6-28-2010

Quantitative structure-property relationships for predicting group IIB metal binding by organic ligands

Aliyar Mousavi

Aliyar Mousavi
Candidate

Chemistry and Chemical Biology
Department

This dissertation is approved, and it is acceptable in quality and form for publication on microfilm:

Approved by the Dissertation Committee:

, Chairperson

QUANTITATIVE STRUCTURE-PROPERTY RELATIONSHIPS FOR PREDICTING GROUP IIB METAL BINDING BY ORGANIC LIGANDS

BY

ALIYAR MOUSAVI

B.S., Chemistry, The University of Maryland at College Park, 1998
M.S., Chemistry, Long Island University (Brooklyn Campus), 2004

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Chemistry

The University of New Mexico
Albuquerque, New Mexico

May, 2010

DEDICATION

To

My dear uncle, Mahmoud Mousavi, for his enduring and loving attention to my life

ACKNOWLEDGMENTS

I would like to extend many thanks to my advisor, Dr. Stephen E. Cabaniss, for sharing his knowledge with me throughout the entire period of my PhD research and for valuing and directing my research interests. Special appreciation is given to Dr. Abdulmehdi S. Ali for helping me in my PhD experimental and theoretical research. I also sincerely thank Dr. Richard A. Kemp, Dr. Laura J. Crossey, and Dr. Mark R. Ondrias for the time and help they have given me despite their own heavy loads of research, teaching, or administrative responsibilities. Finally, I would like to acknowledge the following people:

1- Dr. David L. Tierney for his experimental contributions to Appendix 2 of which he is a coauthor.

2- Kimberly Gugliotta, from the Geoanalytical Chemistry Laboratory at the Department of Earth and Planetary Sciences, University of New Mexico, for preparing the figure showing the biogeochemical cycling of mercury in Wisconsin Lakes.

3- Rose D. Chávez for all her proof-reading work put into the manuscript titled *Mercury in Natural Waters: Toxicity, Medical and Environmental Preventive Regulations, and Geochemistry – A Mini-Review* of which she is one of the coauthors, submitted to **Journal of Water and Health** for publication.

4- Mina Moeini, from the School of Architecture and Planning, University of New Mexico, for assistance in reading the data from the original figure showing experimental results for mercury complexation by natural organic matter.

5- Alexandra R. Priewisch, from the Department of Earth and Planetary Sciences, University of New Mexico, for translating a part of a reference from German to English.

6- Yafa Dayan, M.S., Certified School Psychologist at Jeffrey M. Rapport School for Career Development, Bronx, NY, for her explanations relating to the psychological problems caused by mercury.

7- Dr. Natasha Kolchevska, Professor of Russian and Chairperson of Foreign Languages Literatures at the University of New Mexico, for her helpful scholarly comments about Russian history.

8- Michael R. Davenport, Ken A. Seal, Monique R. Cordova, and Lidong Wang for providing or assisting me with various computer software at the University of New Mexico Department of Chemistry and Chemical Biology computer lab.

QUANTITATIVE STRUCTURE-PROPERTY RELATIONSHIPS FOR PREDICTING GROUP IIB METAL BINDING BY ORGANIC LIGANDS

BY

ALIYAR MOUSAVI

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Chemistry

The University of New Mexico
Albuquerque, New Mexico

May, 2010

QUANTITATIVE STRUCTURE-PROPERTY RELATIONSHIPS FOR PREDICTING GROUP IIB METAL BINDING BY ORGANIC LIGANDS

BY

ALIYAR MOUSAVI

B.S., Chemistry, The University of Maryland at College Park, 1998
M.S., Chemistry, Long Island University (Brooklyn Campus), 2004
Ph.D., Chemistry, The University of New Mexico, 2010

ABSTRACT

Mercury (Hg), cadmium (Cd), and zinc (Zn) in the environment are all of toxicological and environmental concern, and the pollution of natural waters by any of these three elements is most serious. Mercury is the most environmentally concerning of the three because of the neurotoxin species monomethylmercury produced in aquatic systems through the methylation of Hg^{2+} by aquatic microorganisms. An important chemical process in natural waters that limits the availability of mercury for methylation is the binding of Hg(II) by natural organic matter (NOM). These associations are exceptionally strong, and as NOM is ubiquitous in aquatic environments, estimating equilibrium constants for Hg(II) binding to NOM in natural waters is important. Cadmium is moderately toxic to all organisms, and skeletal damage caused by exposure to cadmium-contaminated water has been reported. Also high concentrations of zinc that are toxic or even lethal to organisms have been observed in natural waters. As the free ion forms of cadmium and zinc in natural waters are thought to be most toxic, Cd(II) and Zn(II) complexation by NOM and estimating the complexation equilibrium constants are, similarly to Hg(II), of interest.

With experimental determination of M(II)-NOM (M = Hg, Cd, Zn) binding constants being costly and time consuming, it is desirable to estimate those constants without the benefit of additional experimental data. This work uses QSPRs (Quantitative Structure-Property Relationships) to predict binding constants from hypothetical structures of NOM molecules. For the first time, to our knowledge, a QSPR for predicting Hg(II) complexation by organic ligands has been developed. Also two QSPRs for predicting Cd(II) and Zn(II) complexation by organic ligands, that had been developed earlier, have been improved to be capable of predicting the binding of Cd(II) and Zn(II) to thiol-containing molecules.

Most of the compounds used in the calibration data sets of the three QSPRs contained some or all of carboxylate, amine, and thiol ligand groups. The Hg(II), Cd(II), and Zn(II) QSPRs respectively have standard error of prediction (S_{pred}) values of 1.60, 0.935, and 0.984 log units and describe 96.5%, 93.1%, and 93.4% of the variability in data. The most noteworthy observation in the developed QSPRs was the exceptionally high affinity Hg(II) had for thiols. Although thiols form a very small fraction of NOM, this binding is considerably important because of its strength. This work also presents certain potential applications of the developed QSPRs in predicting M(II)-NOM binding as well as predicting M(II) binding to organic molecules which would be synthesized for M(II) remediation and chelation therapy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Mercury in Natural Waters: Toxicity, Environmental Preventive Regulations, and Geochemistry

This chapter, with some changes and additions, forms a major part of a manuscript titled *Mercury in Natural Waters: Toxicity, Medical and Environmental Preventive Regulations, and Geochemistry – A Mini-Review*, prepared by Aliyar Mousavi, Rose D. Chávez, Abdul-Mehdi S. Ali, and Stephen E. Cabaniss, and submitted to **Journal of Water and Health** for publication.

1.1 Introduction

The symbol of mercury, Hg, is derived from the Latin word *hydrargyrum* meaning "liquid silver" (Greenwood & Earnshaw 1994). The symbol refers to the fact that mercury in elemental form exists as a liquid at room temperature (Holleman – Wiberg 2001). The compounds of mercury, with the exception of derivatives of Hg_2^{2+} ion, involve Hg(II) (Greenwood & Earnshaw 1994). However, in addition to existing in elemental form and in inorganic compounds, mercury also has an enormous number of known organometallic compounds (Greenwood & Earnshaw 1994). From an environmental point of view monomethylmercury and dimethylmercury species are of special importance.

With all the forms of mercury being variously toxic to aquatic biota and humans, mercury in natural waters is an issue. Understanding this issue well requires familiarity with details about the health hazards of mercury in food and water, aquatic ecosystem pollution by mercury, and the environmental processes associated with the toxicity of mercury in natural waters. This chapter first addresses the toxicity of mercury to aquatic life and humans and

the regulations concerning mercury in aquatic environments and then briefly reviews the geochemistry of mercury in natural waters.

1.2 Mercury Toxicity

Both inorganic and organic mercury compounds are toxic to aquatic plants (Moore & Ramamoorthy 1984; Boening 2000). Aquatic plants are affected by mercury in water at 1 mg/L of inorganic mercury and at much lower concentrations of organic mercury (Boening 2000). Mercury may intoxicate the pre-hatch (egg) stage of fish (Moore & Ramamoorthy 1984; Hammerschmidt & Sandheinrich 2005) and can cause poisoning leading to decrease in the ability of fish to survive (Moore & Ramamoorthy 1984; Vieira *et al.* 2009). Mercury compounds also can cause reproductive abnormality in fish (Moore & Ramamoorthy 1984; Boening 2000). Mercury in fish is of environmental concern especially because fish in the food supply is where most of the mercury present in humans originates (Ravichandran 2004). Consumption of contaminated fish products, fish, and wildlife that are at the top of aquatic food chains is almost entirely what causes human exposure to monomethylmercury (U.S. Geological Survey 2000; Fitzgerald 2003), a chemical species with which the principal mercury-related human health concern is associated today (Fitzgerald 2003).

Monomethylmercury is a highly neurotoxic species, and its poisoning is known as "Minamata disease" and is named after the major industrially related mass poisonings, severe debilitation, and many deaths which occurred in Minamata and Niigata, Japan between 1950 and 1975 (Fitzgerald 2003). In 1952 only, the poisonings led to the death of 52 people (Greenwood & Earnshaw 1994). Monomethylmercury not only damages the nervous system, but also affects the immune system and alters genetic and enzyme systems (U.S. Geological Survey 2000).

Although the only form of mercury that accumulates appreciably in fish is monomethylmercury (U.S. Geological Survey 2000; Ravichandran 2004), maternal exposure to methylmercury-contaminated fish is also a possible cause of breast milk being a potential source of exposure to inorganic mercury (Risher & De Rosa 2007). Autoimmune effects have been reported in humans following prolonged exposure to inorganic mercury, and long-term exposure to both high and low amounts of inorganic mercury can cause renal damage (Risher & De Rosa 2007). Still organomercury compounds are more dangerous than inorganic mercury compounds (Greenwood & Earnshaw 1994). Because of their greater permeability of biomembranes than Hg(II) they are more readily absorbed in the gastrointestinal tract and, acting by binding to the –SH groups in proteins, have a more immediate and permanent effect on the brain and central nervous system (Greenwood & Earnshaw 1994).

It is, however, important to mention that qualitatively similar specific neurotoxic symptoms have been reported, irrespective of the mercury compound to which one has been exposed (Risher & De Rosa 2007). Some of the symptoms are memory loss, neuromuscular changes (weakness, muscle atrophy, and muscle twitching), insomnia, and emotional lability which may be behaviorally manifested as irritability (Risher & De Rosa 2007). Here a scientific observation related to history might be worth noting as a matter of fact open to scholarly interpretation. The autopsy on Ivan the Terrible has revealed spinal disease (Perrie 2006). Whether certain pain killers were the source of exposure or something else was, the autopsy has also revealed large amounts of mercury in Ivan's body (Perrie 2006).

The Minamata poisonings in Japan resulted from consumption of locally caught fish, that formed the staple diet of the small fishing community (Greenwood & Earnshaw 1994),

and seafood that had been contaminated principally by monomethylmercury discharged with wastewater from factories synthesizing it as a by-product (Fitzgerald 2003) as Hg(II) salts were used inefficiently (Greenwood & Earnshaw 1994). Still research has shown that mercury can be a threat to the health of people and wildlife in many environments that are not even obviously polluted (U.S. Geological Survey 2000). It is important noting that because animals accumulate monomethylmercury faster than they eliminate it, small environmental concentrations of it can readily accumulate to potentially harmful concentrations in fish, fish-eating wildlife, and people (U.S. Geological Survey 2000). As the only form of mercury that accumulates appreciably in fish is monomethylmercury (U.S. Geological Survey 2000; Ravichandran 2004), the fact that fish accumulate monomethylmercury faster than they eliminate it is a physiological basis for fish consumption advisories for mercury. However, consumption of contaminated fish products, fish, and wildlife that are at the top of aquatic food chains has not always been the only significant source of human exposure to methylmercury. Both methylmercury and ethylmercury have been used previously as fungicides on seeds used for growing crops, but such use is discouraged worldwide by the World Health Organization, subject to severe regulatory restriction worldwide, and currently cancelled in the United States (U.S. EPA, TEACH Chemical Summary).

1.3 Environmental Preventive Regulations

With fish consumption dominating the pathway for exposure to methylmercury for most human populations, many governments around the world provide recommendations or legal limits for the maximum allowable amount of mercury and/or methylmercury in fish to be sold on the market (WHO-UNEP 2008) (See Table 1.1).

Table 1.1 Regulatory limits of methylmercury in fish (as mg methylmercury/kg in fish) set by the Food and Agriculture Organization (FAO)/World Health Organization (WHO) Codex Alimentarius Commission, U.S. Food and Drug Administration (FDA), and Japan. The information is taken from WHO-UNEP 2008.

Fish Species	*Guideline Level Set by the FAO/WHO Codex Alimentarius Commission*	*Action Level Set by the U.S. FDA*	*Allowed Level Set by Japan*
Non-Predatory Fish	0.5	1	0.3
Predatory Fish	1	1	0.3

Although no legally binding international treaty on mercury pollution is in effect as of today, according to a decision made on February 20, 2009 during a meeting of the governing council of the United Nations Environment Programme (UNEP), which was held in Nairobi, more than 140 countries have agreed to begin negotiations on a global treaty to control mercury (Hogue 2009). The U.S. Environmental Protection Agency (EPA) has determined the safe level of mercury in drinking water to be 2 μg/L, referring to μg/L as parts per billion (ppb), and has also set that to be the enforceable standard because EPA believes, given present technology and resources, that is the lowest level to which water systems can reasonably be required to remove mercury should it occur in drinking water (U.S. EPA 1995). As for fresh waters and salt waters, EPA has set required mercury criteria for the U.S. states too, unless a state has its own criteria which meet the EPA criteria. The EPA mercury criterion (maximum contaminant level) for fresh waters is 2.1 μg/L, and that for salt waters is 1.8 μg/L (ECFR). Table 1.2 lists the regulatory limits of mercury in water set by EPA.

Table 1.2 Regulatory limits of mercury in water set by the U.S. Environmental Protection Agency (EPA).

Water Type	*Allowed Maximum Concentration in μg/L*
Drinking Water	2 (U.S. EPA, 40 CFR)
Fresh Water	2.1 (ECFR)
Salt Water	1.8 (ECFR)

1.4 Geochemical Occurrence

Mercury occurs naturally in the earth's crust. Its only important ore and natural source is cinnabar, HgS, found along lines of previous volcanic activity. The most famous and extensive deposits of cinnabar are at Almaden in Spain and have been worked since Roman times. Some of other deposits of cinnabar are situated in Algeria, Mexico, and Italy. While the deposits at Almaden contain up to 6-7% Hg, other deposits usually contain < 1% Hg (Greenwood & Earnshaw 1994).

Mercury is released into the environment by both natural and anthropogenic processes. Natural processes, such as volcanic activity, release mercury primarily in the form of Hg^0 (g) into the air (Risher & De Rosa 2007). Anthropogenic inputs of mercury to the environment are numerous and widespread. The most problematic are atmospheric, principally coal and municipal waste burning (Fitzgerald 2003). Mining of gold is another important anthropogenic source (U.S. Geological Survey 2000; Risher & De Rosa 2007), especially as mercury was used in the Mediterranean world for extracting metals by amalgamation as early as 500 BC (Greenwood & Earnshaw 1994). The use of mercury in the chlor-alkali industry, where mercury is used as an electrode in the electrochemical

process of manufacturing chlorine, is a major source (Järup 2003). Cinnabar was widely used in the ancient world as a pigment (vermilion) (Greenwood & Earnshaw 1994), and mercury is used today in paints and tattoo inks. Some other examples of anthropogenic sources of nonoccupational exposure to mercury are dental amalgams (still used for filling teeth in many countries), barometers, instruments for measuring blood pressure, gas regulators, fluorescent bulbs, wall light switches, camera batteries, thermostats, and thermometers (Järup 2003; Risher & De Rosa 2007).

Greater than 95% of the mercury found in the atmosphere is gas-phase elemental mercury, while the mercury in precipitation is predominantly ionic (Fitzgerald 2003). Elemental mercury in the environment can combine with a number of elements (including chlorine, sulfur, and phosphorus) to form compounds of inorganic mercury (Risher & De Rosa 2007). The redox couple of $Hg^{(0)}$ with the stable Hg^{2+} ($Hg^{(0)}/Hg^{2+}$; $E^0 = 0.85$ V) provides the potential for dynamic oxidation and reduction cycling in the environment (Fitzgerald 2003). Because of the stability of Hg^{2+} in water Hg^0 (g) is not a major species in aquatic systems, and Hg_2^{2+} is a minor species at concentrations less than 450 mg/L of total mercury, a level that is unlikely in natural waters (Panel on Mercury 1978). The approximate concentration of mercury in freshwater is 0.1-2 ng/L, and that in ocean water is 0.2-1 ng/L. The mercury species of highest concentrations in natural waters in general is Hg^{2+} (Fitzgerald 2003). Yet Hg^{2+} in aquatic systems can be readily methylated predominantly through the action of aquatic microorganisms (Fitzgerald 2003; Risher & De Rosa 2007).

Since 2000 a whole-ecosystem experiment (termed METAALICUS, Mercury Experiment to Assess Atmospheric Loading in Canada and the United States) has been designed and in progress at the Experimental Lakes Area in northwestern Ontario, Canada, to

study the relationship between atmospheric deposition of mercury and mercury accumulation in fish (Sandilands *et al.* 2005). Changes in inorganic mercury loading (increase or decrease) will yield a response in fish monomethylmercury (Munthe *et al.* 2007), and significant and even linear relationships have been observed between inorganic mercury loading rates and spike monomethylmercury concentrations in zooplankton, benthic invertebrates, and fish (Orihel *et al.* 2007). Understanding the biogeochemistry of Hg^{2+} methylation process is therefore of great environmental importance.

1.5 Mercury Biogeochemistry

Sulfate-reducing bacteria have been identified as the primary organisms responsible for the production of monomethylmercury in aquatic environments (Ekstrom & Morel 2008). The methyl group in monomethylmercury has been transferred to Hg^{2+}, and this transferring is done by Co-CH_3 groups (Greenwood & Earnshaw 1994). One chemical species is known to be the Co-CH_3 source, and that is metabolically produced methylcobalamin (Choi *et al.* 1994), a Co(III) compound produced in biological systems through the methylation of the super-reduced form of vitamin B_{12} (Huheey *et al.* 1993). Following methylation, monomethylmercury-containing bacteria may be consumed by the next higher level in the food chain or may excrete the monomethylmercury to the water (U.S. Geological Survey 2000) where it can adsorb to plankton and so move to the next level in the food chain (Greenwood & Earnshaw 1994; U.S. Geological Survey 2000). Monomethylmercury is the only observed methylmercury species in common freshwaters, although it is important to note that in ocean waters dimethylmercury and (to a lesser extent) monomethylmercury are common constituents of the dissolved mercury pool (Fitzgerald 2003). However, at a low

pH, dimethylmercury is unstable and rapidly decomposes to monomethylmercury (Trevors 1986; Black *et al.* 2009).

