Niklaus Sidler
14 August

Lead and Proteins: A Study of Coordination Sites

Proteins

The proteins for this study were found in the RCSB protein data base (PDB), with the help of the *MetalPDB* from the University of Florence (CERM). A total of 57 protein crystal structures were identified, which contained at least one lead (Pb) atom, leading to a total of 165 documented Pb atoms within these structures. Among the total Pb binding sites, 54 (=33%) were occupied with the complex $[Pb(Me)_3]^+$ and the other 111 (=67%) sites with discrete Pb(II) ions. The proteins themselves range from RNA pol. holoenzymes for *thermus thermophilus*,[1] over calmodulin for *homo sapiens*,[2] to NEF proteins for *HIV*,[3] and therefore form a diverse array of proteins. Special emphasis was put on proteins that already contained metal centres,[4] which were now occupied by Pb and also on proteins that were specifically designed for Pb(II) coordination.[4]

General Pb Coordination Sites

No clear assumptions, about stable Pb complexes, could be made when looking at all 165 documented PDB Pb atoms at once. Pb is mainly used to aid crystallization for X-Ray diffraction spectroscopy (xrd).[5] These Pb atoms are usually loosely bound to the proteins, with low coordination numbers (CN) being favoured over high CN. It is notable that one of the most abundant CN, when looking at all ligands, is five (*Figure 1*, left). This is due to the $[Pb(IV)Me_3]^+$ species, where a majority of the Pb has the three Me ligands and also coordinates to one amino acid (bidentate), giving these Pb atoms a CN of five.

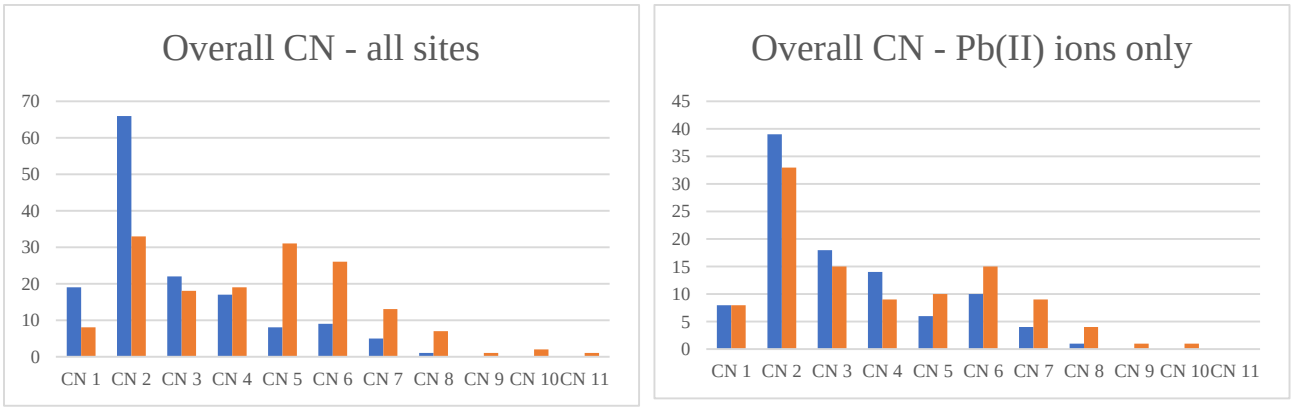


Figure 1. CN of all Pb atoms (left) and only the Pb(II) atoms (right), the CN when looking at the protein interactions are displayed in blue and the overall CN are displayed in orange.

There is also a noticeable difference between the CN of the two Pb species. Pb(II) has on average higher CN than $Pb(IV)Me_3$, when looking at the coordination with the proteins themselves. The opposite is observed when all ligands are taken into account.

