, *B1M* co-5-methoxybenzimidazolylcobamide, *CNC* co-cyanocobalamin, *B12* cobalamin, *NH3* ammonia. *Single letters* refer to the one-letter codes of amino acids

six amine molecules of the 2PTM model have much weaker interactions with Co(III) than the corrin ring, 575 kcal mol^{-1}, probably due to the high overall charge of this model, +3. Nevertheless, this interaction is still much stronger than that observed for Co(II) systems.

Overview of metal–ligand distances

Table 1 presents the average distances between the Co ion and the ligand atoms that coordinate to it for all of the systems considered in this work. The average oxygen–Co distance is 2.10 Å with a standard deviation (SD) of 0.12. The discriminated analysis of these distances shows that there is an inverse relation between the average oxygen–Co distance and the formal charge on the oxygen atom. The average distance for oxygen atoms in water molecules (which are the only neutral oxygen atoms considered here) is 2.13 Å, while that for partially negatively charged oxygen atoms (belonging to carboxylic groups) is 2.07 Å. This difference is of little significance, however, due to the high SD. Here, other variables such as the overall charge of the model and the total number of ligands may exert greater influences on the oxygen–Co distances. For negatively charged oxygen atoms (hydroxide ions), the difference is significant and the average distance is much shorter, 1.84 Å.

For nitrogen–Co bonds, the average length is 2.01 Å. The SD is also relatively high, at 0.14 Å, indicating a heterogeneous set of coordination spheres. Once more, different coordination numbers and overall charges lead to a wide range of bond lengths to the metal. Corrin nitrogen atoms are a distinct group among nitrogen atoms. They have shorter bond lengths,

Table 1 Average distances between coordinating atoms and Co

Atom	Average distance (Å)	SD
All types of oxygen	2.10	0.12
Oxygen in water molecules	2.13	0.12
Partially negatively charged oxygen	2.07	0.10
Negatively charged oxygen	1.84	0.02
All types of nitrogen	2.01	0.14
Corrin nitrogen	1.92	0.02
Non-corrin nitrogen	2.11	0.14
All types of carbon	1.94	0.05

1.92 Å on average, due to the negative charge distributed over the entire corrin ring.

Finally, carbon atoms have an average distance to the metal of 1.94 Å. A distinction can be made between the two aliphatic carbon atoms (distances: 1.96 and 1.98 Å) and the carbon atom of the cyano group (1.87 Å).

Potassium study

Statistical analysis

Number of K atoms by protein Since several of the proteins contain more than one potassium ion, Fig. 5 displays the number of proteins in the PDB that have a particular number of K coordination spheres. A total of 29 proteins (37 %) in the PDB are mononuclear K systems, while 29 % (23 occurrences) are binuclear K systems. Metalloproteins with three and four potassium coordination spheres comprise 5 % (4 proteins) and 14 % (11 proteins) of the K-proteins, respectively, whilst the remaining 15 % (12 occurrences) have more than four potassium ions. The analysis of the number of K coordination spheres indicates that coordination numbers of 2, 3, and 5 are prevalent in mononuclear K-proteins, as 18 such proteins were found (6+6+6, corresponding to 21 % each coordination sphere). Next more common are coordination numbers of 6 and 4, with five (17 %) and four (14 %) occurrences, respectively. The least prevalent K coordination spheres are those with coordination numbers of 1 and 7, with only one observed structure each (less than 4 %). In binuclear K-proteins, the most common coordination number combinations, with eight (35 %) and four (17 %) occurrences, respectively, are 6/6 and 5/5 (i.e., with both metal spheres having coordination numbers of 6 and 5, respectively). The proteins with three potassium ions show four different combinations of coordination numbers (1/7/7, 6/7/7, 3/3/4, and 2/4/6), each with only one occurrence. For K-proteins with four metal coordination spheres, a total of seven different coordination number combinations are observed. The most common combinations are 6/6/6/6 and 5/5/5/5 with four and two proteins, respectively, which illustrate the great prevalence of hexa- and pentacoordinate spheres for the potassium cations present in these metalloproteins. Figure 5 also shows the percentage occurrence for each coordination number. This analysis again highlights the prevalence of hexacoordinate (41 %), heptacoordinate (21 %), and pentacoordinate (14 %) geometries.

The nature of K coordination spheres

Figure 5 also shows an overview of potassium-binding ligands and atoms in the K proteome present in the PDB. As can be seen, the water molecule is the most common ligand in K coordination spheres (27 %), followed by the negatively charged residues Asp (9 %) and Glu (8 %). Thr represents 8 % of all K–ligand interactions, while Ser and Ala account for 6 % of the ligands present in K coordination spheres. Less common ligands are: Asn (5 %), Val, Gly, and Ile (4 %), Leu and His (3 %), Lys and Gln (2 %), and Phe (1 %). It is interesting to note that all neutral residues interact with K through their carbonyl groups, and the sum of all these residues corresponds to 22 % of all K–ligand interactions. The remaining K–ligand interactions are with other amino acid residues (7 %) and other ligands such as inhibitors (3 %). These percentages agree with those found in a previous study on the K proteome present in the PDB and Cambridge Structural Database, in which the water molecules and peptidic carbonyl groups were noted to be the most common ligands in K coordination spheres [11]. These results indicate that solvent molecules are of great importance in the K proteome due to their participation in K coordination spheres, which play a significant structural role in 3D protein folding.

As far as atoms are concerned, we can observe that oxygen (from K-binding water molecules and carbonyl groups) is the most common K-coordinating atom (96 %). Nitrogen atoms from His or N-inhibitors represent 3 % of the K-coordinating atoms, whilst all other K-coordinating atoms comprise only 1 %.

Figure 6 presents an analysis of the ligand combinations seen for the most common coordination numbers: 5, 6, and 7. For a coordination number of 5, the combination of one alanine, one glutamate, one leucine, one threonine, and one valine (A-E-L-T-V) is the most common (six occurrences), comprising 15 % of all such coordination spheres. Another combination representing a significant percentage of pentacoordinate K-proteins (10 %) is A-D-I-WAT-WAT. Interestingly, 19 of the pentacoordinate K-proteins (49 %) show unique ligand combinations, indicating high ligand variability in the pentacoordinate potassium spheres.