What environmental factors affect the methylation of Hg^{2+} in aquatic ecosystems? Methylation ability in pure culture has been demonstrated by only a few sulfate-reducing bacteria and a few closely related iron-reducing bacteria. Those sulfate-reducing bacteria are stimulated by sulfate, and the activity of microorganisms is often stimulated by organic matter in sediments and soils. As such, both sulfate and organic matter in sediments and soils are effective biogeochemical factors on Hg^{2+} methylation in aquatic ecosystems. Also uptake of inorganic mercury by microorganisms that use facilitated transport for inorganic mercury uptake is enhanced with decreasing pH. It is noteworthy that many studies have linked lake acidity to increased monomethylmercury bioaccumulation (Munthe *et al.* 2007).

Reviewing the aquatic mercury cycle is in no way complete if reverse processes such as demethylation and reduction are not addressed. In fact, aside from grossly polluted environments, mercury is normally a problem only where the rate of natural formation of monomethylmercury from inorganic mercury is greater than the reverse reaction (U.S. Geological Survey 2000). Monomethylmercury demethylation process can result in the formation of Hg^{2+} (Winfrey & Rudd 1990; U.S. Geological Survey 2000) and is done by sunlight that can also break down monomethylmercury to Hg^{0} (U.S. Geological Survey 2000) or is mediated by methylmercury-resistant bacteria in which case the final product is elemental mercury (Winfrey & Rudd 1990). The reduction of Hg^{2+} to Hg^{0} is a notable reaction in the biogeochemical cycling of mercury in natural waters (Winfrey & Rudd 1990; Fitzgerald 2003). The elemental mercury formed in aquatic environments can be volatilized to the atmosphere (Winfrey & Rudd 1990; U.S. Geological Survey 2000; Fitzgerald 2003).

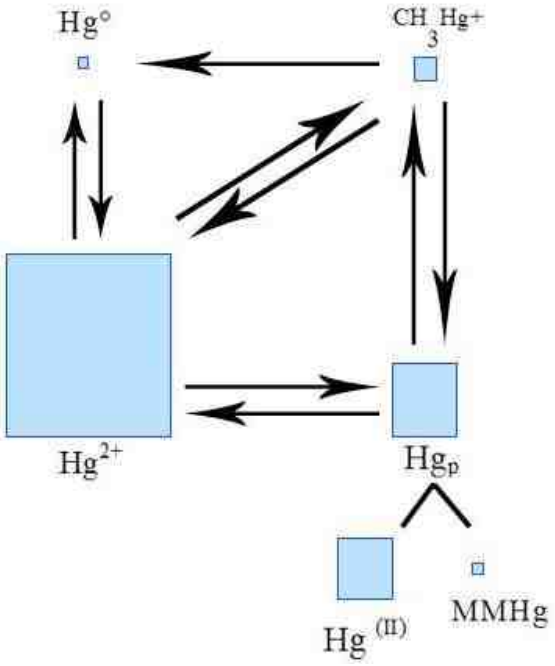


Figure 1.1 The biogeochemical cycling of mercury in Wisconsin Lakes, USA. The size of each box represents the percentage of its corresponding species in the lakes. The mercury pool consists of 3.25% Hg^0, 8.3% CH_3Hg^+ (monomethylmercury), 25% Hg_p (particulated mercury), consisting of 20.8% $Hg^{(II)}$ and 4.2% MMHg (monomethylmercury), and 63.4% Hg^{2+}. The species percentages are based on the data reported by Fitzgerald 2003.

Figure 1.1 is a pictorial representation of the biogeochemical cycling of mercury in freshwater lakes. Other than methylation, demethylation, and reduction, there are two important sets of reactions in the biogeochemical cycling of mercury in fresh waters. In the presence of hydrogen sulfide, Hg^{2+} precipitates as HgS. This precipitation has generally been assumed to inhibit the availability of mercury for methylation and is believed to do so (Winfrey & Rudd 1990; Fitzgerald 2003). The second is the set of reactions in which Hg^{2+} is particulated or adsorbs to natural organic matter (NOM) (Winfrey & Rudd 1990; Fitzgerald 2003), a heterogeneous mixture of organic compounds of ill-defined chemical structure (Ravichandran 2004). Mercury forms exceptionally strong associations with NOM (Fitzgerald 2003), and the binding of Hg(II) to NOM limits the availability of mercury for

methylation (Ravichandran 2004). Therefore, as NOM is ubiquitous in aquatic environments (Ravichandran 2004), this binding deserves increasing attention. The purpose of the next chapter is to address the relationships between mercury toxicity and Hg(II)-NOM bindings and to discuss a potentially most useful methodology to predict the bindings.

1.6 References

Black, F. J., Conaway, C. H. & Flegal, A. R. 2009 Stability of dimethyl mercury in seawater and its conversion to monomethyl mercury. *Environ. Sci. Technol.* **43**(11), 4056-4062.

Boening, D. W. 2000 Ecological effects, transport, and fate of mercury: a general review. *Chemosphere* **40**(12), 1335-1351.

Choi, S. C., Chase, T. Jr. & Bartha, R. 1994 Enzymatic catalysis of mercury methylation by *Desulfovibrio desulfuricans* LS. *Appl. Environ. Microbiol.* **60**(4), 1342-1346.

Ekstrom, E. B. & Morel, F. M. 2008 Cobalt limitation of growth and mercury methylation in sulfate-reducing bacteria. *Environ. Sci. Technol.* **42**(1), 93-99.

Electronic Code of Federal Regulations (ECFR). Title 40: Protection of Environment, Part 131-Water Quality Standards, Subpart D-Federally Promulgated Water Quality Standards.

Fitzgerald, W. F. Geochemistry of mercury in the environment. In *Environmental Geochemistry* (ed. B.S. Lollar) 2003 Vol. 9 *Treatise on Geochemistry* (eds. H.D. Holland and K.K. Turekian). Elsevier-Pergamon, Oxford, pp 107-148.

From Early Rus' to 1689 (ed. M. Perrie) 2006 Vol. 1 *The Cambridge History of Russia*. Cambridge University Press, Cambridge, p. 252.

Greenwood, N. N. & Earnshaw, A. 1994 *Chemistry of the Elements*. Pergamon Press, Tarrytown, NY, pp 1395-1422.

Hammerschmidt, C. R. & Sandheinrich, M. B. 2005 Maternal diet during oogenesis is the major source of methylmercury in fish embryos. *Environ. Sci. Technol.* **39**(10), 3580-3584.

Hogue, C. 2009 Mercury Countries agree to start talks to control the metal globally. *Chemical & Engineering News* Mar 2, 2009, 13.

Holleman, A. F. – Wiberg, E. 2001 *Inorganic Chemistry.* Academic Press, San Diego, CA, pp 1303-1316.

Huheey, J. E., Keiter, E. A. & Keiter, R. L. 1993 *Inorganic Chemistry: Principles of Structure and Reactivity*. HarperCollins College Publishers, New York, NY, pp 889-964.

Järup, L. 2003 Hazards of heavy metal contamination. *Br. Med. Bull.* **68**(1), 167-182.

Moore, J. W. & Ramamoorthy, S. 1984 *Heavy Metals in Natural Waters*. Springer-Verlag New York Inc., New York, NY, pp 125-160.

Munthe, J., Bodaly, R. A. (Drew), Branfireun, B. A., Driscoll, C. T., Gilmour, C. C., Harris, R., Horvat, M., Lucotte, M. & Malm, O. 2007 Recovery of mercury-contaminated fisheries. *Ambio.* **36**(1), 33-44.

Orihel, D. M., Paterson, M. J., Blanchfield, P. J., Bodaly, R. A. (Drew) & Hintelmann, H. 2007 Experimental evidence of a linear relationship between inorganic mercury loading and methylmercury accumulation by aquatic biota.
Environ. Sci. Technol. **41**(14), 4952-4958.

Panel on Mercury of the Coordinating Committee for Scientific and Technical Assessments of Environmental Pollutants 1978 *An Assessment of Mercury in the Environment*. National Academy of Sciences, Washington, DC, pp 30-39.

Ravichandran, M. 2004 Interactions between mercury and dissolved organic matter-a review. *Chemosphere* **55**(3), 319-331.

Risher, J. F. & De Rosa, C. T. 2007 Inorganic: the other mercury. *J. Environ. Health* **70**(4), 9-16.

Sandilands, K. A., Rudd, J. W. M., Kelly, C. A., Hintelmann, H., Gilmour, C. C. & Tate, M. T. 2005 Application of enriched stable mercury isotopes to the Lake 658 watershed for the METAALICUS project, at the Experimental Lakes Area, northwestern Ontario, Canada. *Can. Tech. Rep. Fish. Aquat. Sci.* 2597: viii + 48 p.

Trevors, J. T. 1986 Mercury methylation by bacteria. *J. Basic Microbiol.* **26**(8), 499-504.

U.S. Environmental Protection Agency (EPA). 1995 *National Primary Drinking Water Regulations.* 40 CFR, Parts 141-143.

U.S. Environmental Protection Agency (EPA). *Organic Mercury.* Toxicity and Exposure Assessment for Children's Health (TEACH) Chemical Summary.

U.S. Geological Survey. 2000 *Mercury in the Environment*. U.S. Dept. of the Interior, U.S. Geological Survey, Reston, VA.

Vieira, L. R., Gravato, C., Soares, A. M., Morgado, F. & Guilhermino, L. 2009 Acute effects of copper and mercury on the estuarine fish Pomatoschistus microps: linking biomarkers to behaviour. *Chemosphere* **76**(10), 1416-1427.

Winfrey, M. R. & Rudd, J. W. M. 1990 Environmental factors affecting the formation of methylmercury in low pH lakes. *Environ. Toxicol. Chem.* **9**(7), 853-869.

World Health Organization (WHO), United Nations Environment Programme (UNEP) 2008 *Guidance for Identifying Populations at Risk from Mercury Exposure*. UNEP DTIE Chemicals Branch and WHO Department of Food Safety, Zoonoses and Foodborne Diseases, Geneva.

CHAPTER 2

A Quantitative Structure-Property Relationship for Predicting Hg(II) Binding by Organic Ligands

This chapter is the main draft of a manuscript titled *A Quantitative Structure-Property Relationship for Predicting Hg(II) Binding by Organic Ligands* in preparation by Aliyar Mousavi and Stephen E. Cabaniss for resubmission to **Environmental Science and Technology** for publication.

2.1 Introduction

Natural Organic Matter (NOM) plays an important role in the environmental fate of mercury (Khwaja *et al.* 2006), and binding of Hg(II) to NOM strongly affects bioavailability of Hg(II) in aquatic ecosystems (Haitzer *et al.* 2002) and limits Hg(II) availability to methylating bacteria (Ravichandran 2004). In that regard, knowing binding constants for Hg(II)-NOM complexes is of special importance.

In order to estimate equilibrium constants for Hg(II) binding to NOM in natural waters, several experimental methods are available. The most commonly employed approaches use competitive ligands, for example Br^- (Skylberg *et al.* 2000), ethylenediaminetretraacetic acid (EDTA) (Haitzer *et al.* 2002; Haitzer *et al.* 2003), diethyldithiocarbamate (Hsu & Sedlack 2003), and DL-penicillamine (Khwaja *et al.* 2006). However, experimental determination of the equilibrium constant of Hg(II) binding by NOM is costly and time consuming. It is, therefore, desirable to estimate those equilibrium constants without the benefit of additional experimental data.

The most common models developed to describe metal-binding equilibria with NOM are the competitive Gaussian distribution model (Perdue & Parrish 1987), the Windermere

humic aqueous model (WHAM) (Tipping 1998), and the non-ideal competitive adsorption-Donnan (NICA-Donnan) model (Milne *et al.* 2003). In a 2003 paper, Milne and co-workers (Milne *et al.* 2003) provided recommended generic NICA-Donnan model parameters for 23 metal ions, including Hg(II), for both fulvic and humic acids. However, they did not consider the published collection of data for Hg(II) extensive enough to describe the full variability of binding properly and, therefore, used the variation in hydrolysis behavior of Hg(II) as an indication of likely relative capabilities for binding to humic substances (Milne *et al.* 2003). Hence, the degree of confidence in the parameters provided for Hg(II) is low (C. Milne, pers. comm.).

An alternative approach is to use QSPRs (Quantitative Structure-Property Relationships) to predict binding constants from hypothetical structures of NOM molecules (Cabaniss 2009). No QSPR useful for predicting Hg(II) binding equilibrium constants has, to our knowledge, been developed. NOM is remarkably heterogeneous in structure and exact structural information is unavailable (Schmitt-Kopplin 2003). Consequently, any QSPR which is to be useful for predicting Hg(II) binding by NOM should require minimal information- preferably constitutional descriptors such as functional group and elemental composition.

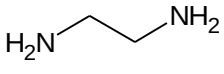
This chapter develops a QSPR for predicting equilibrium constants of Hg(II) binding by small organic molecules similar in ligand group composition to NOM. The equilibrium constants are related to elemental and functional group composition of the molecules without relying on stereochemistry or even overall connectivity. The QSPR is then applied to hypothetical structures of NOM and the predicted constants compared to literature data. In

future work, the QSPR will be interfaced with a dynamic, agent-based model of NOM for predicting Hg(II) complexation in different environments (Cabaniss 2009).

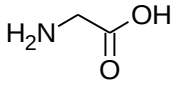
2.2 The Calibration Data Set

2.2.1 Selecting the Ligand Molecules and the Thermodynamic Equilibrium Constants

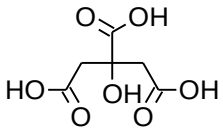
Ligand molecules for the calibration data set used in this work were selected so that (a) molecules contain one or more of the functional groups: alcohol, carboxylic acid, amine, thioether, and thiol, (b) a 1:1 binding constant for Hg(II) is available, and (c) no molecule has more than one thiol group.



(i) ethylenediamine



(ii) glycine



(iii) citric acid

(iv) D-penicillamine

Figure 2.1 Example Ligands

The calibration data set contains 44 distinct ligand molecules with 1:1 binding constants (Figure 1 shows example molecules; Supplementary Table 2.1 gives the complete set). The equilibrium constant (K_{HgL}) for each 1:1 complex was selected at 25 ^{o}C and ionic strength 0.1, unless no constant at such conditions was available in which case a constant determined at similar conditions was selected. For most cases, the thermodynamic

equilibrium constants were obtained from NIST Standard Reference Database 46 Version 8.0 (NIST 2004).

However, throughout the process of collecting stability constants for the calibration model an inconsistency was apparent for the complexes Hg(II) forms with thiol-containing molecules (Khwaja *et al.* 2006). The equilibrium constants (K_{HgL} values) reported in the NIST database for Hg(II) binding to thiol-containing molecules are much (~20 log units) smaller than those reported by some sources listed in the database yet excluded by it (Basinger *et al.* 1981; Casas & Jones 1980). These differences appear to be systematically related to the analytical methods used for determination of Hg(II) complex stability constants (see Table 2.1).

Table 2.1 Inconsistency in Thiol Equilibrium Binding (K_{HgL})

Molecule	**log K_{HgL} (pH Titrations)**	**log K_{HgL} (Hg Electrode Titrations)**
Penicillamine	18.86 (Strand & Lund 1983)	38.3 (Casas & Jones 1980)
L-cysteine	14.21 (Lenz & Martell 1964)	37.8 (Basinger *et al.* 1981)

The method used in studies accepted by the NIST database is pH-potentiometric titration, an indirect titration of Hg-L systems monitored by pH electrode. In this method the stability constant is calculated from the equilibrium pH of a solution of the thiol-containing molecule after protons have been replaced by Hg(II) in a titration process. An example of this method is the work of Strand and Lund (Strand & Lund 1983), who reported log K_{HgL} = 18.86 as the formation constant for Hg(II) 1:1 complex with D-penicillamine. In contrast,

Basinger et al. (Basinger *et al.* 1981) reported log K_{HgL} = 37.8 as the formation constant for the same complex, using the method of Hg electrode titrations.

The indirect pH titration method is not suitable for measuring very strong Hg(II) complex stability constants, because the measured pH is relatively insensitive to differences in log K_{HgL} above ~15. Using the acid dissociation constants Strand and Lund (Strand & Lund 1983) reported for D-penicillamine and their total concentrations, we simulated pH titrations with various values of the 1:1 equilibrium constant (see Figure 2) using the chemical equilibrium calculator Titrator (Cabaniss 1987). For example, in the absence of added base the Hg(II)-D-penicillamine mixture has pH = 2.700 if log K_{HgL} = 16, and pH = 2.699 if log $K_{HgL} \geq 17$. Even at the titration equivalence point (V = 2.00 mL), where the pH difference is the highest, the difference is <0.002 pH units. This methodological problem applies to all the 'accepted' Hg(II) equilibrium constants for thiol-containing ligands we found in the NIST database (NIST 2004).

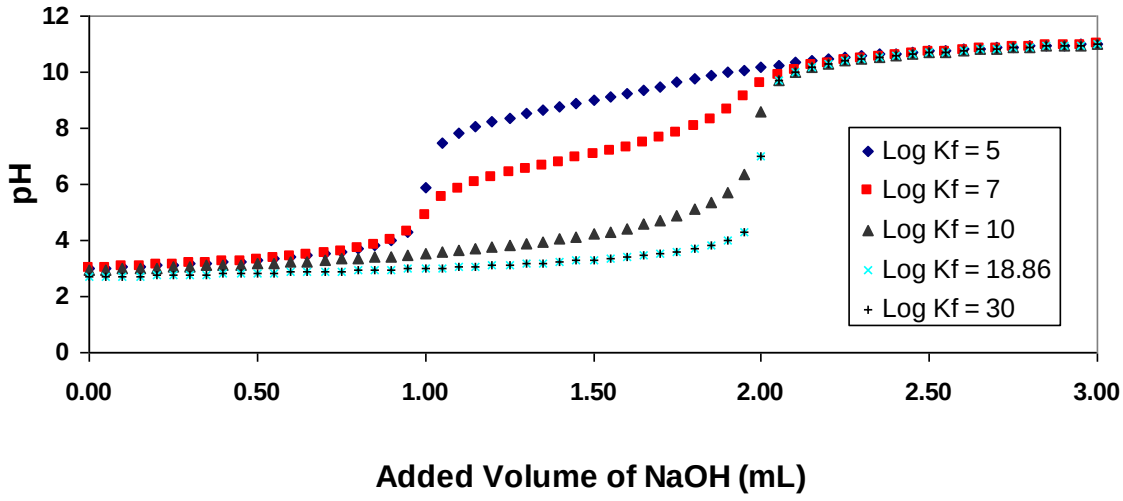


Figure 2.2 Simulated pH titrations of Hg(II)-Penicillamine with NaOH at differing values of Hg(II)-PA binding strength.

The Hg electrode results were consistent (within 2 log units) with those of Koszegi-Szalai and Paal (Koszegi-Szalai & Paul 1999), who used a method with a completely different chemical basis, equilibrium phase partitioning. For the thiol-containing ligands, therefore, equilibrium constants based on Hg electrode titrations were used. The thermodynamic equilibrium constants for Hg(II) complexes of D-penicillamine, L-cysteine, N-acetyl-cysteine, and mercaptoacetic acid were taken from Basinger et al. (Basinger *et al.* 1981), and that for N-acetyl-D,L-penicillamine from Casas and Jones (Casas & Jones 1980).