Still, a few observations can be made, mainly with regards to which AA coordinate to the Pb atoms and how. The relative abundance of the different AAs can be seen below (*Table 1*). Asp and Glu, both with carboxylic acid in their side chains, are the most abundant AAs and make up almost 60% of the AAs found in Pb

coordination sites (CS). Cys can also be found frequently and the coordination of Cys to Pb(II) will be further examined in a later chapter (*Pb Specific Coordination Sites*). The high affinity for both, a hard ligand and a soft ligand is in agreement with Pb(II) being classified as a borderline acid, according to the HSAB concept. The theory that Pb(IV)Me_3 might be softer than Pb(II), is not supported by the collected data as Cys has about the same abundancy for both species.[6]

Other AAs such as His, Asn, *etc.* can also be encountered in rare cases. Pb can also bind to the backbone (BB) of a protein, which it does exclusively to the oxygen atom of the amide bond and not to the nitrogen. This is due to the lone pair of the latter being delocalized in a resonating double bond.[7] With relation to all other AAs this type of coordination makes up 13.7%.

Amino Acid	Relative Abundance
Asp	32.3%
Glu	26.5%
Cys	15.3%
His	4.5%
Asn	2.6%
Arg	1.3%
Tyr	1.0%
Ser	1.0%
Lys	1.0%
Gln	0.6%
Thr	0.3%

Table 1. Relative abundance of different AAs in Pb coordination sites (through side chains only).

A differentiation was made between the monodentate (mono) and bidentate (bi) coordination to the Pb atom if such was possible (*Table 2*). For Asp, the ration is almost 1.1, whereas for Glu the ratio is 0.4 in favour of the bi-coordinating carboxylic acid group. For Asn, Arg and Gln the ratio is always in favour of the mono-coordination, but it is difficult to assess the precision of statement, as the sample size for the last three AAs is rather small.

Amino Acid	Mono to bi ratio
Asp	1.1
Glu	0.4
Asn	7
Gln	2*
Arg	3

Table 2. Ratio between mono and bi coordination for different AAs (* no bidentate Gln coordination was observed).

The difference between the Glu and Asp mono to di ratios might be explained by the additional methylene group, which leads to a higher flexibility when coordinating and therefore lowering the energy penalty, by

avoiding steric effects. This is under the assumption that the bidentate coordination is energetically more favoured, as HOMO of the carboxyl group might have a more favourable overlap with the LUMO of Pb when coordinating in a bidentate manner.[8] This assumption is further supported by the fact that the bidentate coordination is more prevalent among low CN sites, where steric effects are less important. This would also agree findings of *Dudev et al.* where electronic effects were determined to have a lower influence on the coordination (when comparing Asp to Glu) than other effects.[9]

Pb Replacing Other Metals

Calcium (Ca)

Ca(II) is the most frequently replaced metal ion (17 instances) in the data set and it is always replaced by Pb(II). The total CN range from three to ten, with a noticeable difference between protein and complete coordination, mainly caused by the additional coordination of H_2O in eleven out of the 17 studied cases. The amount of H_2O molecules per CS ranges from one to two.