Among K-proteins with a coordination number of 6 (the most prevalent number for K-proteins), the ligand

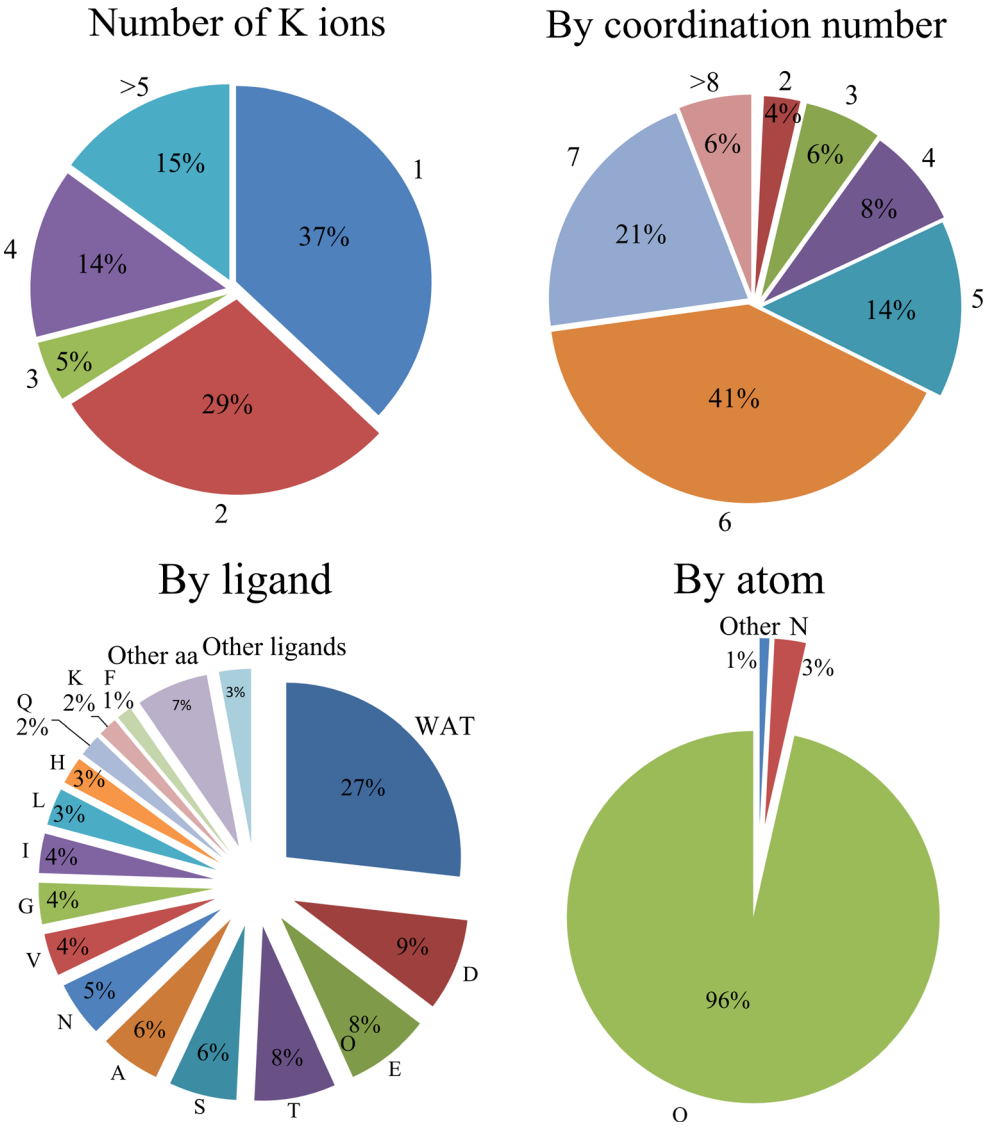

Fig. 5 Percentages of K-proteins with one, two, three, for, or five or more K coordination spheres, percentage occurrence of each K-sphere coordination number, as well as the average composition of the K coordination spheres by atom and ligand

combination D-N-S-T-WAT-WAT occurs in 7 % of all such systems (eight occurrences). This is the most common ligand combination observed in hexacoordinate K-proteins, such as muscle pyruvate kinase enzyme [69]. The combinations D-L-S-S-T-V and A-E.2 O-G-P-T are also very common in K-proteins, occurring six times each (5 %). The combination of three carbonyl backbones (from alanine, glyine, and proline), one threonine, and a bidentate glutamate is typical of formyltransferase enzymes [57]. Twenty-six of the K-proteins with this coordination number (24 %) present unique combinations of ligands, again illustrating great ligand variability.

For a K-protein coordination number of 7, the most common metal coordination sphere is K-Ot. O-T-WAT-WAT-WAT-WAT, which is present in the GroEL chaperonin protein of *E. coli* [60] and occurs in 22 % of the heptacoordinate K-proteins (13 occurrences).

Enzymatic classification

Among the 79 K-proteins with an Enzyme Commission (EC) number identified in the PDB, 54 are K-enzymes (68.4 %). Figure 7 illustrates the number of K coordination spheres for each enzyme class. There are 13, 18, 14, 5, 3, and 1 K sphere for classes I, II, III, IV, V, and VI, respectively. Enzymes in class II (transferases) catalyze the transfer of a functional group, and are the most prevalent class of enzymes among the K-enzymes present in the PDB. It was observed that the combination D-N-S-T-WAT-WAT is the most common, reinforcing the conclusion that it is also the most typical hexacoordinate potassium sphere. Enzymes in class III (hydrolases) catalyze processes involving the hydrolysis of chemical bonds, and are the second most prevalent enzyme class among the K-enzymes considered in this study. The most common combination of ligands found in this class is D-L-S-