2.2.2 Calculating and Using the Conditional Equilibrium Constants

The equation for the formation of each complex and the corresponding equilibrium constant K_{HgL} is:

$$Hg^{2+} + L^{n-} \rightleftarrows HgL^{2-n} \qquad K_{HgL} = [HgL^{2-n}] / [Hg^{2+}][L^{n-}] \qquad (1)$$

To facilitate integration with models of NOM and to make predictions at neutral pH, it is useful to know the pH-adjusted conditional constant. For each organic molecule in the data set, therefore, K_{HgL} is used to calculate a corresponding conditional equilibrium constant, K_{HgL}'.

$$K_{HgL}' = [HgL^{2-n}] / [Hg^{2+}][L'] \qquad (2)$$

where $[L']$ is the sum of the concentrations of all ligand species which are not complexed by Hg(II). Combining equations (1) and (2), it may be written

$$K_{HgL}' = K_{HgL} \cdot [L^{n-}]/[L'] = K_{HgL}\,\alpha_n \qquad (3)$$

where α_n is the fraction of the uncomplexed ligand which exists in the $-n$ charge state (Harris 2007). The use of the word "conditional" in naming K_{HgL}' reflects the fact that it varies with pH, ionic strength, and temperature. In this study, pH 7.0 is chosen, but as α_n can be

calculated for other pH values, so can K_{HgL}'. This enables us to use the model for any environmental pH of interest provided acid pKa's are known.

2.3 Descriptor Selection and Calibration of QSPR

The QSPR presented here follows the general multiple linear equation

$$\text{Log } K_{HgL}' = a_1x_1 + a_2x_2 + \ldots + a_nx_n \qquad (4)$$

where x_i represents the ith descriptor variables and a_i parameters represent the coefficients of those variables. Since there was no good theoretical justification for a nonzero intercept, the intercept was set to zero.

In order to choose the QSPR descriptor variables, multiple linear regression was performed on an Excel spreadsheet using several structure-property-related variables. The candidate descriptors were constitutional variables, including molecular weight, numbers of sulfur atoms, hydrogen atoms, carbon atoms, oxygen atoms, nitrogen atoms, thiol groups, hydroxyl groups, carboxylic acid groups, primary amine groups, secondary amine groups, tertiary amine groups, thioether sulfur atoms, and many of their numerous combinations. Regression initially used the full set of candidate descriptors, and eliminated any candidates for which the p values were not < 0.05. The regression was then repeated with the reduced set of descriptors until all variables had satisfactory p values. Two additional criteria were that a) the standard error of the prediction (S_{pred}) should be < 2.0 log units and b) the adjusted R^2 of the prediction should be > 0.95.

These criteria led to the selection of the four descriptor variables below- each of which accounts either for a certain effect on the strength of ligands in binding Hg^{2+} or for a contribution to such an effect.

The independent nitrogen variable, # N, is the total number of amine and amide groups in each ligand.

The size-dependent nitrogen variable, # N * # C, is the product of # N and the number of carbon atoms in each ligand, # C, and accounts for the contribution of ligand size to the former effect.

The oxygen variable, # OH + 8.3 # COOH, is a linear combination of the number of OH and number of COOH groups in each ligand. The coefficient 8.3, obtained by trial and error in order to minimize the standard error of the QSPR, represents the much greater effectiveness of the numbers of carboxylic acid groups compared to those of alcohol groups.

Finally, the thiol variable, # Thiol, is the number of thiol groups in each ligand. (Note: for this data set # Thiol has only the values 0 and 1.)

The tolerable maximum predictive uncertainty of a QSPR depends on the requirements of the application, and the reasonably expectable minimum one on the uncertainties and variations in the calibration data. Cabaniss (Cabaniss 2008) achieved an uncertainty of ~1.0 log unit for QSPRs developed to predict the complexation of Al(III), Ca(II), Cd(II), Cu(II), Ni(II), Pb(II), and Zn(II) by organic ligands containing carboxylate, phenol, amine, ether, and alcohol functional groups. For Hg(II), however, the log K_{HgL}' data have a much larger range of values (from ~5 to ~35, that is ~30 orders of magnitude) and the individual data points are less reliable due to analytical difficulties discussed above. Given that the range in log K_{HgL}' is nearly double that of the metals examined previously, the acceptable S_{pred} may also be larger.

For the 'best fit' QSPR using all 44 ligands, S_{pred} = 1.60 log units and adjusted R^2 = 0.965, with the largest prediction errors of 3.59 log units for the overprediction of methionine

and 2.99 log units for the underprediction of 1,7-dithia-4-oxa-10-azacyclodecane (see Figures 2.3 and 2.4). Table 2.2 shows the descriptor coefficients and Table 2.5 the overall fit statistics. Further, in order to see the degree to which the descriptor variables correlated to one another, a correlation analysis was performed (see Table 2.3). In the correlation analysis, the variables are correlated using the coefficient of determination (R^2) square-root (that is R). The correlation analysis results show that the two descriptor variables of # N and #N * # C are highly correlated. This is chemically expectable because # N * # C accounts for the contribution of ligand size to the effect of # N on the strength of ligands in binding Hg^{2+}.

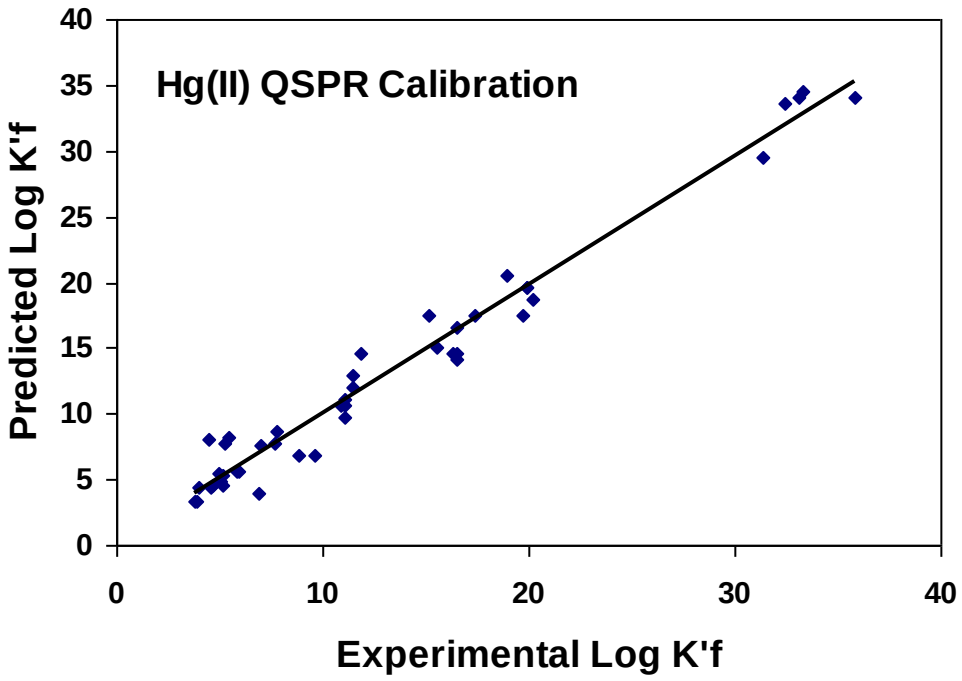


Figure 2.3 Hg(II) QSPR predictions versus experimental log K'_f values

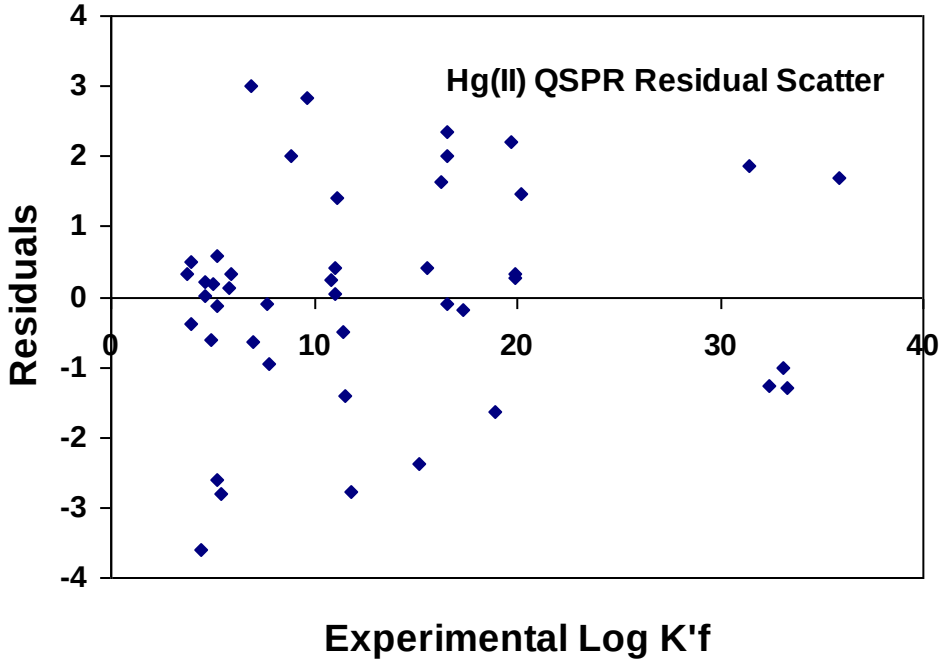


Figure 2.4 The QSPR calibration residuals versus experimental log K'_f

Table 2.2 QSPR Coefficients and their Standard Deviations.

Descriptor	Coefficient	Standard Deviation	p Value
# N	5.77	0.33	3.48×10^{-20}
# N * # C	-0.232	0.039	4.69×10^{-7}
# OH + 8.3 # COOH	0.411	0.020	8.85×10^{-23}
# Thiol	26.1	0.74	1.06×10^{-31}

Table 2.3 QSPR Descriptor Correlations.

	# N	# N * # C	# OH + 8.3 # COOH	# Thiol
# N	1			
# N * # C	0.890	1		
# OH + 8.3 # COOH	-0.264	-0.047	1	
# Thiol	-0.217	-0.207	-0.014	1

2.4 Internal Validation

The model was tested using leave-many-out (LMO) internal validation. Using a stratified, 5-fold-8-held-out approach, five sub-models were developed, each using 36 of the 44 ligand molecules and the same descriptor variables as the calibrated QSPR. The five sets of 8 test molecules included one thiol-containing molecule each (stratified selection). Each sub-model was calibrated on the set of 36 molecules using linear regression and the resulting descriptors used to calculate a root mean square error, $RMSE = [\Sigma (x_{exp} - x_{pred})^2 / N]^{(1/2)}$, for each of the 8 test molecules. The fraction of variance represented by the model, $q^2 = 1 - (RMSE^2/V_{exp})$, was also calculated.

Table 2.4 QSPR Sub-Model Descriptor Coefficients.

Validation #	# N	# N * # C	# OH + 8.3 # COOH	# Thiol
1	5.75	-0.231	0.421	25.6
2	5.48	-0.214	0.417	26.5
3	5.93	-0.232	0.398	25.7
4	5.84	-0.239	0.415	26.3
5	5.79	-0.235	0.409	26.4

Table 2.5 QSPR Sub-Model Statistics (N = 36).

Validation #	q^2	RMSE
1	0.974	1.52
2	0.949	2.02
3	0.945	2.14
4	0.974	1.55
5	0.977	1.39

Table 2.6 QSPR Statistics.

QSPR Calibration Fit Statistics		QSPR Validation Test Statistics	
$N_{calibration}$	44	$N_{validation}$	36
Adjusted R^2	0.965	q^2 (average)	0.964
S_{pred}	1.60	RMSE (average)	1.72

The test statistics are comparable to the fit statistics, but not identical. In regression fit statistics, the number of fitting parameters is one degree of freedom and is therefore subtracted from the number of data points when calculating the error of prediction (S_{pred}). In the test statistics, however, since the test is pure prediction, there are no fitting parameters, and the error of prediction is simply RMSE. The appropriate comparisons are thus between S_{pred} (calibration) and average RMSE (validation) and between adjusted R^2 (calibration) and q^2 (validation). The average fraction of variability described by the QSPR, 0.964, is almost equal to the adjusted R^2 of 0.965. The average RMSE of 1.72 obtained in the validation is

only slightly higher than S_{pred} of 1.60 obtained in the calibration. The close similarity indicates that the QSPR developed is robust.

2.5 Descriptor Coefficients

The test statistics are comparable to the fit statistics, but not identical. In regression fit statistics, the number of fitting parameters is one degree of freedom and is therefore subtracted from the number of data points when calculating the error of prediction (S_{pred}). In the test statistics, however, since the test is pure prediction, there are no fitting parameters, and the error of prediction is simply RMSE. The appropriate comparisons are thus between S_{pred} (calibration) and average RMSE (validation) and between adjusted R^2 (calibration) and q^2 (validation). The average fraction of variability described by the QSPR, 0.964, is almost equal to the adjusted R^2 of 0.965. The average RMSE of 1.72 obtained in the validation is only slightly higher than S_{pred} of 1.60 obtained in the calibration. The close similarity indicates that the QSPR developed is robust.

2.6 References

Basinger, M. A., Casas, J. S., Jones, M. M. & Weaver, A. D. 1981 Structural requirements for Hg(II) antidotes. *J. Inorg. Nucl. Chem.* **43**(6), 1419-1425.

Cabaniss, S. E. 1987 Titrator: An interactive program for aquatic equilibrium calculations. *Environ. Sci. Technol.* **21**(2), 209-210.

Cabaniss, S. E. 2008 Quantitative structure-property relationships for predicting metal binding by organic ligands. *Environ. Sci. Technol.* **42**(14), 5210-5216.

Cabaniss, S. E. 2009 Forward modeling of metal complexation by NOM: I. *A prori* prediction of conditional constants and speciation. *Environ. Sci. Technol.* **43**(8), 2838-2844.

Casas, J. S. & Jones, M. M. 1980 Mercury(II) complexes with sulfhydryl containing chelating agents: stability constant inconsistencies and their resolution. *J. Inorg. Nucl. Chem.* **42**(1), 99-102.

Haitzer, M., Aiken, G. R. & Ryan, J. N. 2002 Binding of mercury(II) to dissolved organic matter: The role of the mercury-to-DOM concentration ratio. *Environ. Sci. Technol.* **36**(16), 3564-3570.

Haitzer, M., Aiken , G. R. & Ryan, J. N. 2003 Binding of mercury(II) to aquatic humic substances: Influence of pH and source of humic substances. *Environ. Sci. Technol.* **37**(11), 2436-2441.

Harris, D.C. 2007 *Quantitative Chemical Analysis.* W. H. Freeman and Company, New York, NY, pp 228-249.

Hsu, H. & Sedlak, D. L. 2003 Strong Hg(II) complexation in municipal wastewater effluent and surface waters. *Environ. Sci. Technol.* **37**(12), 2743-2749.

Khwaja, A. R., Bloom, P, R. & Brezonik, P, L. 2006 Binding constants of divalent Mercury (Hg^{2+}) in soil humic acids and soil organic matter. *Environ. Sci. Technol.* **40**(3), 844-849.

Koszegi-Szalai, H. & Paal, T. L. 1999 Equilibrium studies of mercury(II) complexes with penicillamine. *Talanta.* **48**(2), 393-402.

Lenz, G. R. & Martell, A. E. 1964 Metal chelates of some sulfur-containing amino acids. *Biochemistry.* **3**(6), 745-750.

Milne, C. J., Kinniburgh, D. G., Van Riemsdijk, W. H. & Tipping E. 2003 Generic NICA-Donnan model parameters for metal-ion binding by humic substances. *Environ. Sci. Technol.* **37**(5), 958-971.

NIST Critically Selected Stability Constants of Metal Complexes Database, 2004 Version 8.0, US Dept. of Commerce, Gaithersburg, MD.

Perdue, E. M. & Parrish, R. S. 1987 Fitting multisite binding equilibria to statistical Distribution models: Turbo Pascal program for Gaussian models. *Computers Geosci.* **13**(6), 587-601.

Ravichandran, M. 2004 Interactions between mercury and dissolved organic matter-a review. *Chemosphere* **55**(3), 319-331.

Schmitt-Kopplin, P. & Junkers, J. 2003 Capillary zone electrophoresis of natural organic matter. *J. Chromatogr., A* **998**(1-2), 1-20.

Skyllberg, U., Xia, K., Bloom, P. R., Nater, E. A. & Bleam, W. F. 2000 Binding of mercury(II) to reduced sulfur in soil organic matter along upland-peat soil transects. *J. Environ. Qual.* **29**(3), 855-865.

Strand, R. & Lund, W. 1983 Complex formation of zinc, cadmium, and mercury with penicillamine. *J. Inorg. Biochem.* **19**(4), 301-309.

Tipping, E. 1998 Humic ion-binding model VI: An improved description of the interactions of protons and metal ions with humic substances. *Aquatic Geochem.* **4**(1), 3-48.