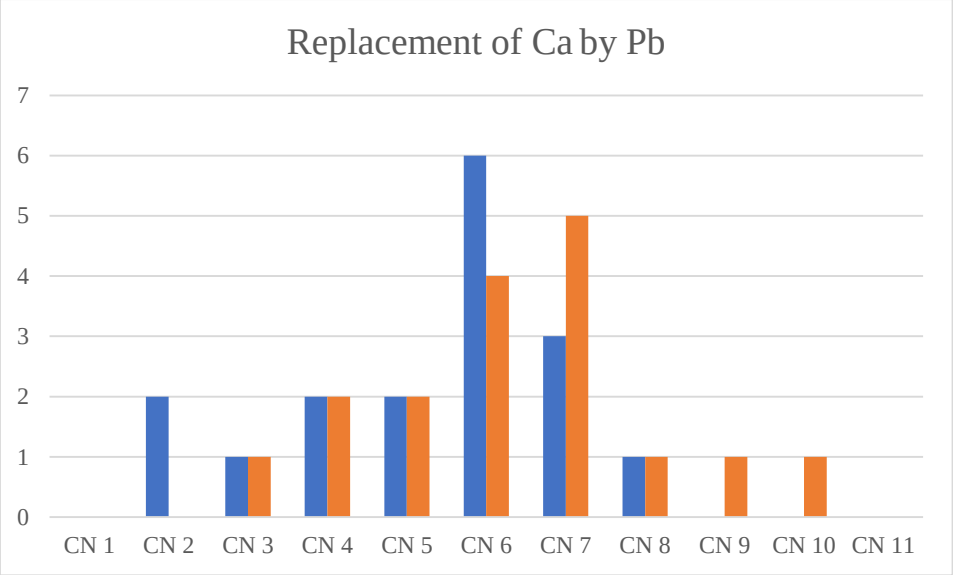


Figure 2. CN of Pb atoms with only the protein (blue) and with the protein and molecules present in the crystal structure (orange).

There is a lower variety of AAs in these cases, with the majority being Asp. The oxygen atoms of the peptide BB also have a notable impact on the coordination, and they can be found in an 18.2% abundance in the CS. The ratios between mono- and di-coordination stay similar to the ratios described for all CS, although small discrepancies can be observed (*Table 4*).

Amino Acid	Relative Abundance
Asp	62.1%
Glu	15.2%
Asn	4.5%

Table 4. Relative abundance of different AAs in Pb coordination sites (through side chains only), when replacing Ca.

The crystal structure of the protein with the PDB code 3twy shows an interesting property of Pb(II), its orbital structure. Pb(II) CS usually have a hemidirected geometry, which means that the full 6s orbital expands on one half of the CS, forcing all other ligands onto the other half.[10]

Magnesium (Mg)

Replacement of Mg(II) by Pb was observed six times, four times with Pb(II) and twice with Pb(IV)Me_3. The overall CN vary between five and eight (*Figure 2*).

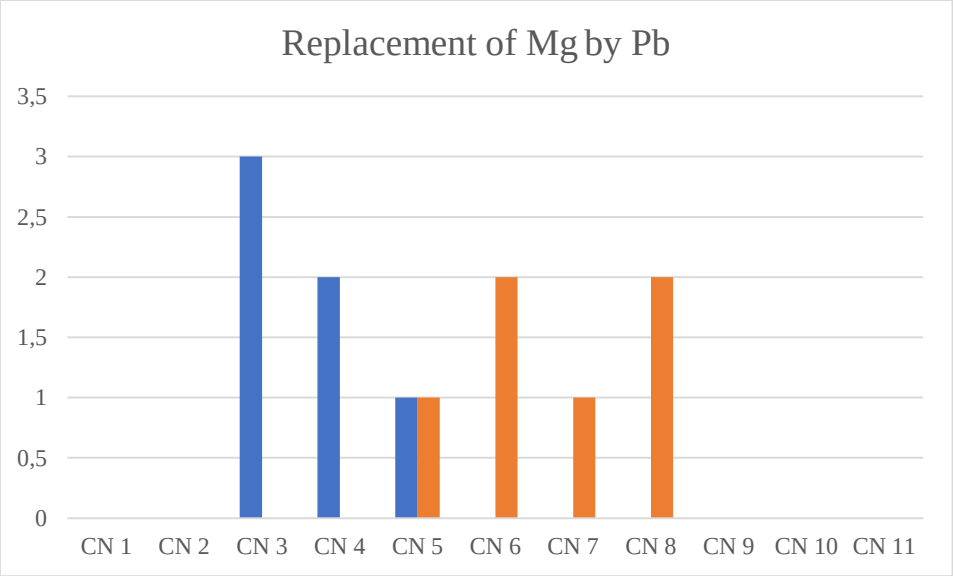


Figure 3. CN of Pb atoms with only the protein (blue) and with the protein and molecules present in the crystal structure (orange).