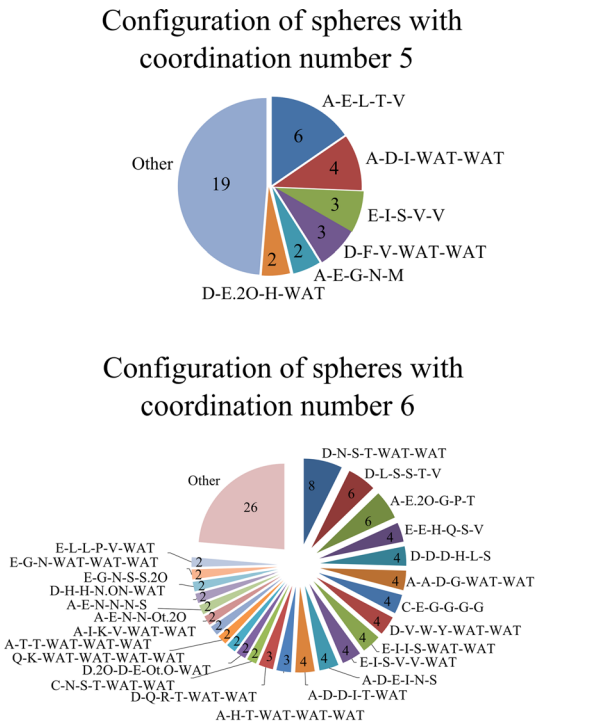

Fig. 6 Ligand configurations for K coordination spheres with coordination numbers of 5, 6, and 7. Each residue is denoted by a one-letter amino-acid code; *WAT* is a water molecule; *Ot. N* and *Ot. O* are other (substrate or inhibitor) ligands bound by a nitrogen or an oxygen atom, respectively; *E.2 O and D.2 O* represent glutamate and aspartate residues that are bound through both carboxylic oxygen atoms

S-T-V (six occurrences), followed by D-D-D-H-L-S, I-N-Ot.2 O-T-T, and D-V-W-WAT-WAT-Y (each of which occur four times). All of these combinations of ligands are associated with enzymes in which the coordination sphere of K plays an important structural role in catalysis. E-E-H-Q-S-V, A-D-D-I-T-WAT, E-I-I-S-WAT-WAT, and D-T-V-Y-WAT-WAT-WAT are the most prevalent ligand combinations in ligases, oxidoreductases, lyases, and isomerases, respectively.

Overall, this statistical analysis shows that there is a high diversity of possible ligand combinations, even for enzymes from the same enzymatic class. In general, when metalloproteins have similar ligand combinations around a

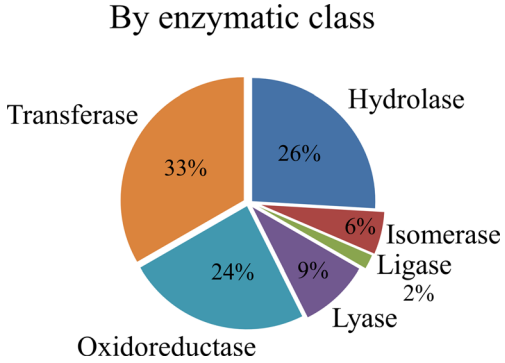

Fig. 7 Proportion of the enzymes in each enzymatic class that are potassium enzymes

particular metal, it suggests that those metalloproteins are all associated with a particular type of enzymatic catalysis. Typically, however, the potassium ion is not directly involved in catalysis; it instead plays an important structural role in proteins, which may be why a high diversity of ligand combinations was observed in the present study.

Quantum mechanical calculations

To evaluate the thermodynamic stabilities and electronic properties of the most representative K-protein coordination spheres, quantum calculations were performed for coordination numbers of 3, 4, 5, 6, and 7. A dielectric constant of 4 was used in the energy calculations to mimic the effects of the enzymatic environment on the spheres studied.

The abundance of each K-sphere configuration was used to select the spheres to be used in the calculations. The models used in the QM calculations are shown in Figs. SI-3, SI-4, and SI-5 of the ESM, while the results for the QM calculations are shown in Figs. 8 and 9.

Evidently, the stability of a specific metal coordination sphere is correlated with several characteristics of its enzymatic environment, such as distortions in the positions of the metal-binding residues, the nature of the residues in the second sphere, the backbone distribution, and the secondary structures of the protein segments. Although these aspects that are not considered in the present QM calculations, the results obtained from them highlight interesting trends regarding the intrinsic nature of each K coordination sphere with respect to its association energy.

Since K is an alkali metal, its presence in proteins is generally associated with crucial structural roles in a wide range of biological actions, such as creating adequate protein folding for enzymatic catalysis. This structural function fits with the high coordination numbers of K coordination spheres, as well as the large number of interactions between the metal and backbone groups (carbonyl or amine) or solvent molecules. According to the statistical analysis performed

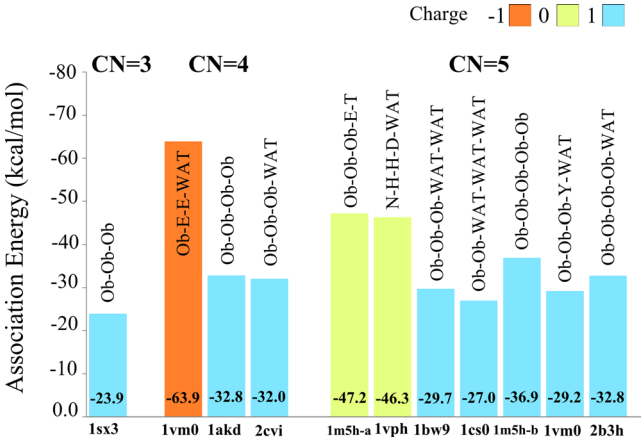

Fig. 8 Association energies of several K spheres with coordination numbers of 3, 4, and 5, as calculated with QM methods. Residues are denoted by one-letter amino-acid codes; *Ob* is an oxygen atom of a carbonyl group from the backbone segment; *CN* is the coordination number

above, the most common K coordination numbers are 6 and 7, which are associated with a wide range of representative ligand combinations (18 and 6 complexes, respectively), as shown in Figs. SI-4 and SI-5 of the ESM.