Supplementary Table 2.1 The Hg(II) QSPR calibration compounds, the functional groups, the descriptors, and the values for log K'_f

Compound Name	# OH	# COOH	# C	# N	# N*#C	# OH+8.3#COOH	# Thiol	log K'_f
Methylamine	0	0	1	1	1	0	0	4.94
Ethylamine	0	0	2	1	2	0	0	5.18
Acetic acid	0	1	2	0	0	8.3	0	3.74
Propanoic acid	0	1	3	0	0	8.3	0	3.90
Ethylenediamine	0	0	2	2	4	0	0	11.02
Diethylenetriamine	0	0	4	3	12	0	0	16.55
1,4,8,11-Tetraazaundecane	0	0	7	4	28	0	0	16.51
Glycine	0	1	2	1	2	8.3	0	7.76
Pyridine	0	0	5	1	5	0	0	5.19
Aniline	0	0	6	1	6	0	0	4.61
Triethylamine	0	0	6	1	6	0	0	3.99
Nitrilotris(ethyleneamine)	0	0	6	4	24	0	0	15.14
Triethanolamine	3	0	6	1	6	3	0	5.96
HEDTA	1	3	10	2	20	25.9	0	17.39
EDTA	0	4	10	2	20	33.2	0	18.92
Nitrilotriacetic acid	0	3	6	1	6	24.9	0	11.84
Oxalic acid	0	2	2	0	0	16.6	0	9.66
D-Tartaric acid	2	2	4	0	0	18.6	0	7.00
Citric acid	1	3	6	0	0	25.9	0	10.88
Pyridine-2-carboxylic acid	0	1	6	1	6	8.3	0	7.69
Triethylenetetramine	0	0	6	4	24	0	0	19.74
Tetramethylenediamine	0	0	4	2	8	0	0	11.10
Butylamine	0	0	4	1	4	0	0	5.04
Diethanolamine	2	0	4	1	4	2	0	5.79
Piperidine	0	0	5	1	5	0	0	4.62
Tetraethylenepentamine	0	0	8	5	40	0	0	19.92
1,1,3,3-Tetrakis(2-aminoethylthio)propane	0	0	11	4	44	0	0	11.49
1,1,4,4-Tetrakis(2-aminoethylthio)butane	0	0	12	4	48	0	0	11.48

1,1,5,5-Tetrakis(2-aminoethylthio)pentane	0	0	13	4	52	0	0	11.08
1,7-Dithia-4-oxa-10-azacyclodecane	0	0	8	1	8	0	0	6.91
Thiodiacetic acid	0	2	4	0	0	16.6	0	8.83
Dienkolic acid	0	2	7	2	14	16.6	0	15.53
L-Ethylenedithio-3,3'-bis(2-aminopropanoic acid)	0	2	8	2	16	16.6	0	16.29
L-Trimethylenedithio-3,3'-bis(2-aminopropanoic acid)	0	2	9	2	18	16.6	0	16.55
Thiobis(ethylenenitrilo)tetraacetic acid (TEDTA)	0	4	12	2	24	33.2	0	19.90
Ethylenebis(thioethylenenitrilo)tetraacetic acid	0	4	14	2	28	33.2	0	20.18
S-Methylcysteine	0	1	4	1	4	8.3	0	5.45
Methionine	0	1	5	1	5	8.3	0	4.44
Ethionine	0	1	6	1	6	8.3	0	5.20
Mercaptoacetic acid	0	1	2	0	0	8.3	1	31.39
L-cysteine	0	1	3	1	3	8.3	1	33.29
N-Acetyl-L-cysteine	0	1	5	1	5	8.3	1	35.82
D-penicillamine	0	1	5	1	5	8.3	1	33.12
N-Acetyl-D,L-penicillamine	0	1	7	1	7	8.3	1	32.40

CHAPTER 3

Quantitative Structure-Property Relationships for Predicting Cd(II) and Zn(II) Binding by Organic Ligands

3.1 Introduction

Cadmium and zinc are also in Group IIB, like mercury. Although common knowledge does not find them, especially zinc, as dangerous as mercury, they are still environmentally important metals. This chapter first introduces cadmium and zinc from an environmental viewpoint, addresses the importance of their binding to organic ligands, and then focuses on quantitative structure-property relationships (QSPRs) for predicting their binding to organic molecules similar in ligand group composition to natural organic matter (NOM).

3.1.1 Cadmium

Despite the existence of weak evidence for ultratrace essentiality of cadmium (Cd) in rats, cadmium is moderately toxic to all organisms and is a cumulative poison in mammals (Huheey et al. 1993). It accumulates in humans mainly in the kidneys and liver, and prolonged intake of even very small amounts leads to dysfunction of the kidneys (Greenwood & Earnshaw 1994). Also long-term high cadmium exposure may cause skeletal damage of which the first report was from Japan, where the itai-itai disease (a combination of osteomalacia and osteoporosis) was discovered in the 1950s (Järup 2003).

Cadmium was discovered in 1817 in a zinc ore and it is found in most zinc ores as CdS (Greenwood & Earnshaw 1994). Currently the main use of cadmium compounds is in rechargeable nickel-cadmium batteries (Järup 2003), and the major use of metallic cadmium is as a protective coating (Greenwood & Earnshaw 1994). Crustal material that is either

weathered on and eroded from the Earth's surface or injected into the Earth's atmosphere by volcanic activity is the principal natural source of cadmium in the environment (Callender 2003). Anthropogenic sources of cadmium in the environment include industrial emissions and the application of fertilizer and sewage sludge to farm land (Järup 2003). Anthropogenic as well as natural sources of cadmium may lead to contamination of soils (Järup 2003). It is especially important to note that pollution caused the itai-itai disease in Japan (Huheey *et al.* 1993) as the exposure to cadmium was caused by cadmium-contaminated water used for irrigation of local rice fields (Järup 2003).

The sorption of cadmium on particulate matter and bottom sediments is considered to be a major factor affecting its concentration in natural waters. The speciation of cadmium is generally considered to be dominated by dissolved forms except in cases where the concentration of suspended particulate matter is high (Callender 2003). The concentration of dissolved cadmium both in average world river water and in ocean water is 0.08 µg/L. In a typical natural aerobic freshwater aquatic system, the predominant inorganic cadmium species are Cd^{2+}, where pH is below 8, $CdCO_3^0$, where pH is between 8 and 10, and $Cd(OH)_2^0$, where pH is above 10 (Callender 2003). In the study of cadmium toxicity in natural waters the speciation of cadmium is significantly important. That is because the free ion form of the metal is thought to be the most available and toxic (Callender 2003). However, up to 70% of cadmium in natural waters is complexed by organic ligands (Morel & Hering 1993). Sander and co-workers (Sander *et al.* 2007) investigated how cadmium binding by natural organic ligands varied with pH, depth, and seasonality for lake waters. They observed that conditional stability constants relative to Cd^{2+} (aq) increased with decrease in depth. Their results also showed a higher percentage of Cd(II) complexed in the

summer. Still, regardless of the variation environmental factors cause in the percentage of complexed cadmium, cadmium complexation by dissolved natural organic matter (NOM) is significant. Xue and Sigg (Xue & Sigg 1998) use a technique involving ligand exchange and free Cd^{2+} concentration measurement to report examples of conditional stability constants of stable cadmium organic complexes in fresh water being $\sim 9 < \log K < \sim 10$. Therefore, because of the relation of cadmium complexation by dissolved NOM to Cd(II) toxicity, with the abundance of NOM in natural waters, estimating thermodynamic constants for it in aquatic systems deserves considerable attention.

3.1.2 Zinc

Until recently it was thought that zinc (Zn) was not harmful to the environment and that health risks were minimal compared to other heavy metals. In fact, it is generally less toxic than other heavy metals (Callender 2003). Still, being an essential element and micronutrient required for normal growth by plants and animals, zinc can be detrimental to organisms at both high and low concentrations (Cheng & Allen 2006). In order to understand zinc toxicokinetics, we need to know of the interactions zinc has with many chemicals to produce altered patterns of accumulation, metabolism, and toxicity (Eisler 1993). From an environmental point of view, however, high concentrations of zinc that are known to be toxic or even lethal to organisms have been observed in natural waters (Cheng & Allen 2006), and this is greatly concerning as natural waters are sources of life in nature.

Zinc is a naturally occurring metal. Its major ores are ZnS and $ZnCO_3$, and its largest supplier of ores is Canada (Greenwood & Earnshaw 1994). Zinc is widely used by humans for domestic and industrial purposes (Cheng & Allen 2006). Most importantly, accounting for 35-40% of output, it is used as an anti-corrosion coating (Greenwood & Earnshaw 1994).

As it is almost always associated with cadmium in mineral deposits and other earth materials, its principal natural source in the environment is the same as cadmium (Callender 2003). Major anthropogenic sources of zinc in the environment include electroplaters, smelting and ore processors, mine drainage, domestic and industrial sewage, combustion of solid wastes and fossil fuels, road surface runoff corrosion of zinc alloys and galvanized surfaces, and erosion of agricultural soils (Eisler 1993).

The concentration of dissolved zinc in average world river water is 0.60 μg/L, and the concentration of zinc in ocean water is 0.39 μg/L. In freshwater, the uncomplexed Zn^{2+} ion is the dominant inorganic zinc species at an environmental pH below 8, and uncharged $ZnCO_3^0$ is the main inorganic zinc species at higher pH (Callender 2003). Like cadmium, the speciation of zinc in natural waters is very important in the study of its toxicity. Some of zinc species most harmful to aquatic life under conditions of low pH, low alkalinity, low dissolved oxygen, and elevated temperatures are zinc aqua ions (Eisler 1993). On the other hand, zinc complexing with organic ligands in streams and lake waters with highly soluble organic carbon concentrations may occur significantly (Callender 2003).

Cheng and Allen (Cheng & Allen 2006) studied and compared the zinc binding characteristics of NOM from several representative surface waters. At the same pH, ionic strength, and temperature, similarity in zinc binding properties of the NOM isolated by reverse osmosis from different surface water sources tested was observed, and it was concluded that zinc complexation characteristics of the NOM used in their study did not depend on the NOM origin. They also observed the conditional binding constants increasing with increasing pH, indicating the zinc-NOM complexes become more stable at higher pH. The conditional binding constants of zinc by NOM (log K) in that study ranged from 4. to 7.

(Cheng & Allen 2006). Overall, with greater than 95% of zinc in oceans and 60-95% of zinc in estuarine waters being complexed by organic ligands (Morel & Hering 1993), the complexation of zinc by natural organic matter (NOM) in natural waters can leave only a small fraction of the total zinc as aqua ions. Therefore, in order to evaluate zinc toxicity in natural waters, we need to know the thermodynamic constants for zinc complexation by NOM in those aquatic systems.

3.2 The Necessity to Develop Quantitative Structure-Property Relationships

Similar to other metals, costliness and high consumption of time often make it desirable to estimate Cd-NOM and Zn-NOM binding constants without the benefit of additional experimental data. On the other hand, developing an alternative non-experimental approach faces a special issue: exact stereochemistry or even overall connectivity is unavailable for NOM ligands. Therefore, the most desirable approach is to use QSPRs to predict binding constants from hypothetical structures of NOM molecules using elemental composition and functional group data only. Recently Cabaniss developed such QSPRs for predicting binding constants of environmentally and geochemically interesting metals, including zinc, with small molecules similar in ligand group composition to NOM (Cabaniss 2008).

The QSPRs presented in Cabaniss' work are suitable for use with synthetic or geochemical ligands for which minimal structural information is available, provided that an accuracy of ±1 log unit in the conditional binding constants is acceptable (Cabaniss 2008). However, the criteria used to select ligand molecules for the calibration data set had a deficiency as for the universal applicability of the QSPRs for predicting metal binding by NOM: absence of S atoms. Sulfur is a minor constituent of most NOM samples (typically

≤1% by weight), and only a portion of the sulfur is present as thiols (Xia *et al.* 1998). However, in a study of cadmium complexation by Swannee River NOM Hertkorn and co-workers (Hertkorn *et al.* 2004) observed not only that cadmium was coordinated by oxygen and nitrogen ligands in NOM, but also that under alkaline conditions a small, but unquantifiable, percentage of cadmium is coordinated to S ligands. Also, in a study of zinc complexation in organic soils, Karlsson and Skyllberg (Karlsson & Skyllberg 2007) concluded that zinc forms mainly inner-sphere complexes with a mixture of 6-fold coordination with O ligands and 4-fold coordination with S and O/N ligands (see Appendix 2). Further, in biological systems, cadmium acts by binding to the –SH group of cysteine residues in proteins and so inhibits SH enzymes (Greenwood & Earnshaw 1994), and metallothioneins play an important role in zinc homeostasis and in protection against zinc poisoning (Eisler 1993). Therefore, thiol groups in NOM are also worthy of special attention in studying both Cd-NOM and Zn-NOM complexation. This work uses the QSPRs developed by Cabaniss for predicting cadmium and zinc binding by organic ligands (Cabaniss 2008) and develops more universally applicable QSPRs by incorporating cadmium-thiol and zinc-thiol complexes.

3.3 The Calibration Data Set

3.3.1 Selecting the Ligand Molecules and the Thermodynamic Equilibrium Constants

Ligand molecules for the calibration data set used in this work were selected so that (a) molecules contain one or more of the functional groups: carboxylic acid, amine, and thiol, (b) a 1:1 binding constant for cadmium and zinc is available, and (c) no molecule has more than one thiol group.

The calibration data set contains 65 distinct ligand molecules with 1:1 binding constants for cadmium and 69 for zinc. Like the Hg(II) QSPR in Chapter 2, the equilibrium constants (K_{CdL} and K_{ZnL}) for each 1:1 complex were selected at 25 ^{0}C and ionic strength 0.1, unless no constant at such conditions was available in which case a constant determined at similar conditions was selected. All the thermodynamic equilibrium constants were obtained from NIST Standard Reference Database 46 Version 8.0 (NIST 2004).

3.3.2 Calculating and Using the Conditional Equilibrium Constants

The equations for the formation of each complex and the corresponding equilibrium constants K_{CdL} and K_{ZnL} are:

$$Cd^{2+} + L^{n-} \rightleftarrows CdL^{2-n} \qquad K_{CdL} = [CdL^{2-n}] / [Cd^{2+}][L^{n-}] \qquad (1)$$

$$Zn^{2+} + L^{n-} \rightleftarrows ZnL^{2-n} \qquad K_{ZnL} = [ZnL^{2-n}] / [Zn^{2+}][L^{n-}] \qquad (2)$$

To facilitate integration with models of NOM and to make predictions at neutral pH, it is useful to know the pH-adjusted conditional constant. For each organic molecule in the data set, therefore, K_{CdL} or K_{ZnL} is used to calculate a corresponding conditional equilibrium constant, K_{CdL}' or K_{ZnL}'.

$$K_{CdL}' = [CdL^{2-n}] / [Cd^{2+}][L'] \qquad (3)$$

$$K_{ZnL}' = [ZnL^{2-n}] / [Zn^{2+}][L'] \qquad (4)$$

where $[L']$ is the sum of the concentrations of all ligand species which are not complexed by Cd(II) or Zn(II). Combining equations (1) and (3) together and (2) and (4) together, it may be written

$$K_{CdL}' = K_{CdL} \cdot [L^{n-}]/[L'] = K_{CdL}\,\alpha_n \qquad (5)$$

$$K_{ZnL}' = K_{ZnL} \cdot [L^{n-}]/[L'] = K_{ZnL}\,\alpha_n \qquad (6)$$

As it was mentioned in Chapter 2, α_n is the fraction of the uncomplexed ligand which exists in the $-n$ charge state. The use of the word "conditional" in naming K_{CdL}' and K_{ZnL}' reflects the fact that it varies with pH, ionic strength, and temperature. In this study, like in Chapter 2, pH 7.0 is chosen, but as α_n can be calculated for other pH values, so can K_{CdL}' and K_{ZnL}'. This enables us to use the model for any environmental pH of interest provided acid pKa's are known.

3.4 Descriptor Selection and Calibration of QSPR

Like in the case of the Hg(II) QSPR in Chapter 2, the QSPRs presented here for Cd and Zn follow the general linear equation

$$\text{Log } K_{ML}' = a_1x_1 + a_2x_2 + \ldots + a_nx_n \qquad (7)$$

where x_i represents the ith descriptor variables and a_i parameters represent the coefficients of those variables. Also, since there was no good theoretical justification for a nonzero intercept, the intercept was set to zero.

On an Excel spreadsheet regressions used some of candidate descriptors Cabaniss (Cabaniss 2008) had used for his Cd(II) and Zn(II) QSPRs; however, the presence of thiol was also taken into account. Any candidates for which the p values were not < 0.05 were eliminated later. The regressions were then repeated with the reduced set of descriptors until all variables had satisfactory p values. Also the QSPRs were accepted only if an uncertainty of ~1.0 log unit was achieved. This was a requirement set by Cabaniss (Cabaniss 2008) for the QSPRs developed by him to predict the complexation of Al(III), Ca(II), Cd(II), Cu(II), Ni(II), Pb(II), and Zn(II) by organic ligands containing carboxylate, phenol, amine, ether, and alcohol functional groups. Cabaniss set the requirement of S_{pred} ~ 1.0 log unit as it would, to

our knowledge, be the smallest uncertainty achieved in predicting metal binding by organic ligands without requiring any steric or connectivity information (Cabaniss 2008).

The error minimization requirements led to the selection of the following descriptor variables:

The carboxylic acid group count variable, # COOH, is the number of carboxylic acid groups in each ligand.

The amine group count variable, # Amine, is the number of amine groups in each ligand.

The thiol group count variable, # Thiol, is the number of thiol groups in each ligand. (Note: for this data set # Thiol has only the values 0 and 1.)

The square of the ligand number multigroup variable, LN^2, is calculated as

$$LN^2 = (\#\ COOH + \#\ Amine + \#\ Thiol)^2 \qquad (8)$$

and partially accounts for the "chelate effect"; the ability of multidentate ligands to form more stable metal complexes than those formed by similar monodentate ligands (Harris 2007).

The charge density descriptor, Z_{dens}, is calculated for pH 7.0 assuming that all carboxylic acids and thiols but no amines were deprotonated, resulting in charge Z = (# COOH + # SH - # Amine) and

$$Z_{dens} = Z/MW \qquad (9)$$

where MW is the molecular weight.

The heteroatom-to-carbon ratio multigroup variable, Het:C, is calculated from elemental composition as

$$Het{:}C = (\#\ O + \#\ N + \#\ S) / \#\ C \qquad (10)$$

where # O, # N, # S, and # C refer to the numbers of atoms of the elements in the molecule.

Taking into account the coexistence of both amine and carboxylic acid ligands in molecules where both ligand groups were present, especially where such molecules of smaller size are taken into account, appears to improve the robustness of the model. This effect is mathematically reflected in the model as an amino acid variable AA calculated as

$$AA = 1/MW \quad (11)$$

if both amine and carboxylic acid groups were present on a molecule. AA was set to 0 otherwise.

For the 'best fit' cadmium QSPR using all 63 ligands, S_{pred} = 0.935 log units and adjusted R^2 = 0.931, with the largest prediction errors of 2.79 log units for the underprediction of imino di-3 propionic acid and 2.38 log units for the overprediction of glutamic acid (see Figures 3.1 and 3.2). Table 3.1 shows the cadmium QSPR descriptor coefficients, and Table 3.2 shows its descriptor correlations. The highest absolute values of correlation coefficients appear for # Amine correlating to Z_{dens} and to AA. These correlations are chemically expectable because # Amine both has a frequently encountered presence in the ligands for which Z is calculated and has a defining role for the AA variable. For the 'best fit' zinc QSPR using all 68 ligands, S_{pred} = 0.984 log units and adjusted R^2 = 0.934, with the largest prediction errors of 2.79 log units for the underprediction of Adrenaline and 2.12 log units for the overprediction of glutamic acid (see Figures 3.3 and 3.4). Table 3.3 shows the zinc QSPR descriptor coefficients, and Table 3.4 shows its descriptor correlations. The most correlated variables are LN^2 and Het:C. This correlation is chemically expectable because LN^2 partially accounts for the "chelate effects" and the numerator in Het:C (that is # O + # N + # S) may be viewed as a mathematical expression

representing the abundance of metal binding sites in organic molecules. The overall fit statistics for both cadmium and zinc QSPRs are shown in Table 3.5.

Table 3.1 Cadmium QSPR Coefficients and their Standard Deviations.

Descriptor	Coefficient	Standard Deviation	p Value
# Amine	1.67	0.40	9.20×10^{-5}
# Thiol	2.91	0.34	1.14×10^{-11}
LN^2	0.278	0.029	1.83×10^{-13}
Z_{dens}	69.7	16.6	9.20×10^{-5}
AA	-148.	47.	0.00264

Table 3.2 Cadmium QSPR Descriptor Correlations.