The CS contain an abundance of Cys and Asp (*Table 3*). The ratios between mono- and di-coordination show the same tendencies as the ratios reported above for all CS. Additionally, the coordination of H_2O was observed in three of the six CS, with the number of water molecules ranging from one to five. The data also shows that there is no coordination to the BB of the proteins. It is unclear if this observation is due to the low sample size, as BB coordination has been reported in Mg(II) coordination sites.[11]

Amino Acid	Relative Abundance
Cys	38.9%
Asp	33.3%
Glu	5.6%
His	5.6%
Asn	5.6%
Thr	5.6%
Ser	5.6%

Table 3. Relative abundance of different AAs in Pb coordination sites (through side chains only), when replacing Mg.

This low sample size and high variety among CS, with regards to the CN and subsequent structure, allows only for a few coherent statements with the given theory. But it can be said, that the CN lager than four suggest the involvement of the 6d orbitals (at least to a certain degree) in the formation of the metal-ligand bonds.[12]

Zinc (Zn)

Only one Zn(II) ion was replaced by a Pb(II) ion in a CS containing three Cys side chains in its inner and one oxygen atom of the BB in its outer coordination sphere. A tetrahedral configuration is usually preferred by Zn(II) ions but not by Pb(II), leading to a rather long distance between the Pb(II) and the oxygen atom of 3.4 Å.[13] An explanation for this phenomenon can be found again when looking at the orbitals of Pb(II). The 6s and 6p hybridization is energetically disfavoured,[12] the ligands therefore coordinate mainly via the three 6p orbitals, pushing the full 6s orbital to the other side of the coordination sphere (hemidirected). The resulting electronic repulsion between the lone pairs leads to the longer distance between the Pb(II) and the oxygen atom (**3.2**, *Figure 6*).[10]

Lead Specific Coordination Sites

Proteins that were specifically designed to bind Pb(II) were reported in five cases.[14,15] All of them have the same binding site, containing three Cys side chains (*Figure 4*). No other coordinating side chains were observed and also no other ligands such as H_2O. *Ruckthong et al.* suggest that there is a high energy penalty for the coordination of H_2O, due to the optimal binding pattern of the CS.[10]

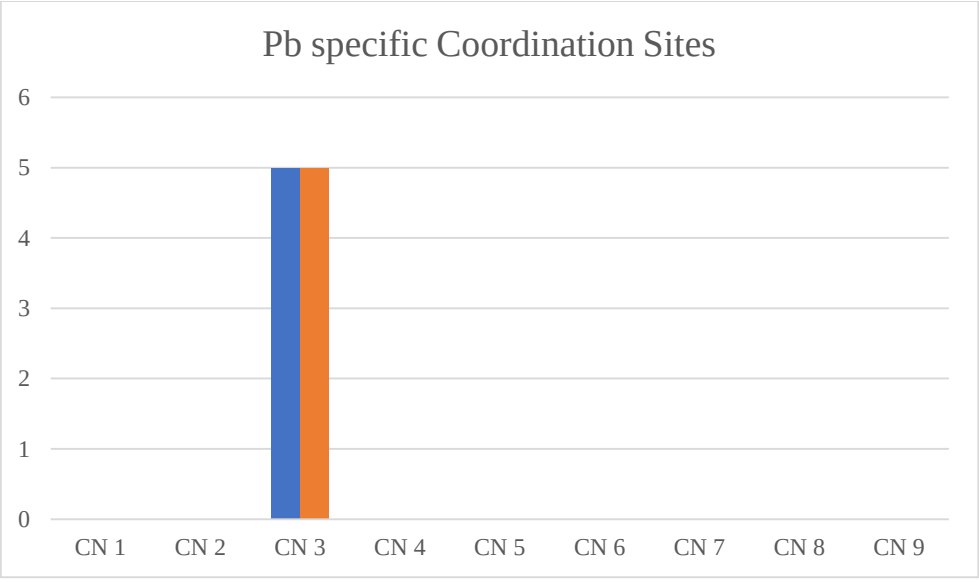


Figure 4. CN of Pb atoms with only the protein (blue) and with the protein and molecules present in the crystal structure (orange).

The geometry of the CS could only be studied in one instance, for the protein with the PDB code 5gpe. This revealed a three-fold coordination with a 90° angel between the ligands (**3.2**, *Figure 6*). Similar to the Zn(II) binding site mentioned above, the outer coordination sphere also contained the oxygen atom of an Asn side-chain with a 3.4 Å between the two. This is further in agreement with the model mentioned when talking about Pb(II) replacing Zn(II).