The overall charge on each K-protein complex is either −1, 0, or +1, and the complexes with a charge of −1 or 0 are more stable than tthe positively charged complexes. The calculated association energies for the four K coordination spheres with an overall charge of −1 range from −61.0 to −72.9 kcal mol^{-1}, whilst those for the eleven neutral K coordination spheres vary from −37.5 to −55.8 kcal mol^{-1}. The positively charged complexes show association energies ranging between −23.9 and −46.6 kcal mol^{-1}. In addition, it was observed that K complexes with coordination numbers of 6 and 7 are more stable than those with smaller coordination numbers (3, 5, and 5), suggesting that this alkali metal preferentially adopts geometries with large coordination numbers. Reinforcing this idea, we noticed that the only ligand combination observed for a K coordination number of 3 has an overall charge of +1 and is the unstable system according to its association energy.

Table 2 Average distances between coordinating atoms and K

Atom type	Average distance (Å)	SD
Carbonyl oxygen	2.75	0.08
Oxygen in a water molecule	2.83	0.07
Carboxylic oxygen	2.72	0.07
Hydroxyl oxygen	2.90	0.08
Phosphate oxygen	2.75	0.02
Indole nitrogen	2.76	0.08

It was also verified that the association energies of different combinations of ligands for same K coordination number can vary greatly. The three different ligand combinations for a K coordination number of 4 vary between −32.8 and −63.9 kcal mol^{-1}, while the seven combinations for a K coordination number of 5 have association energies of between −27.0 and −47.2 kcal mol^{-1}. The ligand combinations observed for coordination numbers of 6 and 7 present energetic values of −27.9 to −72.9 kcal mol^{-1} and −28.6 to −61.0 kcal mol^{-1}, respectively. It was also noted that, across all coordination numbers, different ligand combinations that result in the same overall charge have similar association energies. This fact is exemplified by the combinations Ob-E-D-Y-WAT-WAT-WAT and Ob-Ob-Ot. O-WAT-WAT-WAT-WAT for a coordination number of 7, which both have an association energy of −61.0 kcal mol^{-1}.

In general, the number of occurrences of each coordination sphere and the association energy of that sphere did not appear to be strongly correlated. However, for a coordination number of 7, the most common combination of ligands (Ob-Ob-Ot. O-WAT-WAT-WAT-WAT, with 13 occurrences) is also the most stable. The next most common combination, Ob-Ob-Ob-Ob-D.2 O-WAT, occurs three times and 5.2 kcal mol^{-1} less stable. For a coordination number of 6, the most stable complex (which has the ligand combination Ob-D.2 O-E-Ot.O-WAT) occurs four times, while the ligand combination that occurs 12 times yields a less stable complex with association energy of

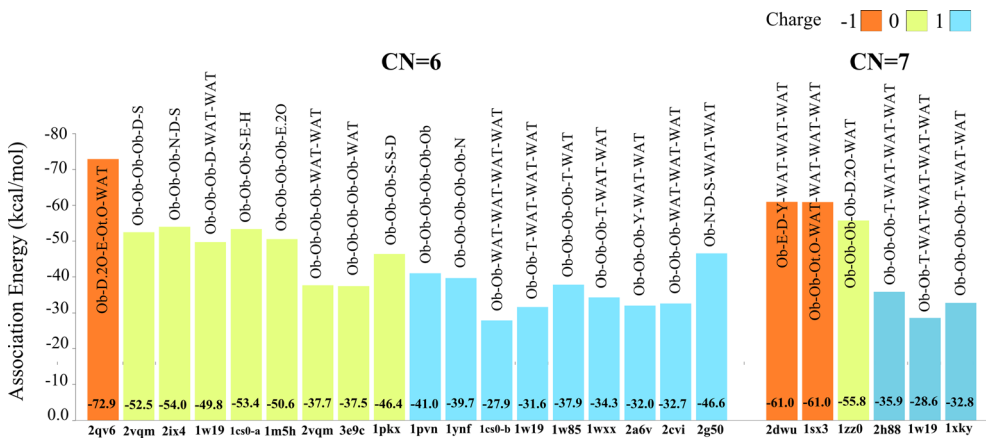

Fig. 9 Association energies of several K coordination spheres with coordination numbers of 6 and 7, as calculated with QM methods. Residues are denoted by one-letter amino-acid codes; *Ob* is an oxygen atom of a carbonyl group from the backbone segment; *CN* is the coordination number

−37.7 kcal mol^{-1}. Interestingly, in the two most stable K coordination spheres obtained for coordination numbers of 6 and 7, respectively, the strongly negative charge of the phosphate group reduces the electrostatic potential at the potassium cation, limiting its tendency to interact with other molecules. Despite their relative rarity in proteins, the ligand combinations Ob-Ob-Ob-N-D-S (which occurs eight times) and Ob-Ob-Ob-S-E-H (four times) are the second and third most thermodynamically stable hexacoordinate K spheres.

In addition to thermodynamic stability, other factors such as the functionality and the potential reactivity of the metal sphere are also important influences on the prevalence of a given K coordination sphere in nature.

Overview of K–ligand distances

Table 2 presents the average distances between the K ion and the ligand atoms coordinated to it for all of the systems examined in this work. Overall, this distance analysis points to an inverse relationship between the average oxygen–K distance and the formal charge on the oxygen atom. The smallest O–K distance is 2.72±0.07 Å, which is associated with oxygen atoms in carboxylic groups. The next smallest distance is to oxygen atoms belonging to carbonyl and phosphate groups, which have O–K distances of 2.75±0.08 Å and 2.75±0.02 Å, respectively. The largest average O–K distances are 2.90±0.08 Å (to hydroxyl oxygens) and 2.83±0.07 Å (to oxygens from water molecules). For N–K bonds, the average length is 2.76±0.08 Å. In general, it was found that the overall charge on the model and the total number of ligands in the coordination sphere exert the strongest influences on the O/N–K distances. In particular, the O–K distances increase as the coordination number increases: the carbonyl O–K distances are 2.65±0.07 Å, 2.68±0.07 Å, 2.78±0.18 Å, 2.83±0.20 Å, and 2.81±0.22 Å for coordination numbers of 3, 4, 5, 6, and 7, respectively; the water molecule O–K distances are 2.76±0.12 Å, 2.79±0.17 Å, 2.87±0.21 Å, and 2.91±0.31 Å for coordination numbers of 4, 5, 6, and 7, respectively.