	# Amine	**# Thiol**	**LN^2**	**Z_{dens}**	**AA**
# Amine	1				
# Thiol	0.042	1			
LN^2	0.428	-0.064	1		
Z_{dens}	-0.654	0.009	0.176	1	
AA	0.657	-0.009	0.246	-0.254	1

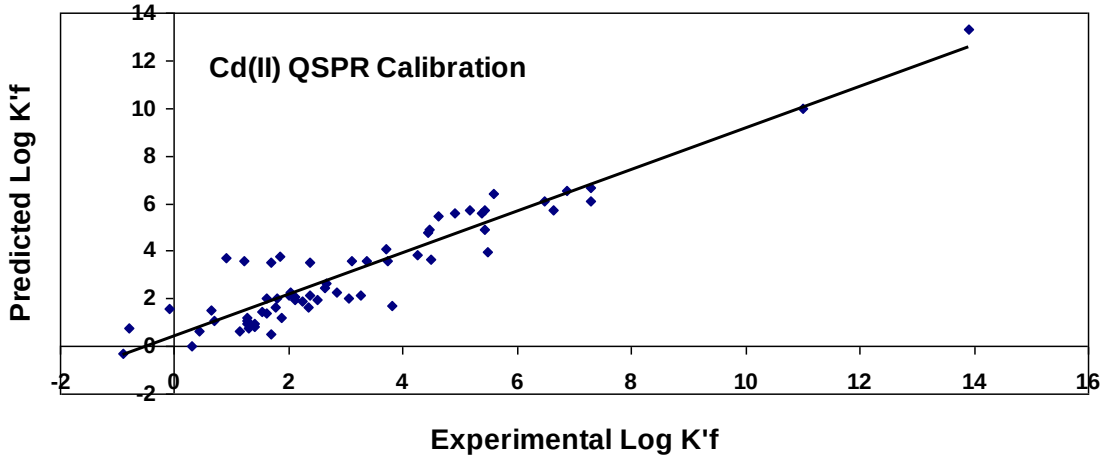


Figure 3.1 Cd(II) QSPR predictions versus experimental log K'_f values

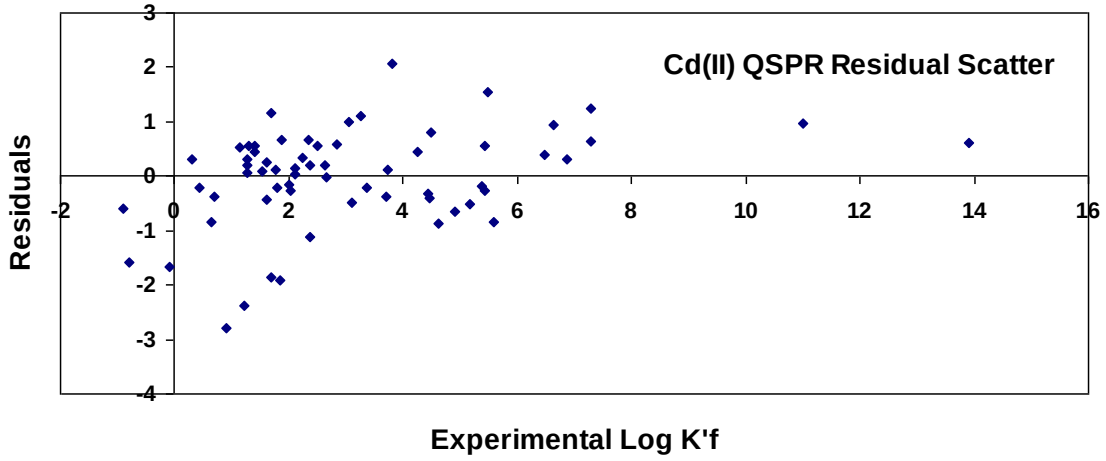


Figure 3.2 Cd(II) QSPR calibration residuals versus experimental log K'_f

Table 3.3 Zinc QSPR Coefficients and their Standard Deviations.

Descriptor	Coefficient	Standard Deviation	p Value
# Thiol	1.14	0.37	0.00323
LN^2	0.363	0.020	3.72×10^{-27}
Het:C	0.993	0.179	5.61×10^{-7}

Table 3.4 Zinc QSPR Descriptor Correlations.

	# Thiol	LN^2	Het:C
# Thiol	1		
LN^2	-0.035	1	
Het:C	0.087	0.310	1

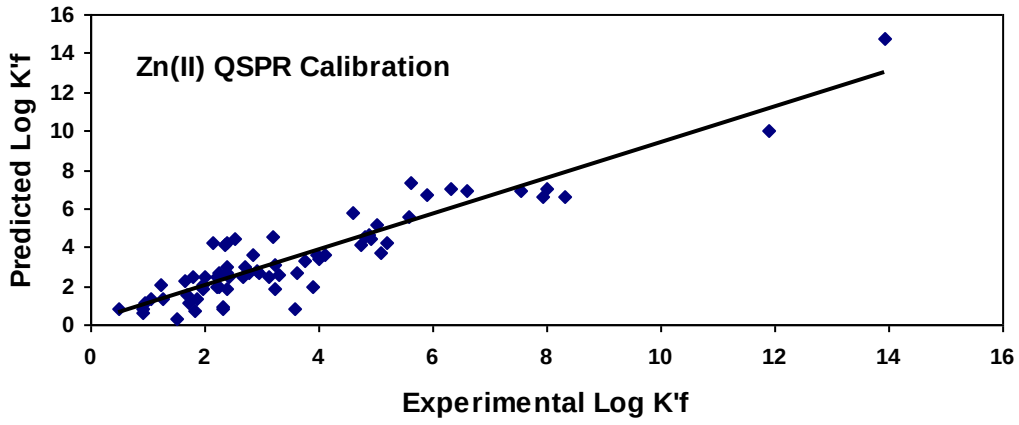


Figure 3.3 Zn(II) QSPR predictions versus experimental log K'_f values

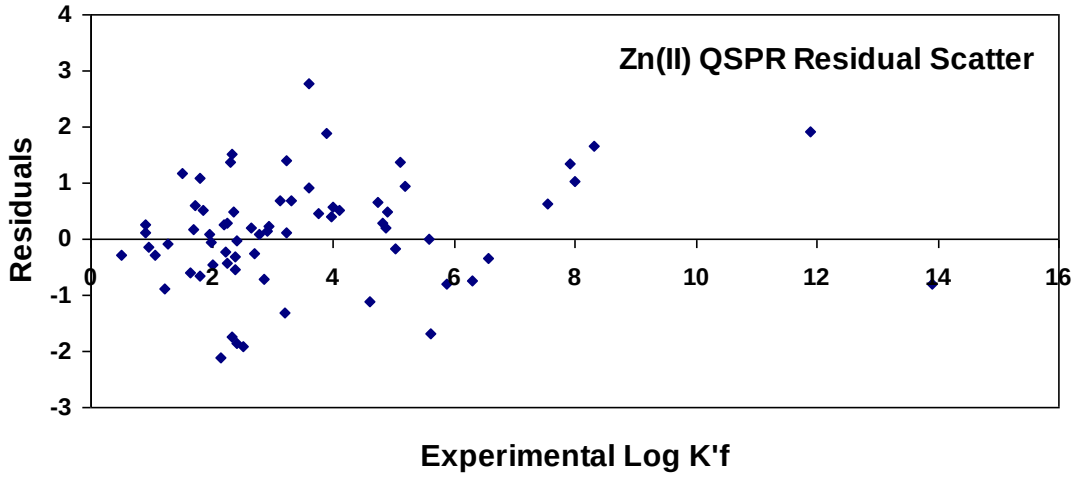


Figure 3.4 Zn(II) QSPR calibration residuals versus experimental log K'_f

Table 3.5 QSPRs Statistics.

Statistics	Cd	Zn
$N_{calibration}$	63	68
Adjusted R^2	0.931	0.934
S_{pred}	0.935	0.984

3.5 Discussion

3.5.1 The Proposed Addition of a Thiol Term to Cabaniss' Cd(II) and Zn(II) QSPRs

Why couldn't the full sets of QSPR descriptor variables for cadmium and zinc used by Cabaniss (Cabaniss 2008) be used with the addition of a thiol term? This approach would be the easiest if the resulting QSPRs were satisfactory. However, the addition of a thiol term to the full set of Cabaniss' Cd(II) QSPR descriptor variables makes one of the descriptor variables (that is # Ether) insignificant. Also, the addition of a thiol term to the full set of Cabaniss' Zn(II) QSPR descriptor variables makes one of the descriptor variables (that is #

COOH) insignificant. Therefore, the outcome models are unsatisfactory, although they have satisfactory values of adjusted R^2 and S_{pred} (see Tables 3.6-3.10 and Figures 3.5-3.8).

Table 3.6 "Proposed" QSPRs Statistics.

Statistics	Cd	Zn
$N_{calibration}$	63	68
Adjusted R^2	0.933	0.940
S_{pred}	0.917	0.916

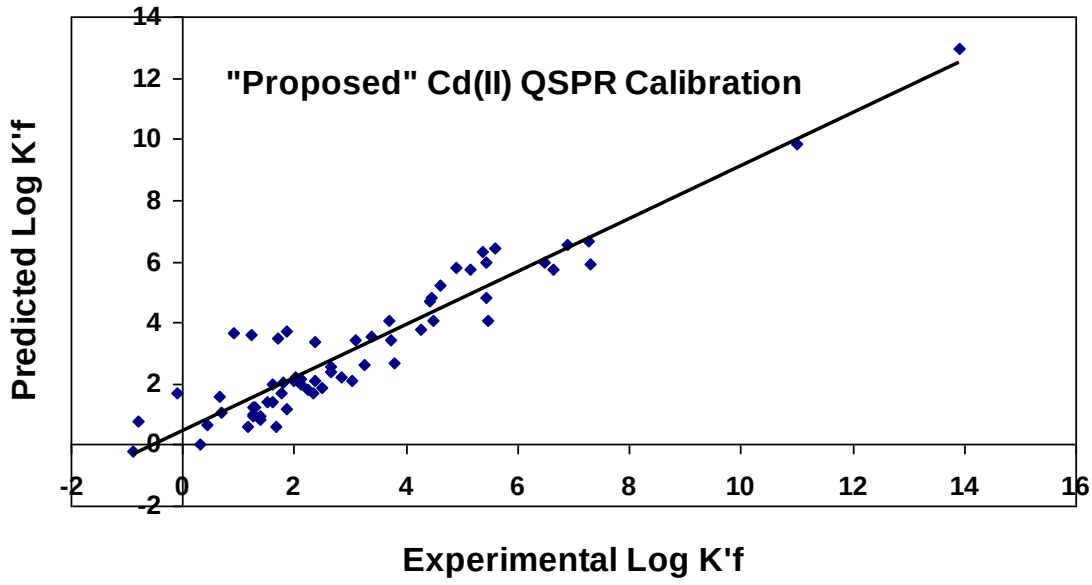


Figure 3.5 "Proposed" Cd(II) QSPR predictions versus experimental log K'_f values

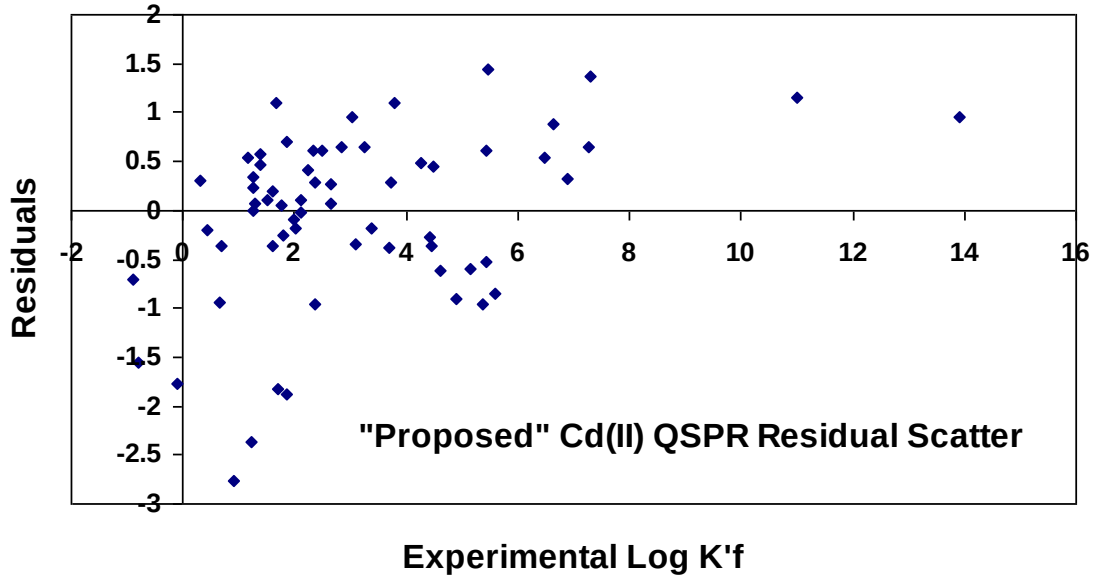


Figure 3.6 "Proposed" Cd(II) QSPR calibration residuals versus experimental log K'_f

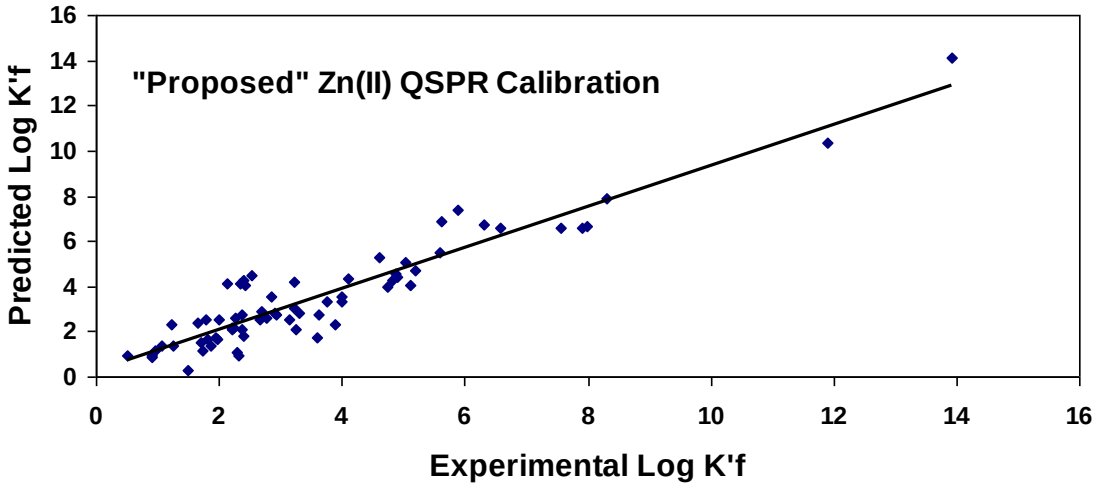


Figure 3.7 "Proposed" Zn(II) QSPR predictions versus experimental log K'_f values

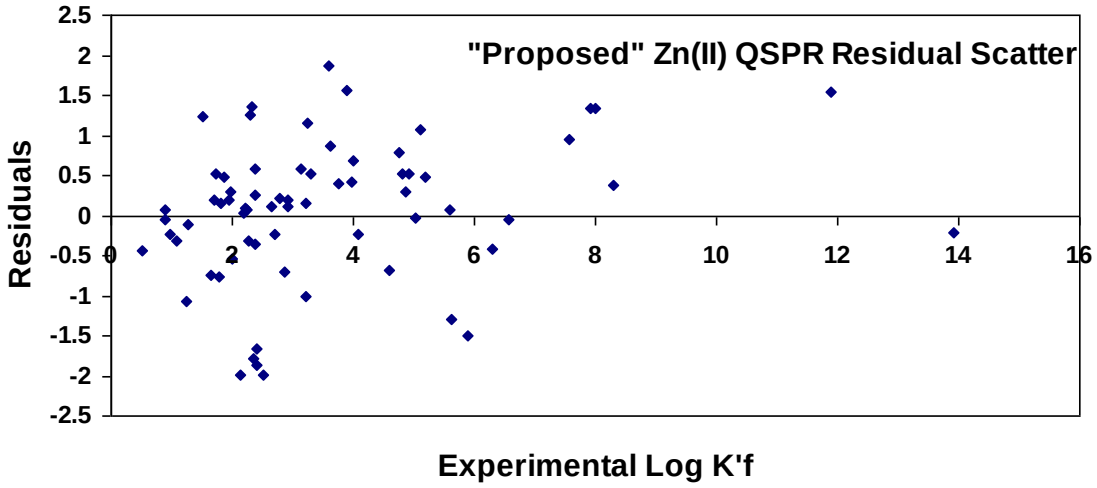


Figure 3.8 "Proposed" Zn(II) QSPR calibration residuals versus experimental log K'_f

Table 3.7 "Proposed" Cadmium QSPR Coefficients and their Standard Deviations.

Descriptor	Coefficient	Standard Deviation	p Value
# Amine	1.78	0.39	3.17×10^{-5}
# Ether	0.521	0.286	0.074
LN^2	0.262	0.030	3.94×10^{-12}
Z_{dens}	69.2	16.3	7.89×10^{-5}
AA	-146.	46.1	0.00242
# Thiol	2.93	0.34	5.51×10^{-12}

Table 3.8 "Proposed" Cadmium QSPR Descriptor Correlations.

	# Amine	# Ether	LN^2	Z_{dens}	1/MW	# Thiol
# Amine	1					
# Ether	-0.266	1				
LN^2	0.428	0.171	1			
Z_{dens}	-0.654	0.205	0.176	1		
1/MW	0.657	-0.214	0.246	-0.254	1	
# Thiol	0.042	-0.135	-0.064	0.009	-0.009	1

Table 3.9 "Proposed" Zinc QSPR Coefficients and their Standard Deviations.

Descriptor	Coefficient	Standard Deviation	p Value
# COOH	0.342	0.25	0.173
# Amine	1.10	0.33	0.00139
LN^2	0.275	0.0455	9.62×10^{-8}
Het:C	0.311	2.46	0.0167
AA	-122.5	43.10	0.00607
# Thiol	1.37	0.40	0.00103

Table 3.10 "Proposed" Zinc QSPR Descriptor Correlations.