The research done by *Huang et al.* also suggests that the specificity of the protein towards a specific metal ion can be modelled by forcing a certain geometry *via* the structure of the CS. The more rigid a system, the more specific it is.[14,15] For Pb(II), this would be a CS with a structure similar to the one described above. In turn can a more flexible system be used when a more diverse group of metal ions is targeted.[15] This might be achieved by placing the Cys ligands in different positions of the sequence, when synthesizing a small polypeptide for heavy metal chelation.

Correlation Between the Ligands

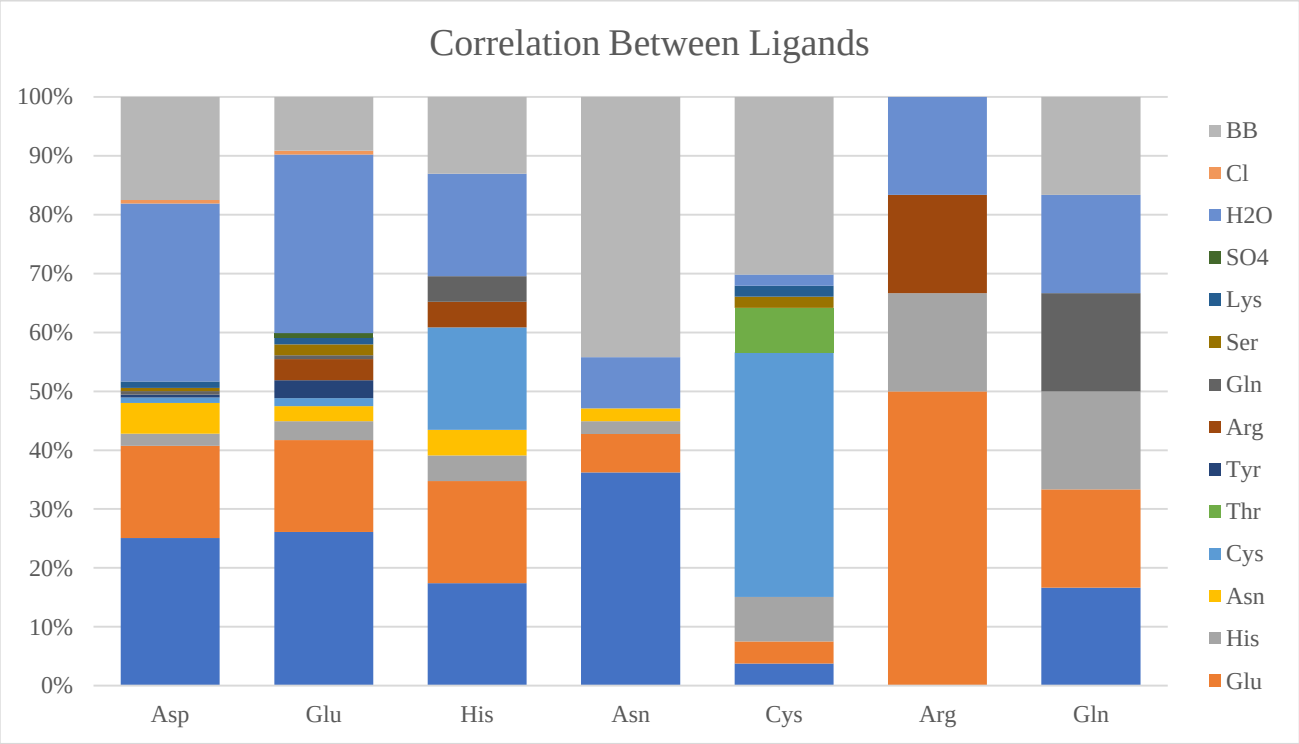
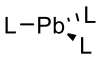


Figure 5. The correlation between certain AAs and the other ligands found in the same CS.