Conclusions

A detailed statistical characterization of the metal coordination spheres present in the cobalt and potassium proteins currently available in the Protein Data Bank has been provided in this work. All of the results reported here should provide valuable guidelines for gaining a deeper understanding of both Co and K proteomes.

We verified that the atoms that most commonly bind cobalt are oxygen (55 %) and nitrogen (40 %). Co–carbon coordination is rarer but remarkable in its own right, as it is one of the few examples of metal–carbon bonding found in organic molecules. In terms of residues, water, histidine, and carboxylic amino acids are the most prevalent in Co coordination spheres. The corrin ring of cobalamins is also a very important ligand, with many biological roles. In the case of Co, we found that coordination spheres are only very rarely repeated in different proteins, although there are structures that contain several instances of the same coordination sphere.

We observed some interesting trends in the calculated association energies. Firstly, we found that models with an overal charge of −1 were the most stable for Co(II) systems, and that the association energy increases steadily as the charge on the model increases. No clear trend was found between the association energy and the coordination number. Secondly, we found that Co(III) systems—and the cobalamin system in particular—are much more stable than Co(II) systems. Analysis of the distances of ligands to Co showed that, in general, as the formal charge on the ligand becomes more negative, the distance to the metal decreases. This is due to the increased covalent character of those bonds, particularly those with carbon atoms and the hydroxide ion.

For potassium proteins, we noted that water molecules and peptidic carbonyl groups are the most common ligands. In terms of atoms coordinated to K, oxygen is by far the most common—it accounts for 96 % of all K–ligand interactions. Nitrogen is the next most common, accounting for 3 % of all K–ligand interactions. In general, K-proteins have one or two alkali metal cations, they also show a marked preference for hexacoordinate and heptacoordinate K-sphere geometries. However, there are also a significant number of proteins with other coordination numbers, including 3, 4, and 5. The most common hexacoordinate and heptacoordinate ligand combinations for K-proteins were DNSTWatWat (present in pyruvate kinase) and K-Ot. O-T-WAT-WAT-WAT-WAT (present in the GroEL chaperonin protein of *E. coli*). The enzyme classes that typically include potassium ions are transferases and hydrolases.

Quantum mechanical studies on the intrinsic thermodynamic stabilities of the most common K spheres for coordination numbers of 3, 4, 5, 6, and 7 shown that hexacoordinate and heptacoordinate ligand combinations are the most energetically stable. In addition, negatively charged complexes are thermodynamically favored over neutral and positively charged ones, probably due to the strong negative charge on the ligands, which reduces the electrostatic potential at the cation and thus makes it less inclined to interact with other molecules.

Acknowledgments The authors would like to acknowledge the program FEDER/COMPETE and the Fundação para a Ciência e Tecnologia (FCT) due to the financial support provided (project PTDC/QUI-QUI/103118/2008).