	# COOH	# Amine	LN^2	Het:C	AA	# Thiol
# COOH	1					
# Amine	-0.128	1				
LN^2	0.706	0.522	1			
Het:C	0.359	0.025	0.310	1		
AA	-0.087	0.671	0.286	0.155	1	
# Thiol	-0.295	-3.600×10^{-18}	-0.035	0.087	0.022	1

3.5.2 Comparison of the Cd(II) and Zn(II) QSPRs

The Cd(II) QSPR has a smaller S_{pred} than that of the Zn(II) QSPR. However, both QSPRs have the desired S_{pred} of ~1 log unit and the difference in adjusted R^2 means that the Cd(II) QSPR and the Zn(II) QSPR describe the variance almost equally well. Also, the Zn(II) QSPR uses fewer descriptor variables than the Cd(II) QSPR does, and this means that the former has the advantage of requiring less input information than the latter. The p values of all Cd(II) QSPR descriptor coefficients show that they are significantly descriptive. Same is shown by the p values of all Zn(II) QSPR descriptor coefficients. Except for the AA variable, the descriptor coefficients of the Cd(II) QSPR are entirely positive. This means that all the descriptor variables used in the Cd(II) QSPR except the AA variable have positive contributions to the strength of Cd(II)-ligand binding. This is chemically expected because increases in the number of ligands increase the stability of metal complexes. The positive contributions of the # Thiol and LN^2 variables in the Zn(II) QSPR are also chemically expected because of the same reason. Why does Het:C have a positive contribution to the strength of Zn(II) binding by organic ligands? # C may be viewed as an approximate representation of organic ligand size (see Chapter 2), and the numerator in Het:C (that is # O + # N + # S) may be viewed as a mathematical expression representing the abundance of metal binding sites in organic molecules. Therefore, Het:C may be viewed as a quantitative expression for "ligand distributed presence" in organic molecules.

The # Thiol descriptor coefficient for the cadmium QSPR is 1.77 units larger than that for the zinc QSPR. This reflects the stronger affinity Cd(II) has for thiols. However, the LN^2 descriptor coefficient for the zinc QSPR is 0.085 units larger than that for the cadmium QSPR. Since LN^2 represents the chelate effect and is based on a collection of carboxylic

acid, amine, and thiol ligands, this difference may be attributed to the stronger affinity Zn(II) has for carboxylic acid and amine ligands than does Cd(II).

3.5.3 Comparison with Cabaniss' Cd(II) and Zn(II) QSPRs

In both Cd(II) and Zn(II) QSPRs developed by Cabaniss (Cabaniss 2008), S_{pred} = 0.91. This number is 0.025 log units smaller than the S_{pred} of the Cd(II) QSPR and 0.074 log units smaller than the Z(II) QSPR which have been developed here. However, both of these numerical differences are less than $3/40^{th}$ of a log unit and may be expected with increasing the number of calibration data points by 24% for Cd(II) and 13% for Zn(II), especially considering that using a smaller number of descriptor variables was considered an advantageous factor.

3.5.4 Comparison with the Hg(II) QSPR

As it was mentioned in the last chapter, the tolerable maximum predictive uncertainty of a QSPR depends on the requirements of the application, and the reasonably expectable minimum one on the uncertainties and variations in the calibration data. For Hg(II), the log K_{HgL}' data have a range of values from ~5 to ~35, that is ~30 orders of magnitude, and the analytical difficulties discussed in Chapter 2 affect the reliability of the individual data points. Given that the ranges in both log K_{ZnL}' and log K_{CdL}' is nearly half that of log K_{HgL}', the highest acceptable S_{pred} should also be smaller than that for Hg(II) QSPR (i.e. S_{pred} = 1.60 log units). Consequently both Zn(II) and Cd(II) QSPRs were developed so that the desired S_{pred} of ~1 log unit was achieved.

A striking difference between the Hg(II) QSPR and the QSPRs developed for Cd(II) and Zn(II) here is about the coefficients of # Thiol descriptor variable. The coefficient of # Thiol in the Hg(II) QSPR is 26.1, while that is 2.91 for the Cd(II) QSPR and 1.14 for the

Zn(II) QSPR. These differences are clearly consistent with the various levels of M(II) (M = Zn, Cd, Hg) affinity for thiols. Looking at a couple of specific examples sheds light on the differences. log K_{ML}', where L = L-cysteine, is 33.29 for Hg(II), 5.59 for Cd(II), and 4.60 for Zn(II). Similarly, where L = D-penicillamine, it is 33.12 for Hg(II), 6.87 for Cd(II), and 5.03 for Zn(II). The facts that R-S^- is a soft base and that Hg^{2+} and Cd^{2+} are soft acids while Zn^{2+} is a borderline acid (Miessler & Tarr 2000) describe some of these observations. More of the observations are described by the fact that Hg^{2+} is more massive than Cd^{2+} and a softer acid.

3.6 References

Cabaniss, S. E. 2008 Quantitative structure-property relationships for predicting metal binding by organic ligands. *Environ. Sci. Technol.* **42**(14), 5210-5216.

Callender, E. Heavy metals in the environment—historical trends. In *Environmental Geochemistry* (ed. B.S. Lollar) 2003 Vol. 9 *Treatise on Geochemistry* (eds. H.D. Holland and K.K. Turekian). Elsevier-Pergamon, Oxford, pp 67-105.

Cheng, T. & Allen, H. E. 2006 Comparison of zinc complexation properties of dissolved natural organic matter from different surface waters. *J. Environ. Manage.* **80**(3), 222-229.

Eisler, R. 1993 *Zinc Hazards to Fish, Wildlife, and Invertebrates: A Synoptic Review.* – Technical rept. series. Patuxent Wildlife Research Center, Laurel, MD.

Greenwood, N. N. & Earnshaw, A. 1994 *Chemistry of the Elements.* Pergamon Press, Tarrytown, NY.

Harris, D.C. 2007 *Quantitative Chemical Analysis.* W. H. Freeman and Company, New York, NY, pp 228-249.

Hertkorn, N., Perdue, E. M. & Kettrup, A. 2004 A potentiometric and ^{113}Cd NMR study of cadmium complexation by natural organic matter at two different magnetic field strengths. *Anal. Chem.* **76**(21), 6327-6341.

Huheey, J. E., Keiter, E. A. & Keiter, R. L. 1993 *Inorganic Chemistry: Principles of Structure and Reactivity.* HarperCollins College Publishers, New York, NY.

Järup, L. 2003 Hazards of heavy metal contamination. *Br. Med. Bull.* **68**(1), 167-182.

Karlsson, T. & Skyllberg, U. 2007 Complexation of zinc in organic soils—EXAFS evidence for sulfur associations. *Environ. Sci. Technol.* **41**(1), 119-124.

Miessler, G. L. & Tarr, D. A. 2000 *Inorganic Chemistry.* Prentice-Hall Inc., New Jersey, NJ.

Morel, F. M. M. & Hering, J. G. 1993 *Principles and Applications of Aquatic Chemistry.* Wiley-Interscience, New York, NY.

NIST Critically Selected Stability Constants of Metal Complexes Database, 2004 Version 8.0, US Dept. of Commerce, Gaithersburg, MD.

Sander, S., Ginon, L., Anderson, B. & Hunter, K. A. 2007 Comparative study of organic Cd and Zn complexation in lake waters – seasonality, depth and pH dependence. *Environ. Chem.* **4**(6), 410-423.

Xia, K., Weesner, F., Bleam, W. F., Bloom, P. R., Skyllberg, U. L. & Helmke, P. A. 1998 XANES studies of oxidation states of sulfur in aquatic and soil humic substances. *Soil Sci. Soc. Am. J.* **62**(5), 1240-1246.

Xue, H. B. & Sigg, L. 1998 Cadmium speciation and complexation by natural organic ligands in fresh water. *Anal. Chim. Acta* **363**(2-3), 249-259.

Supplementary Table 3.1 The Cd(II) QSPR calibration compounds, the functional groups, the descriptors, and the values for log K'_f

Compound Name	MW	# COOH	# Amine	# Thiol	LN2	Zdens	AA	log K'_f
L-cysteine	121	1	1	1	9	0.008254	0.008254	5.59
N-Acetyl-L-cysteine	163	1	0	1	4	0.012255	0	4.47
D-penicillamine	149	1	1	1	9	0.006702	0.006702	6.87
N-Acetyl-D-penicillamine	191	1	0	1	4	0.010458	0	4.43
2-Mercapto ethylamine	77	0	1	1	4	0	0	5.16
L-cysteinyl glycine	178	1	1	1	9	0.005611	0.005611	7.28
DL-(2-Mercapto propionyl) glycine	163	1	0	1	4	0.012255	0	5.44
L-cysteine methyl ester	135	0	1	1	4	0	0	6.64
2-Mercapto ethanol	78	0	0	1	1	0.012798	0	3.70
Methylamine	31	0	1	0	1	-0.03226	0	-0.9
Salicylic Acid	138	1	0	0	1	0.007246	0	-0.8
Adrenaline	183	0	1	0	1	-0.00546	0	-0.1
Catechol	98	0	0	0	0	0	0	0.3
Caffeic acid	180	1	0	0	1	0.005556	0	0.44
Dopamine	153	0	1	0	1	-0.00654	0	0.65
Pyruvic Acid	88	1	0	0	1	0.011364	0	0.69
Imino di-3 propanoic acid	161	2	1	0	9	0.006211	0.006211	0.9
Gluconic acid	196	1	0	0	1	0.005102	0	1.15
Glutamic	137	2	1	0	9	0.007299	0.007299	1.22
2 hydroxy butanoic	104	1	0	0	1	0.009615	0	1.26
Lactic	90	1	0	0	1	0.011111	0	1.26
Leucine	93	1	1	0	4	0	0.010753	1.26
2 methoxy benzoic	152	1	0	0	1	0.006579	0	1.3
4 hydroxy butanoic	104	1	0	0	1	0.009615	0	1.39
Benzoic acid	122	1	0	0	1	0.008197	0	1.4
Acetic	60	1	0	0	1	0.016667	0	1.52
Tartaric	152	2	0	0	4	0.013158	0	1.6
Glutamine	104	1	1	0	4	0	0.009615	1.62
Glycine	65	1	1	0	4	0	0.015385	1.68
Aspartic	123	2	1	0	9	0.00813	0.00813	1.69
Glycylglycine	132	1	1	0	4	0	0.007576	1.77
TriGlycine	189	1	1	0	4	0	0.005291	1.8
Glycylaspartic	190	2	1	0	9	0.005263	0.005263	1.85
Glycolic	76	1	0	0	1	0.013158	0	1.87
Pentanedioic	132	2	0	0	4	0.015152	0	2
Succinic	118	2	0	0	4	0.016949	0	2.03

Hexanedioic	146	2	0	0	4	0.013699	0	2.1
Ethylene diamine	56	0	2	0	4	-0.03571	0	2.12
Phenyl malonic	180	2	0	0	4	0.011111	0	2.23
Asparagine	132	1	1	0	4	0	0.007576	2.34
Malic	134	2	0	0	4	0.014925	0	2.36
Hemimellitic	210	3	0	0	9	0.014286	0	2.38
Phthalic Acid	166	2	0	0	4	0.012048	0	2.5
Malonic	104	2	0	0	4	0.019231	0	2.64
Oxalic	90	2	0	0	4	0.022222	0	2.65
Tartronic	120	2	0	0	4	0.016667	0	2.85
DOPA (2-amino-3,4 dihydroxyphenyl propanoic	197	1	1	0	4	0	0.005076	3.04
Isocitric	192	3	0	0	9	0.015625	0	3.1
Diglycolic	134	2	0	0	4	0.014925	0	3.25
Iminodiacetic acid (IDA)	133	2	1	0	9	0.007519	0.007519	3.37
Citric	192	3	0	0	9	0.015625	0	3.72
Benzene 1,2 bis oxyacetic acid	226	2	0	0	4	0.00885	0	3.8
N-benzyl imino diacetic	213	2	1	0	9	0.004695	0.004695	4.26
Carboxy methoxymalonic	178	3	0	0	9	0.016854	0	4.49
TMS (3 hydroxy oxapentane 1,2,4,5 tetracarboxylic)	266	4	0	0	16	0.015038	0	4.62
rac Oxybisbutanedioic	250	4	0	0	16	0.016	0	4.91
meso Oxybisbutanedioic	250	4	0	0	16	0.016	0	5.38
Ditartronic	222	4	0	0	16	0.018018	0	5.44
EDMA	114	1	2	0	9	-0.00877	0.008772	5.47
cis 2,6 dicarboxy piperidine acetic acid	231	3	1	0	16	0.008658	0.004329	6.48
NTA	159	3	1	0	16	0.012579	0.006289	7.3
HEDTA	278	3	2	0	25	0.003597	0.003597	10.99
EDTA	204	4	2	0	36	0.009804	0.004902	13.92

Supplementary Table 3.2 The Zn(II) QSPR calibration compounds, the functional groups, the descriptors, and the values for log K'_f

Compound Name	# COOH	# Amine	# Thiol	LN^2	Het:C	log K'_f
L-cysteine	1	1	1	9	1.333	4.60
N-Acetyl-L-cysteine	1	0	1	4	1	2.86
D-penicillamine	1	1	1	9	0.8	5.03
N-Acetyl-D-penicillamine	1	0	1	4	0.714	3.75
2-Mercaptoethylamine	0	1	1	4	1	4.09
L-cysteinylglycine	1	1	1	9	1.2	5.59
DL-(2-Mercaptopropionyl) glycine	1	0	1	4	1	3.98
Mercaptoacetic acid	1	0	1	4	1.5	4.75
Salicylic Acid	1	0	0	1	0.429	0.5
2 methoxy benzoic	1	0	0	1	0.429	0.9
Benzoic acid	1	0	0	1	0.286	0.9
4 hydroxy butanoic	1	0	0	1	0.75	0.96
Acetic	1	0	0	1	1	1.07
Hexanedioic	2	0	0	4	0.667	1.23
Pyruvic Acid	1	0	0	1	1	1.26
Catechol	0	0	0	0	0.333	1.5
Pentanedioic	2	0	0	4	0.8	1.64
Gluconic acid	1	0	0	1	1.167	1.7
2 hydroxy butanoic	1	0	0	1	0.75	1.72
Succinic	2	0	0	4	1	1.79
Dopamine	0	1	0	1	0.375	1.81
Lactic	1	0	0	1	1	1.86
Glycolic	1	0	0	1	1.5	1.95
Leucine	1	1	0	4	0.6	1.98
Benzene 1,2 bis oxyacetic acid	2	0	0	4	1	2
Glutamic	2	1	0	9	1	2.14
Phthalic Acid	2	0	0	4	0.5	2.2
Glutamine	1	1	0	4	1	2.22
Cyclohexane 1,1 dicarboxylic acid	2	0	0	4	0.5	2.24
Glycylglycine	1	1	0	4	1.25	2.26
Protocatechuic acid	1	0	0	1	0.571	2.3
Caffeic acid	1	0	0	1	0.444	2.32
Imino di-3 propanoic acid	2	1	0	9	0.833	2.34
Phenyl malonic	2	0	0	4	0.444	2.37
OxaloAcetic	2	0	0	4	1.25	2.38
Glycine	1	1	0	4	1.5	2.39
Tricarballylic	3	0	0	9	1	2.4
Glycylaspartic	2	1	0	9	1.167	2.52
Ethylenebisoxyacetic	2	0	0	4	1	2.65
Tartaric	2	0	0	4	1.5	2.69
Asparagine	1	1	0	4	1.25	2.78
Malonic	2	0	0	4	1.333	2.91
Malic	2	0	0	4	1.25	2.93

Oxopropane 1,2 dicarboxylic	2	0	0	4	1	3.13
Aspartic	2	1	0	9	1.25	3.21
Tartronic	2	0	0	4	1.667	3.22
BenzylMalonic	2	0	0	4	0.4	3.24
TriGlycine	1	1	0	4	1.167	3.3
Adrenaline	0	1	0	1	0.444	3.59
Diglycolic	2	0	0	4	1.25	3.61
DOPA (2-amino-3,4 dihydroxyphenyl propanoic	1	1	0	4	0.556	3.9
Oxalic	2	0	0	4	2	4
Iminodiacetic acid (IDA)	2	1	0	9	1.25	4.81
Carboxymethoxymalonic	3	0	0	9	1.4	4.87
Citric	3	0	0	9	1.167	4.91
N-benzyl imino diacetic	2	1	0	9	0.455	5.1
EDMA	1	2	0	9	1	5.19
Ethylenediamine	0	2	0	4	1	2.41
Ditartronic	4	0	0	16	1.5	5.62
Oxybis(ethyleneiminoacetic	2	2	0	16	0.875	5.88
TMS (3 hydroxyoxapentane 1,2,4,5 tetracarboxylic)	4	0	0	16	1.25	6.3
rac Oxybisbutanedioic	4	0	0	16	1.125	6.58
meso Oxybisbutanedioic	4	0	0	16	1.125	7.56
cis 2,6 dicarboxy piperidine acetic acid	3	1	0	16	0.778	7.92
NTA	3	1	0	16	1.167	7.99
Diethyl triamine acetic acid	1	3	0	16	0.833	8.31
HEDTA	3	2	0	25	0.9	11.89
EDTA	4	2	0	36	1.667	13.92

CHAPTER 4

Examples of the Quantitative Structure-Property Relationships Applications for Predicting Hg(II), Cd(II), and Zn(II) Binding to Organic Molecules Similar in Ligand Group Composition to Natural Organic Matter

4.1 Heavy/Trace Metal Binding by Natural Organic Matter

Predicting M(II) (M= Hg, Cd, Zn) binding using the QSPR approach requires structural information about NOM. The complexity of NOM and the analytical problems it poses (Perdue & Ritchie 2003) preclude obtaining exact and complete structural knowledge of the NOM mixture, but hypothetical structures created to illustrate various properties can be used as an initial step towards predictive modeling. For example, Atalay and co-workers (Atalay *et al.* 2009) collected hypothetical structures from several authors and used these to estimate pK_a values and pH titration curves for comparison with experimental data.

In this example, Hg(II), Cd(II), and Zn(II) binding by NOM in a hypothetical dissolved organic matter solution are predicted. The case of Hg(II) complexation is complicated by the importance of thiol ligands, which several authors have suggested dominates complexation at environmental levels of mercury (Ravichandran 2004). Sulfur is a minor constituent of most NOM samples (typically $\leq$1% by weight), and only a portion of the sulfur is present as thiols (Xia *et al.* 1998). Thus, using a single molecule to represent NOM will lead either to uncharacteristically high (or low) sulfur content (in a molecule of MW 1000, for example, a single S atom would contribute >3% of the mass) or to uncharacteristically large MW. Keeping in mind that the chemical complexity of NOM is unrivaled (Perdue & Ritchie 2003), representing NOM as a mixture of hypothetical structures is therefore more reasonable.

Here, NOM is represented as a mixture of the five hypothetical structures collected by Atalay et al. (Atalay *et al*. 2009) with one addition. Since these five molecules, designated *a*-*e* by those authors, lack any thiol structure, a sixth molecule is added by replacing an alcohol group on molecule *e* with a thiol group to make molecule *f*. Table 1 shows the predicted log K'_{f} values for these structures, which range from ~3-14 for smaller structures containing only oxygen ligand atoms (*a, b,* and *e*) to ~12-22 for ligands containing oxygen and nitrogen ligands (*c* and *d*) to ~10-40 for the thiol-containing molecule *f*. It is important to note about calculating predicted log K_{HgL}' values that the calibration data set in our model limits the highest number of the carbon atoms which affect Hg(II) binding to 14 and the highest numbers of the amine and amide groups, alcohol groups, COOH groups, and thiol groups which affect Hg(II) binding to 5, 3, 4, and 1 respectively.