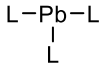
The correlation between the different ligands can be seen above (*Figure 5*). The distribution for each AA generally follows the relative abundance of the AAs in the different CS. Exceptions can mainly be observed for ligands that only have been observed a few times, such as Asn, Thr, Gln. And for the high abundance ligands His, which is rarely found in the same CS as Asp and more frequently with Cys and also Cys, which is more likely to be present with another Cys, His and also a BB bond. It is also interesting to note that H_2O and Cys are rarely found in the same CS.

Geometries

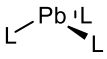
CN = 3



1: trigonal planar



2: T-shaped

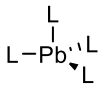


3.1: trigonal pyramidal

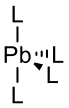
3.2: bent 3.1

CN = 4

4.1: tetrahedral

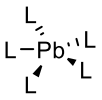


4.2: bent 4.1

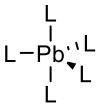


5: seesaw

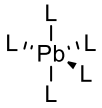
CN = 5



6: pentagonal planar

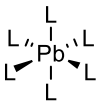


7.1: trigonal bipyramida

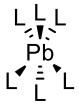


7.2: bent 7.1

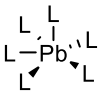
CN = 6



8: octahedral



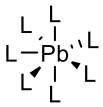
9: trigonal antiprism



10: pentagonal pyramidal

CN = 7

11: tetragonal base-trigonal top



12: pentagonal bipyramidal

CN = 8

13: tetragonal antiprism

CN = 9

14: pentagonal base-tetragonal top

Figure 6. All observed structures when looking at all CS for Pb(II) with CN>2.

For this part, a total of 56 Pb(II) CSs that possess CN of 3 or above were analysed and their geometries were determined. The assignment of the crystal structures was made in accordance with the VSEPR model,[16] the 14 possible geometries can be seen in *Figure 6*. The octahedral (Oh) geometry **8** was the most abundant structure followed by the pentagonal bipyramidal **12** and the tetragonal antiprism **13**, together with the trigonal bipyramidal geometry **7.X** (*Figure 7*; X refers to all options). There were also four structures that could not be resolved as their crystal structures were incomplete. These were mainly specially designed proteins consisting of three Cys CS, giving them a hypothetical trigonal pyramidal **3.X** or tetrahedral **4.X** geometry. They were marked as unknown as neither one nor the other could be confirmed.

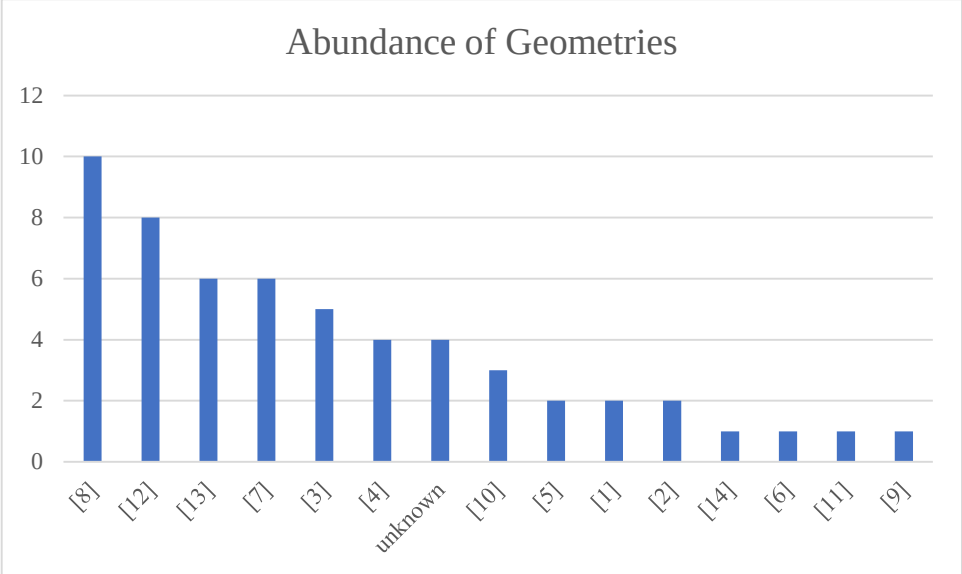


Figure 7. Distribution and abundance of the different geometries of the surveyed CS.