References

1. Harding MM, Nowicki MW, Walkinshaw MD (2010) Metals in protein structures: a review of their principal features. Crystallogr Rev 16:247–302. doi:10.1080/0889311X.2010.485616
2. Maret W (2010) Metalloproteomics, metalloproteomes, and the annotation of metalloproteins. Metallomics 2:117–125. doi:10.1039/B915804A
3. Yamashita MM, Wesson L, Eisenman G, Eisenberg D (1990) Where metal ions bind in proteins. Proc Natl Acad Sci USA 87:5648–5652. doi:10.1073/pnas.87.15.5648
4. Harding MM (2006) Small revisions to predicted distances around metal sites in proteins. Acta Crystallogr D 62:678–682. doi:10.1107/S0907444906014594
5. Dokmanić I, Šikić M, Tomić S (2008) Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination. Acta Crystallogr D 64:257–263. doi:10.1107/S090744490706595X
6. Zheng H, Chruszcz M, Lasota P, Lebioda L, Minor W (2008) Data mining of metal ion environments present in protein structures. J Inorg Biochem 102:1765–1776. doi:10.1016/j.jinorgbio.2008.05.006
7. Tamames JAC, Ramos MJ (2011) Metals in proteins: cluster analysis studies. J Mol Model 17:429–442. doi:10.1007/s00894-010-0733-5
8. Tamames B, Sousa SF, Tamames J, Fernandes PA, Ramos MJ (2007) Analysis of zinc–ligand bond lengths in metalloproteins: trends and patterns. Proteins 69:466–475. doi:10.1002/prot.21536
9. Sousa SF, Lopes AB, Fernandes PA, Ramos MJ (2009) The zinc proteome: a tale of stability and functionality. Dalton Trans 7946–7956. doi:10.1039/B904404C
10. Abriata LA (2013) Investigation of non-corrin cobalt(II)-containing sites in protein structures of the Protein Data Bank. Acta Crystallogr B 69:176–183. doi:10.1107/S2052519213002959
11. Harding MM (2002) Metal–ligand geometry relevant to proteins and in proteins: sodium and potassium. Acta Crystallogr D 58:872–874. doi:10.1107/S0907444902003712
12. Berman H, Henrick K, Nakamura H, Markley JL (2007) The Worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res 35:D301–D303. doi:10.1093/nar/gkl971
13. Allen FH (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. Acta Crystallogr B 58:380–388. doi:10.1107/S0108768102003890
14. Hsin K, Sheng Y, Harding MM, Taylor P, Walkinshaw MD (2008) MESPEUS: a database of the geometry of metal sites in proteins. J Appl Crystallogr 41:963–968. doi:10.1107/S002188980802476X
15. Hemavathi K, Kalaivani M, Udayakumar A, Sowmiya G, Jeyakanthan J, Sekar K (2009) MIPS: metal interactions in protein structures. J Appl Crystallogr 43:196–199. doi:10.1107/S002188980903982X
16. Tus A, Rakipovic A, Peretin G, Tomic S, Sikic M (2012) BioMe: biologically relevant metals. Nucleic Acids Res 40:W352–W357. doi:10.1093/nar/gks514
17. Karlin S, Zhu Z-Y, Karlin KD (1997) The extended environment of mononuclear metal centers in protein structures. Proc Natl Acad Sci USA 94:14225–14230
18. Irving H, Williams RJP (1953) The stability of transition-metal complexes. J Chem Soc 3192–3210. doi:10.1039/JR9530003192
19. Brown KL (2005) Chemistry and enzymology of vitamin B12. Chem Rev 105:2075–2150. doi:10.1021/cr030720z
20. Randaccio L, Geremia S, Demitri N, Wuerges J (2010) Vitamin B12: unique metalorganic compounds and the most complex vitamins. Molecules 15:3228–3259. doi:10.3390/molecules15053228
21. Clardy SM, Allis DG, Fairchild TJ, Doyle RP (2011) Vitamin B12 in drug delivery: breaking through the barriers to a B12 bioconjugate pharmaceutical. Expert Opin Drug Deliv 8:127–140. doi:10.1517/17425247.2011.539200
22. Kobayashi M, Shimizu S (1999) Cobalt proteins. Eur J Biochem 261:1–9. doi:10.1046/j.1432-1327.1999.00186.x
23. Payne LR (1977) The hazards of cobalt. Occup Med 27:20–25. doi:10.1093/occmed/27.1.20
24. Tower SS (2012) Arthroprosthetic cobaltism associated with metal on metal hip implants. BMJ 344:e430–e430. doi:10.1136/bmj.e430
25. Abriata LA, González LJ, Llarrull LI, Tomatis PE, Myers WK, Costello AL, Tierney DL, Vila AJ (2008) Engineered mononuclear variants in *Bacillus cereus* metallo-β-lactamase BcII are inactive. Biochemistry (Mosc) 47:8590–8599. doi:10.1021/bi8006912
26. Llarrull LI, Tioni MF, Vila AJ (2008) Metal content and localization during turnover in *B. cereus* metallo-β-lactamase. J Am Chem Soc 130:15842–15851. doi:10.1021/ja801168r
27. Maret W, Vallee B (1993) Cobalt as probe and label of proteins. Methods Enzymol 226:52–71
28. Bhosale SH, Rao MB, Deshpande VV (1996) Molecular and industrial aspects of glucose isomerase. Microbiol Rev 60:280–300
29. Page MJ, Cera ED (2006) Role of Na^+ and K^+ in enzyme function. Physiol Rev 86:1049–1092. doi:10.1152/physrev.00008.2006
30. Cera ED (2006) A structural perspective on enzymes activated by monovalent cations. J Biol Chem 281:1305–1308. doi:10.1074/jbc.R500023200
31. Harding MM (2001) Geometry of metal–ligand interactions in proteins. Acta Crystallogr D 57:401–411. doi:10.1107/S0907444900019168
32. Shibata N, Masuda J, Tobimatsu T, Toraya T, Suto K, Morimoto Y (1993) Yasuoka N (1999) A new mode of B_{12} binding and the direct participation of a potassium ion in enzyme catalysis: X-ray structure of diol dehydratase. Struct Lond Engl 7:997–1008
33. Evans HJ, Sorger GJ (1966) Role of mineral elements with emphasis on the univalent cations. Annu Rev Plant Physiol 17:47–76. doi:10.1146/annurev.pp.17.060166.000403
34. Andersson CE, Mowbray SL (2002) Activation of ribokinase by monovalent cations. J Mol Biol 315:409–419. doi:10.1006/jmbi.2001.5248
35. Ahmad A, Akhtar MS, Bhakuni V (2001) Monovalent cation-induced conformational change in glucose oxidase leading to stabilization of the enzyme. Biochemistry (Mosc) 40:1945–1955
36. Babu CS, Dudev T, Casareno R, Cowan JA, Lim C (2003) A combined experimental and theoretical study of divalent metal ion selectivity and function in proteins: application to *E. coli* ribonuclease H1. J Am Chem Soc 125:9318–9328. doi:10.1021/ja034956w
37. Ryde U (2007) Accurate metal-site structures in proteins obtained by combining experimental data and quantum chemistry. Dalton Trans 607–625. doi:10.1039/B614448A
38. Brás NF, Fernandes PA, Ramos MJ (2010) QM/MM studies on the β-galactosidase catalytic mechanism: hydrolysis and transglycosylation reactions. J Chem Theory Comput 6:421–433. doi:10.1021/ct900530f
39. Alberto ME, Marino T, Ramos MJ, Russo N (2010) Atomistic details of the catalytic mechanism of Fe(III)–Zn(II) purple acid phosphatase. J Chem Theory Comput 6:2424–2433. doi:10.1021/ct100187c
40. Ribeiro AJM, Ramos MJ, Fernandes PA (2012) The catalytic mechanism of HIV-1 integrase for DNA 3′-end processing established by QM/MM calculations. J Am Chem Soc 134:13436–13447. doi:10.1021/ja304601k
41. Ramos MJ, Fernandes PA (2008) Computational enzymatic catalysis. Acc Chem Res 41:689–698. doi:10.1021/ar7001045
42. Sousa SF, Fernandes PA, Ramos MJ (2009) The search for the mechanism of the reaction catalyzed by farnesyltransferase. Chem Eur J 15:4243–4247. doi:10.1002/chem.200802745
43. Blomberg MRA, Borowski T, Himo F, Liao R-Z, Siegbahn PEM (2014) Quantum chemical studies of mechanisms for metalloenzymes. Chem Rev (in press). doi:10.1021/cr400388t