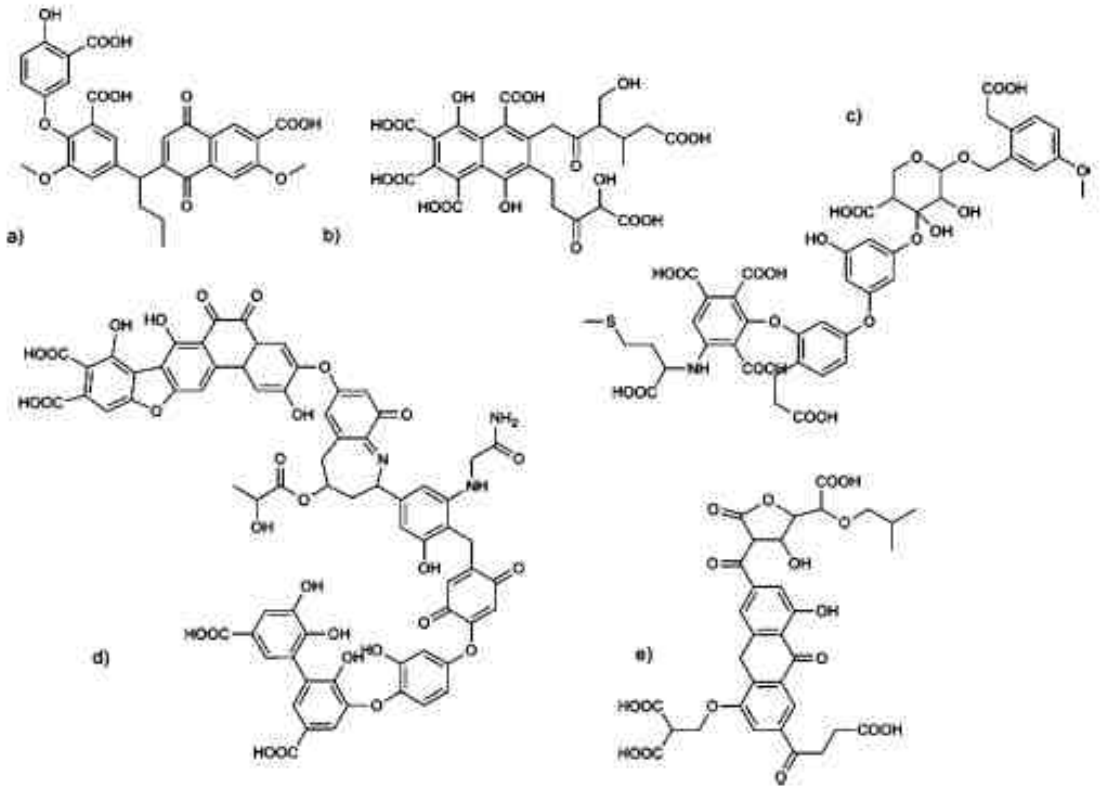


Figure 4.1 A hypothetical NOM mixture "prepared" by "mixing" 2 mol *a*, 2 mol *b*, 2 mol *c*, 2 mol *d*, 1 mol *e*, and 1 mol *f* (that is *e* where the alcohol group is replaced by a thiol group). Structures *a*-*e* are selected by Atalay et al., 2009.

Table 4.1 Hypothetical Ligand Structures.

Designation	Original source	Representing	Predicted log K_{ML}' (M = Hg, Cd, Zn)		
			Hg	Cd	Zn
a	Neiderer & Goss 2007	humic acid	10.23	2.86	3.67
b	Aiken *et al.* 1985	fulvic acid	14.06	10.67	13.75
c	Dialo *et al.* 2003	humic acid	16.99	11.81	13.65
d	Aiken *et al.* 1985	humic acid	21.62	13.34	13.56
e	Leenheer *et al.* 1995	fulvic acid	14.06	4.84	6.31
f	this work*	fulvic acid	39.74	10.33	10.75

*created by replacing an alcohol on structure *e* with a thiol

To simulate an NOM mixture, two copies of molecules *a-d* and one copy each of molecules *e* and *f* were combined. The resulting mixture is 57.% C, 36.4% O, 4.0% H, 1.2% N and 1.1% S by weight, within the range of elemental composition typically reported for NOM although slightly high in sulfur (Perdue & Ritchie 2003). The average molecular weight (866 amu) and fraction of aromatic carbons (30%) also fall within normal range.

To simulate Hg(II), Cd(II), and Zn(II) bindings by this mixture, we assumed a dissolved organic carbon concentration (DOC) of 5.0 mg C L^{-1}, a typical value for fresh waters. Concentrations of each ligand were normalized on a carbon basis as in Cabaniss (Cabaniss 2009) and adjusted to give the desired DOC; this resulted in individual molecule concentrations of 2.02 x 10^{-6} M for molecules *a-d* and 1.01 x 10^{-6} M for molecules *e* and *f*. The total concentration of Hg(II) varied from 0.10 to 5.1 x 10^{-6} M, so did the total

concentration of Cd(II) and that of Zn(II). Equilibrium speciation was calculated with Titrator (Cabaniss 1987) assuming pH 7.0 and ionic strength 0.10.

Combination with an Existing Model of NOM. A stochastic model for the synthesis and degradation of NOM, developed by Cabaniss et al. (Cabaniss *et al.* 2005), referred to as the agent-based model (ABM) represents NOM as a set of individual stochastically-reacting molecules of varying elemental and functional group composition. Ongoing work to incorporate sulfur chemistry in this model, combined with the QSPR developed in this work, will enable an *a priori* model for predicting Hg(II), Cd(II), and Zn(II) binding by NOM in natural waters. Sections 4.1.1-4.1.3 provide comparisons between experimental data and predictions based on the hypothetical NOM model discussed above. In one case, the ABM model of Cabaniss (2009) is modified by including the QSPR equations from chapters 2 and 3, but no thiol-containing molecules are simulated (ABM no thiols). In the other case, the same ABM is used (with updated QSPRs) but 10 molecules containing cysteine residues were added as precursors (ABM w/thiols).

4.1.1 Comparisons with Experimental Data for Hg(II) Binding by Natural Organic Matter

For ligand *f*, the predicted log K_{HgL}' = 39.74 is very close to the mean thiol binding constant of Hg(II) bound to IHSS Pahokee peat humic acid (log K = 38.5) reported by Khwaja et al. (Khwaja *et al.* 2006). Also, the changes in pHg as a function of increasing the loading of total Hg(II) in the hypothetical NOM mixture shown in Figure 4.1 follow the same pattern that Haitzer and co-workers (Haitzer *et al.* 2002) observed when they studied the relationship between the changes in the binding of Hg(II) to NOM as a function of Hg(II):carbon ratio of dissolved organic matter isolated from the Florida Everglades. They

concluded that where the amount of Hg(II) in the mixture is extremely small (below approximately 1 μg of Hg(II)/mg of dissolved organic matter), very strong interactions indicative of Hg(II)-thiol bonds are observed. That conclusion is verified in Figure 4.3 where loadings of Hg(II) below 1 μM result in pHg of 40 or higher and correspond to Hg(II) binding to molecule *f*.

Figure 4.5 shows the log of the moles of Hg bound per gram of dissolved organic matter calculated as a correlating function of increased presence of Hg^{2+} (in M) in a hypothetical solution of 5.0 mg/L DOC, pH 7, based on structures in Table 4.1, the thiol-missing and thiol-containing ABMs. It also shows the same correlation at the same pH based on the experimental data obtained by Haitzer and co-workers (Haitzer *et al.* 2002) as presented by Tipping (Tipping 2007). The root mean square error (RMSE) relative to the experimental results is 1.7, 3.9, and 0.6 log units for the hypothetical NOM model, the thiol-missing ABM, and the thiol-containing ABM respectively. The RMSE values show the importance of including thiols in any predictive model for Hg(II) binding by NOM. The second important source of disagreement with the experimental data is compositional differences between the NOM of the hypothetical NOM solution and the isolated hydrophobic fraction of dissolved organic matter used by Haitzer and co-workers (Haitzer *et al.* 2002). Some of the disagreement must also be attributed to the predictive error of the Hg(II) QSPR.

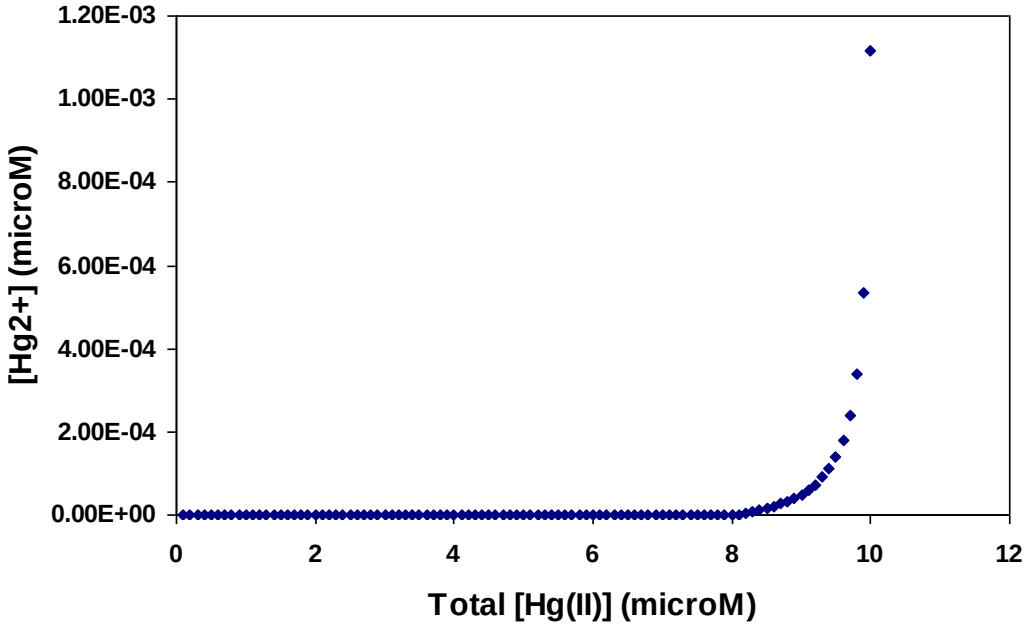


Figure 4.2 Hg^{2+} calculated as a function of increasing loading of Hg(II) in a hypothetical solution of 5.0 mg/L DOC, pH 7, based on structures in Table 4.1

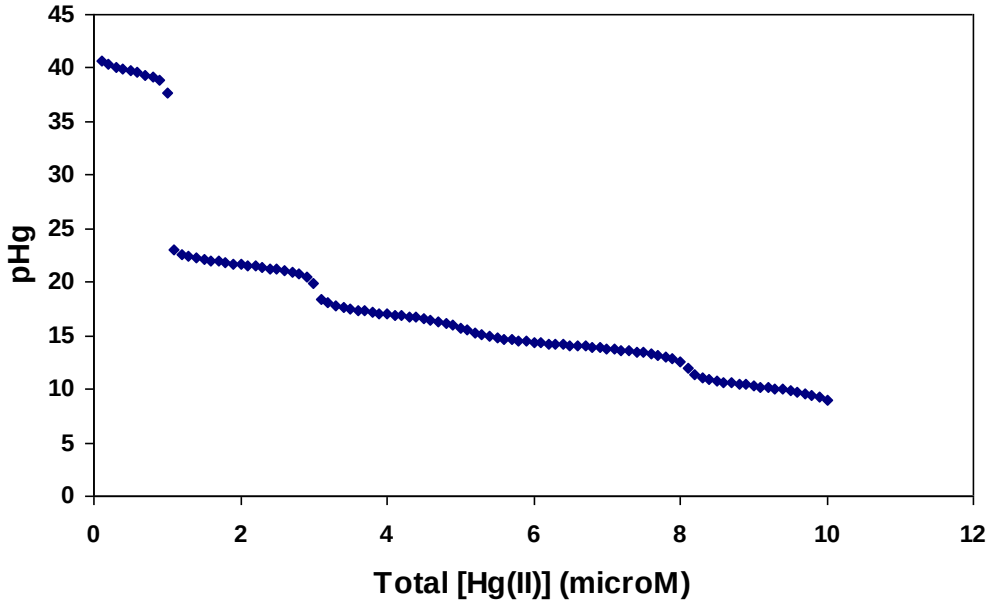


Figure 4.3 pHg calculated as a function of increasing loading of Hg(II) in a hypothetical solution of 5.0 mg/L DOC, pH 7, based on structures in Table 4.1

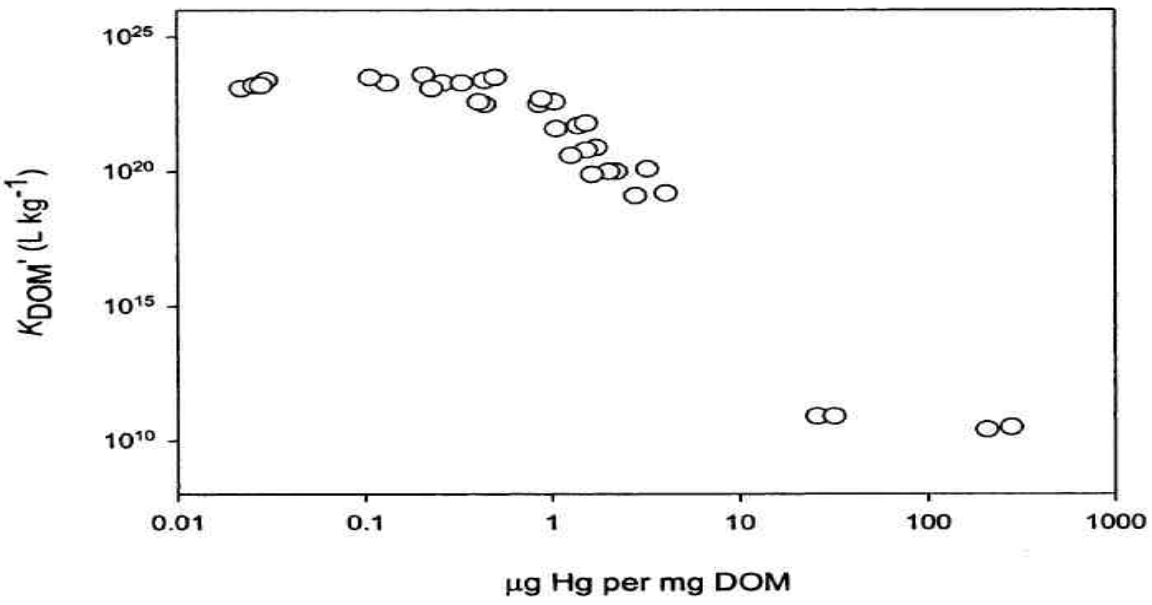


Figure 4.4 Data from Haitzer et al. 2002 for Hg(II) Binding by Dissolved Organic Matter

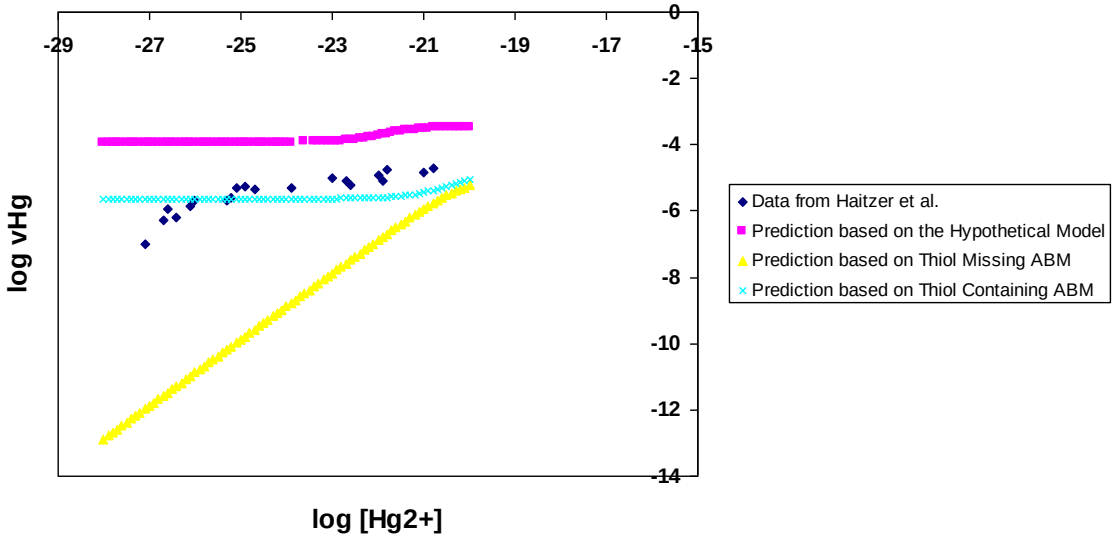


Figure 4.5 The log of the moles of Hg bound per gram of dissolved organic matter calculated as a correlating function of increased presence of Hg^{2+} (in M)

4.1.2 Comparisons with Experimental Data for Cd(II) Binding by Natural Organic Matter

The average of the six log K' values predicted for Cd(II) binding to molecules *a-f* is log K_{CdL}' = 9.0. This is consistent with the range of the conditional stability constants of stable cadmium organic complexes in fresh water reported by Xue and Sigg (Xue & Sigg 1998) (that is ~9 < log K < ~10). Also Figure 4.6 showing the concentration of free Cd^{2+} calculated as a function of increasing loading of Cd(II) in the hypothetical solution of 5.0 mg/L DOC at pH 7 closely resembles the plot of the activity of free Cd^{2+} ions, which are in equilibrium with Cd(II) adsorbed to the organic soil at pH 4.6, versus total Cd(II) concentration according to a model addressed by Karlsson and co-workers (Karlsson *et al.* 2007).

Figure 4.7 shows pCd calculated as a function of increasing loading of Cd(II) in a hypothetical solution of 5.0 mg/L DOC, pH 7, based on structures in Table 4.1, the thiol-missing and thiol-containing ABMs. It also shows pCd as a function of increasing loading of Cd(II) in a solution of purified peat humic acid (PPHA), pH 6, based on data from Benedetti and co-workers (Benedetti *et al.* 1995). The root mean square error (RMSE) relative to the experimental results is 5.8, 0.3, and 0.3 log units for the hypothetical NOM model, the thiol-missing ABM, and the thiol-containing ABM respectively. The RMSE values show that thiols are much less important to be included in a predictive model for Cd(II) binding by NOM than they are for Hg(II) binding by NOM. The most important source of disagreement with the experimental data is compositional differences between the NOM of the hypothetical NOM solution and the purified peat humic acid. Some of the disagreement must also be attributed to the predictive error of the Cd(II) QSPR.