The three additional structures **3.2**, **4.2** and **7.2** are the most extreme cases of their respective **X.1** structures and together with other geometries (**1**, **2**, **5**, **6**, *etc.*) show the hemidirected geometry of Pb(II) CS, as they require at least one lone pair to achieve their conformation.[16] It must also be noted that many of the other observed geometries had slight to heavy distortions due to this effect.

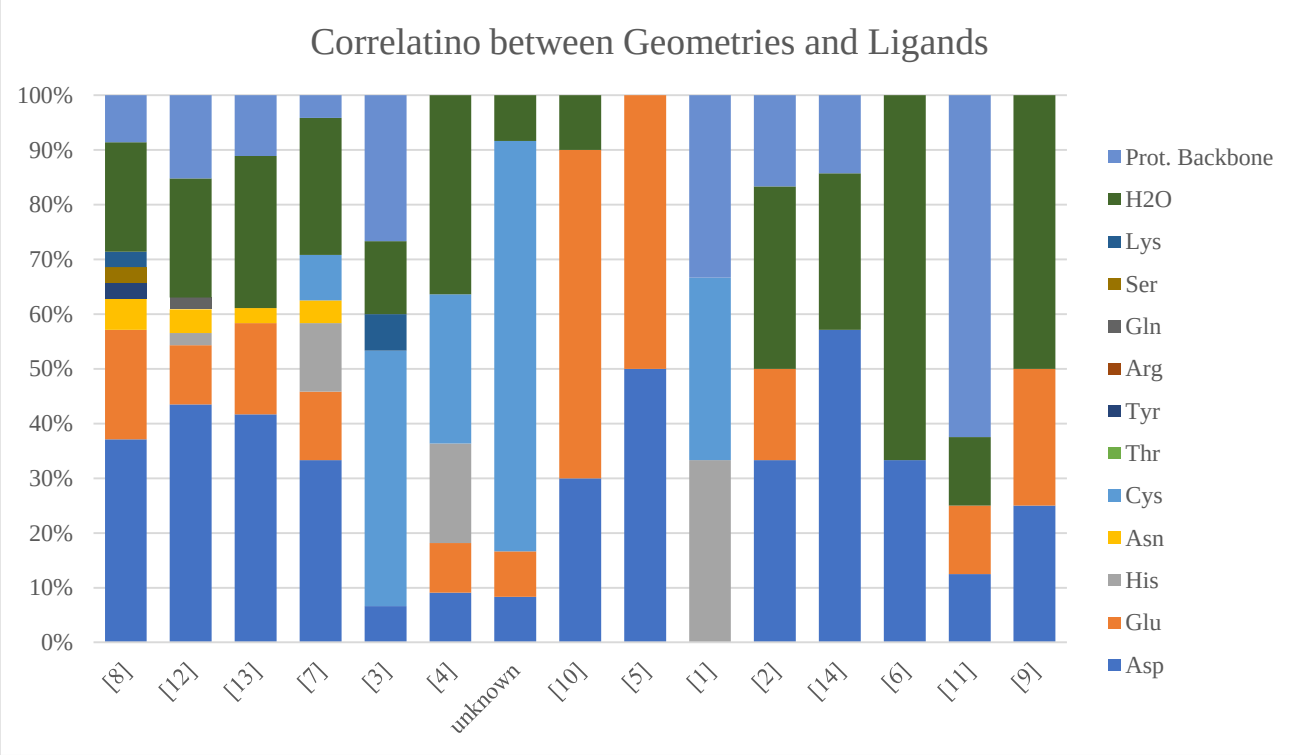


Figure 8. Correlation between the ligands and the observed geometry of the CS.