44. Román-Meléndez GD, von Glehn P, Harvey JN, Mulholland AJ, Marsh ENG (2014) Role of active site residues in promoting cobalt–carbon bond homolysis in adenosylcobalamin-dependent mutases revealed through experiment and computation. Biochemistry (Mosc) 53:169–177. doi:10.1021/bi4012644
45. Zhu X, Teng M, Niu L, Xu C, Wang Y (1999) Structure of xylose isomerase from *Streptomyces diastaticus* no. 7 strain M1033 at 1.85 Å resolution. Acta Crystallogr D 56:129–136. doi:10.1107/S0907444999015097
46. Masuda J, Shibata N, Morimoto Y, Toraya T, Yasuoka N (2000) How a protein generates a catalytic radical from coenzyme B (12): X-ray structure of a diol-dehydratase-adeninylpentylcobalamin complex. Struct Fold Des 8:775–788. doi:10.1016/S0969-2126(00)00164-7
47. Iverson TM, Alber BE, Kisker C, Ferry JG, Rees DC (1999) A closer look at the active site of gamma-class carbonic anhydrases: high-resolution crystallographic studies of the carbonic anhydrase from *Methanosarcina thermophila*. Biochemistry (Mosc) 39:9222–9231. doi:10.1021/bi000204s
48. Hall PR, Zheng R, Antony L, Pusztai-Carey M, Carey PR, Yee VC (2003) Transcarboxylase 5S structures: assembly and catalytic mechanism of a multienzyme complex subunit. Embo J 23:3621–3631. doi:10.1038/sj.emboj.7600373
49. Casares S, Lopez-Mayorga O, Vega MC, Camara-Artigas A, Conejero-Lara F (2005) Cooperative propagation of local stability changes from low-stability and high-stability regions in a SH3 domain. Proteins 67:531–547. doi:10.1002/prot.21284
50. Svetlitchnaia T, Svetlitchnyi V, Meyer O, Dobbek H (2006) Structural insights into methyltransfer reactions of a corrinoid iron–sulfur protein involved in acetyl-CoA synthesis. Proc Natl Acad Sci USA 103:14331–14336. doi:10.1073/pnas.0601420103
51. Flynn GE, Black KD, Islas LD, Sankaran B, Zagotta WN (2007) Structure and rearrangements in the carboxy-terminal region of SpIH channels. Structure 15:671–682. doi:10.1016/j.str.2007.04.008
52. Palm GJ, Lederer T, Orth P, Saenger W, Takahashi M, Hillen W, Hinrichs W (2007) Specific binding of divalent metal ions to tetracycline and to the Tet repressor/tetracycline complex. J Biol Inorg Chem 13:1097. doi:10.1007/S00775-008-0395-2
53. Puorger C, Eidam O, Capitani G, Erilov D, Grutter MG, Glockshuber R (2007) Infinite kinetic stability against dissociation of supramolecular protein complexes through donor strand complementation. Structure 16:631–642. doi:10.1016/j.str.2008.01.013
54. Schlichting I, Jung C, Schulze H (1997) Crystal structure of cytochrome P-450cam complexed with the (1*S*)-camphor enantiomer. FEBS Lett 415:253–257. doi:10.1016/S0014-5793(97)01135-6
55. Vanhooke JL, Thoden JB, Brunhuber NM, Blanchard JS, Holden HM (1998) Phenylalanine dehydrogenase from *Rhodococcus sp*. M4: high-resolution X-ray analyses of inhibitory ternary complexes reveal key features in the oxidative deamination mechanism. Biochemistry (Mosc) 38:2326–2339. doi:10.1021/bi982244q
56. Thoden JB, Huang X, Raushel FM, Holden HM (1999) The small subunit of carbamoyl phosphate synthetase: snapshots along the reaction pathway. Biochemistry (Mosc) 38:16158–16166. doi:10.1021/bi991741j
57. Mamat B, Roth A, Grimm C, Ermler U, Tziatzios C, Schubert D, Thauer RK, Shima S (2002) Crystal structures and enzymatic properties of three formyltransferases from archaea: environmental adaptation and evolutionary relationship. Protein Sci 11:2168–2178. doi:10.1110/ps.0211002
58. Wolan DW, Cheong CG, Greasley SE, Wilson IA (2003) Structural insights into the human and avian IMP cyclohydrolase mechanism via crystal structures with the bound XMP inhibitor. Biochemistry (Mosc) 43:1171–1183. doi:10.1021/bi030162i
59. Gan L, Seyedsayamdost M, Shuto S, Matsuda A, Petsko GA, Hedstrom L (2003) The immunosuppressive agent mizoribine monophosphate forms a transition state analogue complex with Inosine monophosphate dehydrogenase. Biochemistry (Mosc) 42:857–863. doi:10.1021/bi0271401
60. Chaudhry C, Horwich AL, Brunger AT, Adams PD (2004) Exploring the structural dynamics of the *E. coli* chaperonin GroEL using translation-libration-screw crystallographic refinement of intermediate states. J Mol Biol 342:229–245. doi:10.1016/j.jmb.2004.07.015
61. Morgunova E, Meining W, Illarionov B, Haase I, Jin G, Bacher A, Cushman M, Fischer M, Ladenstein R (2004) Crystal structure of lumazine synthase from *Mycobacterium tuberculosis* as a target for rational drug design: binding mode of a new class of purinetrione inhibitors. Biochemistry (Mosc) 44:2746. doi:10.1021/BI047848A
62. Pioszak AA, Murayama K, Nakagawa N, Ebihara A, Kuramitsu S, Shirouzu M, Yokoyama S (2005) Structures of a putative RNA 5-methyluridine methyltransferase, *Thermus thermophilus* TTHA1280, and its complex with *S*-adenosyl-L-homocysteine. Acta Crystallogr F 61:867–874. doi:10.1107/S1744309105029842
63. Blagova E, Levdikov V, Milioti N, Fogg MJ, Kalliomaa AK, Brannigan JA, Wilson KS, Wilkinson AJ (2004) Crystal structure of dihydrodipicolinate synthase (BA3935) from *Bacillus anthracis* at 1.94 Å resolution. Proteins 62:297–301. doi:10.1002/prot.20684
64. Tocilj A, Schrag JD, Li Y, Schneider BL, Reitzer L, Matte A, Cygler M (2005) Crystal structure of *N*-succinylarginine dihydrolase AstB, bound to substrate and product, an enzyme from the arginine catabolic pathway of *Escherichia coli*. J Biol Chem 280:15800–15808. doi:10.1074/jbc.M413833200
65. Nielsen TK, Hildmann C, Dickmanns A, Schwienhorst A, Ficner R (2005) Crystal structure of a bacterial class 2 histone deacetylase homologue. J Mol Biol 354:107–120. doi:10.1016/j.jmb.2005.09.065
66. Satoh T, Sato K, Kanoh A, Yamashita K, Yamada Y, Igarashi N, Kato R, Nakano A, Wakatsuki S (2005) Structures of the carbohydrate recognition domain of Ca^{2+}-independent cargo receptors Emp46p and Emp47p. J Biol Chem 281:10410–10419. doi:10.1074/jbc.M512258200
67. Addlagatta A, Hu X, Liu JO, Matthews BW (2005) Structural basis for the functional differences between type I and type II human methionine aminopeptidases. Biochemistry (Mosc) 44:14741–14749. doi:10.1021/bi051691k
68. May M, Mehboob S, Mulhearn DC, Wang Z, Yu H, Thatcher GRJ, Santarsiero BD, Johnson ME, Mesecar AD (2006) Structural and functional analysis of two glutamate racemase isozymes from *Bacillus anthracis* and implications for inhibitor design. J Mol Biol 371:1219–1237. doi:10.1016/j.jmb.2007.05.093
69. Williams R, Holyoak T, McDonald G, Gui C, Fenton AW (2006) Differentiating a ligand's chemical requirements for allosteric interactions from those for protein binding. Phenylalanine inhibition of pyruvate kinase. Biochemistry (Mosc) 45:5421–5429. doi:10.1021/bi0524262
70. Scrima A, Wittinghofer A (2006) Dimerisation-dependent GTPase reaction of MnmE: how potassium acts as GTPase-activating element. Embo J 25:2940–2951. doi:10.1038/sj.emboj.7601171
71. Huang LS, Shen JT, Wang AC, Berry EA (2006) Crystallographic studies of the binding of ligands to the dicarboxylate site of Complex II, and the identity of the ligand in the "oxaloacetate-inhibited" state. Biochim Biophys Acta 1757:1073–1083. doi:10.1016/j.bbabio.2006.06.015
72. Christensen CE, Kragelund BB, Von Wettstein-Knowles P, Henriksen A (2006) Structure of the human beta-ketoacyl [Acp] synthase from the mitochondrial type II fatty acid synthase. Protein Sci 16:261. doi:10.1110/PS.062473707
73. Morrison SD, Roberts SA, Zegeer AM, Montfort WR, Bandarian V (2007) A new use for a familiar fold: the X-ray crystal structure of GTP-bound GTP cyclohydrolase III from *Methanocaldococcus jannaschii* reveals a two metal ion catalytic mechanism. Biochemistry (Mosc) 47:230–242. doi:10.1021/bi701782e