The figure above (*Figure 8*) shows the correlation between the different ligands and the geometries of the CS. It is interesting to note that H_2O can almost exclusively be found in low CN sites, but is almost non-existent in trigonal pyramidal complexes. The most abundant geometry (Oh) is mainly made up from Asp, Glu and the BB, but does not contain Cys. Which is not in agreement with the relative abundance of these AAs.

Conclusion

A total of 165 Pb atoms in various protein CS were studied. The data showed a high frequency of low CN that are originated from the protein itself and no explanation can be given for this phenomenon at this point in time. It is also evident that the most frequently observed AAs are Asp, Glu and Cys, which is in accordance with the HSAB concept, classifying Pb(II) and Pb(IV) as borderline *Lewis* acids.[6] The differences between the mono- and bidentate coordination of the two AAs might arise from steric and not electronic effects.[8] Three metals (Ca, Mg, Zn) were observed to be replaced by Pb, of which the latter shows the hemidirecting properties of Pb(II).[10] A phenomenon which is also observed in proteins that are suggested to be Pb specific. Where all CS consist of three Cys side chains and have a trigonal pyramidal geometry. The correlation between the ligands shows a high likeliness of Cys and His and also Cys and Cys being present in the same CS, in comparison with what would be expected from the abundance of these AAs. The reason for this behaviour is unknown. The geometries of the CS show a high abundance of hemdirected structures, which is in agreement with the expected behaviour of the electronic structure of Pb(II). It also suggests that the adaptability of the CS to the orbital structure of a metal ion plays and important role towards its selectivity.[14,15]

References

[1] M. J. Adams, B. R. Howard, *J. Utha Acad.* **2016**, *93*, 77–96.

[2] P. Kursula, V. Majava, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2007**, *63*, 653–656.

[3] C.-H. Lee, K. Saksela, U. A. Mirza, B. T. Chait, J. Kuriyan, *Cell* **1996**, *85*, 931–942.

[4] S. Katti, B. Her, A. K. Srivastava, A. B. Taylor, S. W. Lockless, T. I. Igumenova, *Metallomics* **2018**, *10*, 1211–1222.

[5] P. T. Beernink, B. W. Segelke, M. Z. Hadi, J. P. Erzberger, D. M. Wilson, B. Rupp, *J. Mol. Biol.* **2001**, *307*, 1023–1034.

[6] R. G. Pearson, *J. Chem. Educ.* **1968**, *45*, 581.

[7] P. Vollhardt, N. Schore, *Organic Chemistry, Structure and Function*, 6th Edition, New York **2011**.

[8] V. E. Anslyn, A. D. Dougherty, *Modern Physical Organic Chemistry*, 1st Edition, Sausalito, **2005**.

[9] T. Dudev, L. Doudeva, *J. Mol. Model.* **2017**, *23*, 45.

[10] K. A. Morales, M. Lasagna, A. V. Gribenko, Y. Yoon, G. D. Reinhart, J. C. Lee, W. Cho, P. Li, T. I. Igumenova, *J. Am. Chem. Soc.* **2011**, *133*, 10599–10611.

[11] J. J. G. Tesmer, *Science* **1999**, *285*, 756–760.

[12] W. Kutzelnigg, *Angew. Chemie Int. Ed. English* **1984**, *23*, 272–295.

[13] J. S. Magyar, T.-C. Weng, C. M. Stern, D. F. Dye, B. W. Rous, J. C. Payne, B. M. Bridgewater, A. Mijovilovich, G. Parkin, J. M. Zaleski, et al., *J. Am. Chem. Soc.* **2005**, *127*, 9495–9505.

[14] L. Ruckthong, M. L. Zastrow, J. A. Stuckey, V. L. Pecoraro, *J. Am. Chem. Soc.* **2016**, *138*, 11979–11988.

[15] S. Huang, X. Liu, D. Wang, W. Chen, Q. Hu, T. Wei, W. Zhou, J. Gan, H. Chen, *Inorg. Chem.* **2016**, *55*, 12516–12519.

[16] C. E. Housecroft, E. C. Constable, *Chemistry*, 4th Edition, Harlow, **2010**.