74. Bottomley MJ, Lo Surdo P, Di Giovine P, Cirillo A, Scarpelli R, Ferrigno F, Jones P, Neddermann P, De Francesco R, Steinkuhler C, Gallinari P, Carfi A (2008) Structural and functional analysis of the human Hdac4 catalytic domain reveals a regulatory zinc-binding domain. J Biol Chem 283:26694–26704. doi:10.1074/jbc.M803514200
75. Brás NF, Moura-Tamames SA, Fernandes PA, Ramos MJ (2008) Mechanistic studies on the formation of glycosidase-substrate and glycosidase-inhibitor covalent intermediates. J Comput Chem 29: 2565–2574. doi:10.1002/jcc.21013
76. Brás NF, Perez MAS, Fernandes PA, Silva PJ, Ramos MJ (2011) Accuracy of density functionals in the prediction of electronic proton affinities of amino acid side chains. J Chem Theory Comput 7:3898–3908. doi:10.1021/ct200309v
77. Liao R-Z, Yu J-G, Himo F (2011) Quantum chemical modeling of enzymatic reactions: the case of decarboxylation. J Chem Theory Comput 7:1494–1501. doi:10.1021/ct200031t
78. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 98:5648–5652. doi:10.1063/1.464913
79. Lee C, Yang W, Parr RG (1988) Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B 37:785–789. doi:10.1103/PhysRevB.37.785
80. Ditchfield R, Hehre WJ, Pople JA (2003) Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. J Chem Phys 54:724–728. doi:10.1063/1.1674902
81. Andrae D, Häußermann U, Dolg M, Stoll H, Preuß H (1990) Energy-adjusted ab initio pseudopotentials for the second and third row transition elements. Theor Chim Acta 77:123–141. doi:10.1007/BF01114537
82. Frisch M, Trucks G, Schlegel H, Scuseria G, Robb M, Cheeseman J, Scalmani G, Barone V, Mennucci B, Petersson G, Nakatsuji H, Caricato M, Li X, Hratchian H, Izmaylov A, Bloino J, Zheng G, Sonnenberg J, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Jr, Peralta J, Ogliaro F, Bearpark M, Heyd J, Brothers E, Kudin K, Staroverov V, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant J, Iyengar S, Tomasi J, Cossi M, Rega N, Millam J, Klene M, Knox J, Cross J, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann R, Yazyev O, Austin A, Cammi R, Pomelli C, Ochterski J, Martin R, Morokuma K, Zakrzewski V, Voth G, Salvador P, Dannenberg J, Dapprich S, Daniels A, Farkas, Foresman J, Ortiz J, Cioslowski J, Fox D (2009) Gaussian 09, revision A.02. Gaussian Inc., Wallingford
83. Cossi M, Rega N, Scalmani G, Barone V (2003) Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. J Comput Chem 24:669–681. doi:10.1002/jcc.10189
84. Barone V, Cossi M (1998) Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. J Phys Chem A 102:1995–2001. doi:10.1021/jp9716997
85. Siegbahn PEM, Eriksson L, Himo F, Pavlov M (1998) Hydrogen atom transfer in ribonucleotide reductase (RNR). J Phys Chem B 102:10622–10629. doi:10.1021/jp9827835
86. Siegbahn PEM (1998) Theoretical study of the substrate mechanism of ribonucleotide reductase. J Am Chem Soc 120:8417–8429. doi:10.1021/ja9736065
87. Blomberg MRA, Siegbahn PEM, Babcock GT (1998) Modeling electron transfer in biochemistry: a quantum chemical study of charge separation in *Rhodobacter sphaeroides* and Photosystem II. J Am Chem Soc 120:8812–8824. doi:10.1021/ja9805268
88. Odintsov SG, Sabala I, Bourenkov G, Rybin V, Bochtler M (2005) *Staphylococcus aureus* aminopeptidase S is a founding member of a new peptidase clan. J Biol Chem 280:27792–27799. doi:10.1074/jbc.M502023200