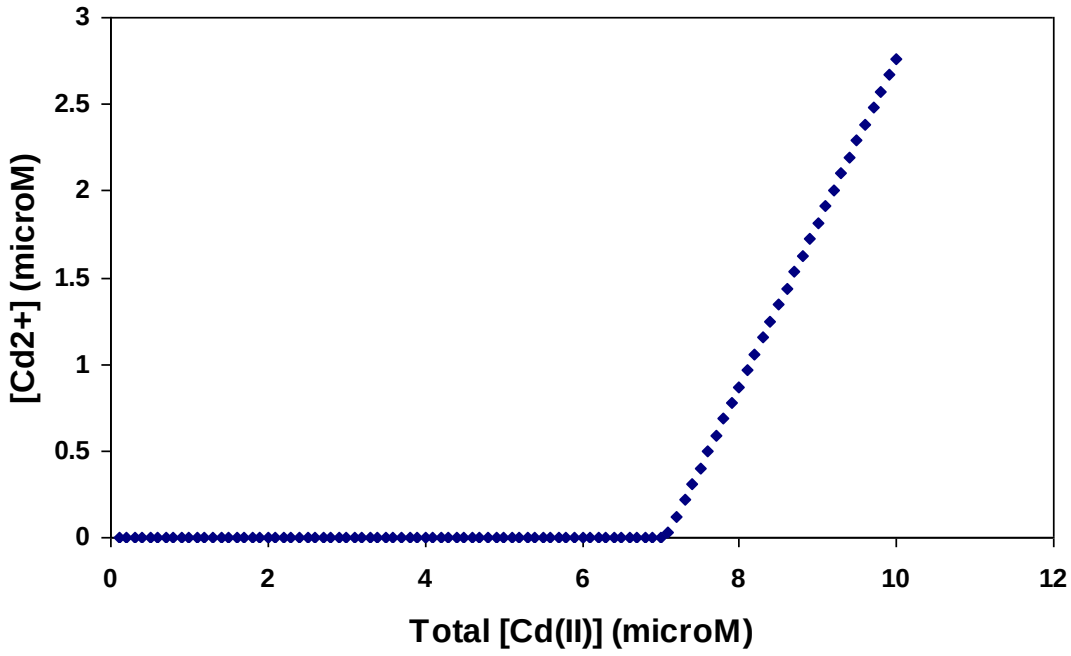


Figure 4.6 Cd^{2+} calculated as a function of increasing loading of Cd(II) in a hypothetical solution of 5.0 mg/L DOC, pH 7, based on structures in Table 4.1

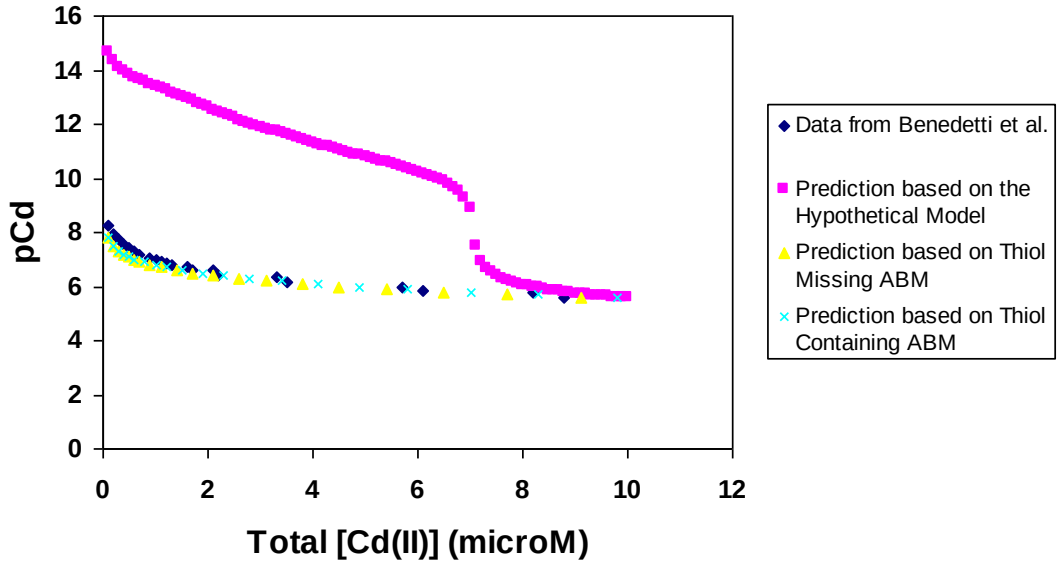


Figure 4.7 pCd as a function of increasing loading of Cd(II)

4.1.3 Comparisons with Experimental Data for Zn(II) Binding by Natural Organic Matter

The predicted log $K_{ZnL}^{'}$ = 3.67 for ligand *a*, and the predicted log $K_{ZnL}^{'}$ = 6.31 for ligand *e* are both within the range of the conditional binding constants of zinc by NOM, ~4 < log K < ~7, reported by Cheng and Allen (Cheng & Allen 2006). Also Figure 4.8, that shows the concentration of Zn^{2+} calculated as a function of increasing loading of Zn(II) in the hypothetical solution of 5.0 mg DOC at pH 7, partially (that is for total Zn(II) concentrations of 5 μM and higher) follows the same pattern as the Zn titration curve Cheng and Allen (Cheng & Allen 2006) show for zinc complexation by dissolved NOM from different surface waters of 10 mg/L DOC at pH 7.0.

Figures 4.9 and 4.10 show pZn calculated as a function of increasing loading of Zn(II) in a hypothetical solution of 5.0 mg/L DOC, pH 7, based on structures in Table 4.1, the thiol-missing and thiol-containing ABMs. They also show pZn as a function of increasing loading of Zn(II) in a solution of 10.0 mg/L DOC from the Edisto River in South Carolina and the Big Moose Lake in New York, pH 7, based on data from Cheng and Allen **(**Cheng & Allen 2006). The root mean square error (RMSE) relative to the experimental results from the Edisto River is 5.7, 0.1, and 0.2 log units for the hypothetical NOM model, the thiol-missing ABM, and the thiol-containing ABM respectively, and the root mean square error (RMSE) relative to the experimental results from the Big Moose Lake is 5.6, 0.1, and 0.1 log units for the hypothetical NOM model, the thiol-missing ABM, and the thiol-containing ABM respectively. The RMSE values show that thiols are much less important to be included in a predictive model for Zn(II) binding by NOM than they are for Hg(II) binding by NOM. The most important source of disagreement with the experimental data is

compositional differences between the NOM of the hypothetical NOM solution and the dissolved NOM of the Edisto River or the Big Moose Lake. Some of the disagreement must also be attributed to the predictive error of the Zn(II) QSPR.

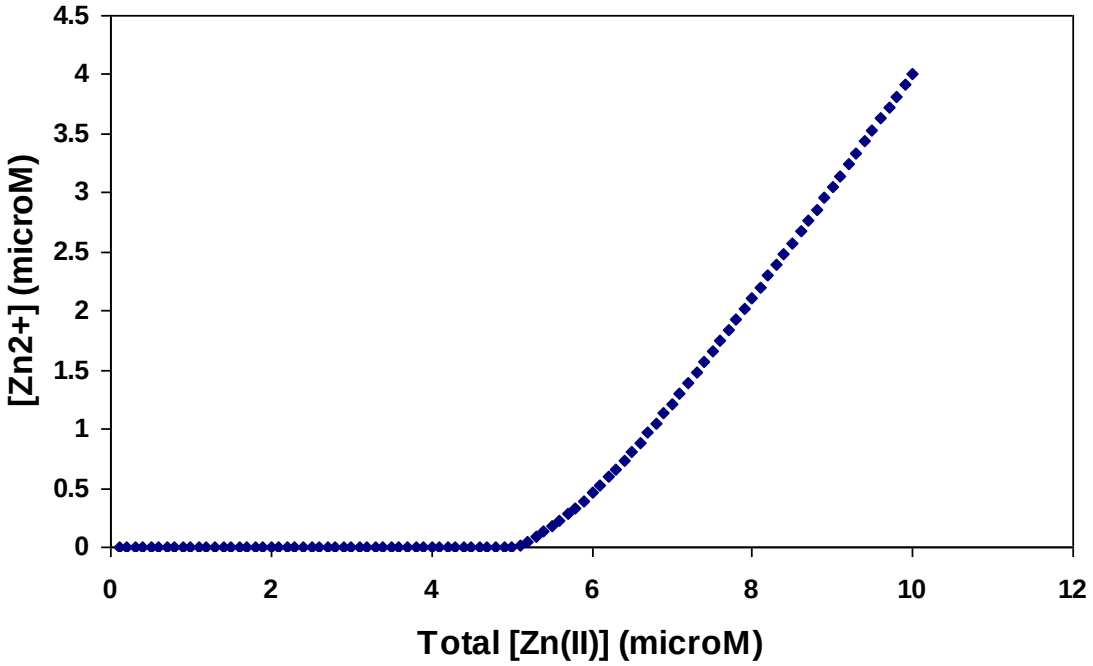


Figure 4.8 Zn^{2+} calculated as a function of increasing loading of Zn(II) in a hypothetical solution of 5.0 mg/L DOC, pH 7, based on structures in Table 4.1

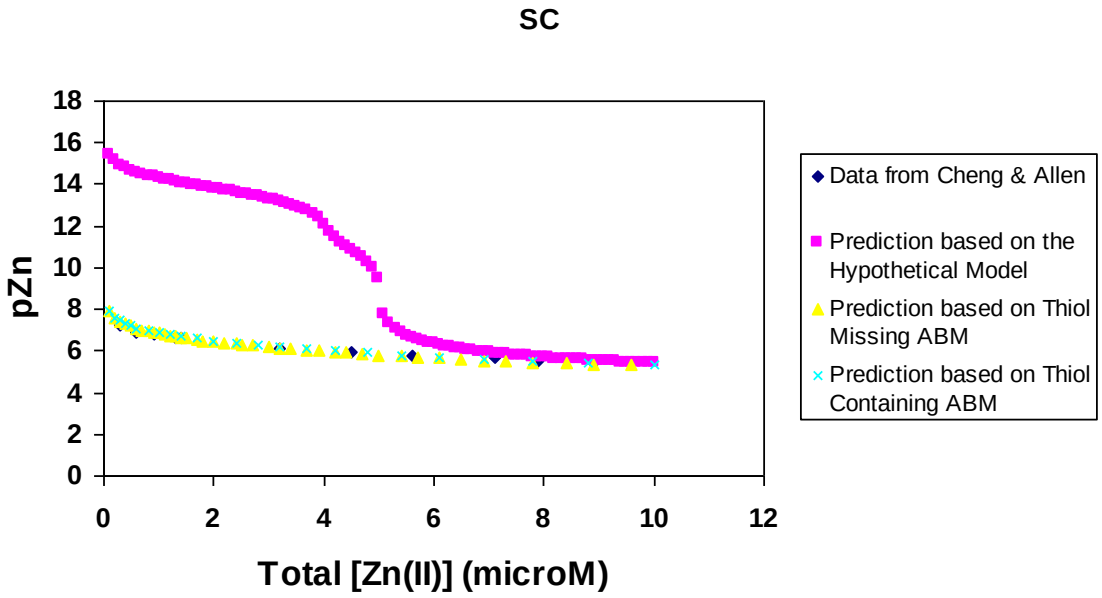


Figure 4.9 pZn as a function of increasing loading of Zn(II) compared to experimental data from the Edisto River in South Carolina (SC)

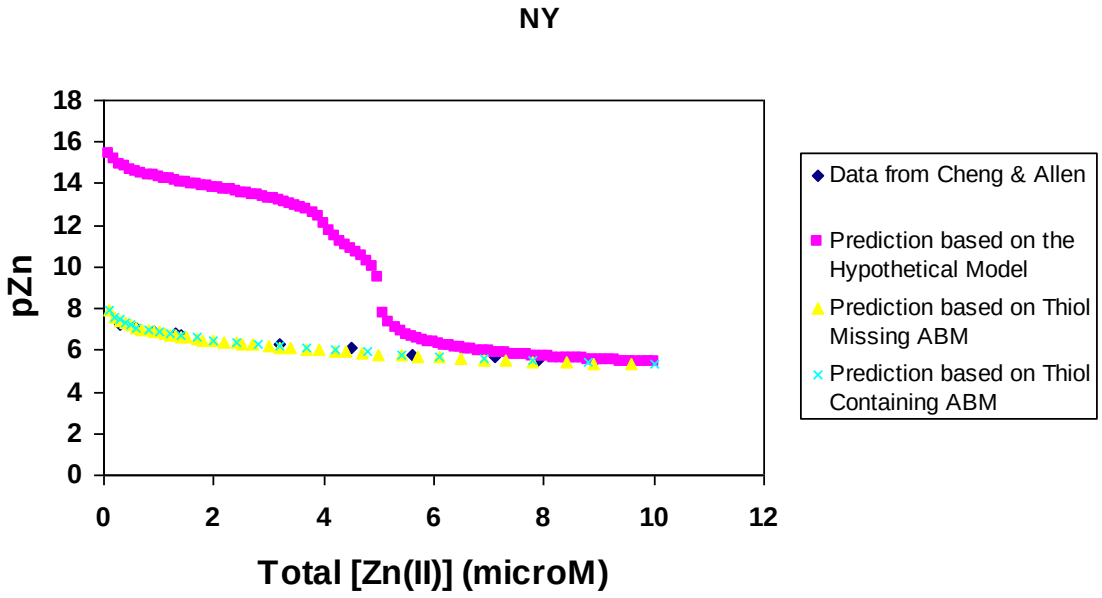


Figure 4.10 pZn as a function of increasing loading of Zn(II) compared to experimental data from the Big Moose Lake in New York (NY)

4.2 Synthesis of Molecules for Environmental Remediation or Biological Chelation of M(II) (M= Hg, Cd, Zn)

Considering the high toxicity of methylmercury, of the aquatic mercury cycle complexation with natural organic matter (NOM) is of special importance because it limits both inorganic mercury availability for methylation and methylmercury availability for bioaccumulation (Ravichandran 2004). The reason for these limitations becomes obvious when one considers that it is the binding of inorganic mercury to methyl groups that forms methylmercury (Greenwood & Earnshaw 1994) and considers that it is the affinity of methylmercury for fatty tissues in animals that causes the bioaccumulation of methylmercury (Ravichandran 2004): In some cases, mercury binds to NOM more strongly than it binds to the methyl group or to the fat. The importance of complexation by NOM in the toxicity of Cd(II) and Zn(II) were also discussed in the last chapter. The "detoxifying" nature of M(II) (M= Hg, Cd, Zn) complexation by NOM motivates us to try to design small molecules similar in ligand group composition to NOM and useful for the environmental remediation or biological chelation of M(II).

4.2.1 Synthesis of Molecules Used for the Remediation of M(II)-Contaminated Waters

Remediation is defined as the process of restoring a contaminated site to a condition where it is no longer a threat to human health and the environment. The typical goal of a remedial effort is to achieve a concentration for the contaminant that is below some legally defined maximum (Encyclopedia of Environmental Pollution and Cleanup 1999). A possible approach in heavy/trace metal remediation is to detoxify the metal in the water system by having it bound to an environmentally friendly substance so strongly that it can neither form a more dangerous substance nor bind to the receptor site of the organism. Therefore,

designing environmentally friendly chemical substances that behave as such to remediate heavy/trace metals in contaminated waters deserves great attention. For mercury, as an example, a number of ligands that precipitate the metal have been suggested, including sodium N,N-dimethyldithiocarbamate (DMDTC), sodium trithiocarbonate (STC), and sodium 1,3,5-triazine-2,4,6-trithiolate (TMT) (Hutchison *et al.* 2008).

One could synthesize small molecules similar in ligand group composition to NOM and determine their binding strength to M(II) after the synthesis. However, costliness and consumption of time makes this approach highly inefficient. Therefore, synthetic chemists need to estimate how strongly the hypothetical M(II) chelates would bind to M(II) before they synthesize them. In other words, it is desirable to estimate the would-be bindings without the benefit of experimental data. What is most preferred is a simple mathematical relationship which allows predicting the bindings robustly using minimal structural information about the hypothetical molecules. Below we demonstrate the applicability of the QSPR which we have developed for Cd(II) for estimating the conditional stability constant of the 1:1 complex Cd(II) would form with 2-heptanethiol, an odorant in red bell pepper extracts identified and synthesized by Simian and co-workers (Simian *et al.* 2004).

According to the cadmium QSPR developed in Chapter 3,

$\text{Log } K_{CdL}' = (1.67 \text{ x \# Amine}) + (2.91 \text{ x \# Thiol}) + (0.278 \text{ x } LN^2) + (69.7 \text{ x } Z_{dens}) + (-148. \text{ x AA})$.

For 2-heptanethiol (MW = 132.2694), # Amine = 0, # Thiol = 1, $LN^2 = 1$, $Z_{dens} = 0.00756$, and AA = 0. Therefore, $\log K_{CdL}' = (1.67 \text{ x } 0) + (2.91 \text{ x } 1) + (0.278 \text{ x } 1) + (69.7 \text{ x } 0.00756) + (-148. \text{ x } 0) = 3.715$.

4.2.2 Synthesis of Molecules Used for Chelation Therapy

A body of experimental evidence has indicated that chelation may be important in the pharmacological action of many drugs. The use of chelating agents to remove certain toxic cations such as lead and plutonium from the body has been widely recognized in medical practice (McGraw-Hill Encyclopedia of Science & Technology 2007). Some chelating agents considered to be useful for treatment against metal poisoning are British Anti-Lewisite (BAL), calcium disodium ethylenediaminetetraacetic acid ($CaNa_2EDTA$), meso 2,3-dimercaptosuccinic acid (DMSA), and sodium 2,3-dimercaptopropane-1- sulfonate (DMPS) (Flora *et al.* 2008). However, a very important issue in the use of a chelating agent for a therapeutic purpose is any side effect it might have. Discovering or developing effective natural or synthetic chelating agents for the removal of toxic quantities of heavy/trace metal deposits in the body has been subject to research (Cai *et al.* 2005; McGraw-Hill Encyclopedia of Science and Technology 2007).

A possibly useful research area in this matter is designing small molecules similar in ligand group composition to NOM and useful for the biological chelation of cations like CH_3Hg^+, Hg(II), Cd(II), Zn(II), etc. Considering the importance of the metal-chelate versus metal-biotic ligand binding strength, in such synthesis one could synthesize the chelates and determine their binding strength to these cations after the synthesis. However, due to the same reasons addressed in 4.2.1, this approach is highly inefficient, and consequently a simple mathematical relationship which allows predicting the bindings robustly using minimal structural information about the hypothetical molecules is most preferred. Below we demonstrate the applicability of the QSPR which we have developed for Zn(II) for estimating the conditional stability constant of the 1:1 complex Zn(II) would form with

diethylenetriaminepentaacetic acid (DTPA), a chelating agent that increases the urinary excretion of zinc and inhibits its lethal effect (Sato & Tsuda 2008).

According to the zinc QSPR developed in Chapter 3,

$\log K_{ZnL}' = (1.14 \times \#\text{ Thiol}) + (0.363 \times LN^2) + (0.993 \times \text{Het:C})$

and for DTPA, $[(HOOCCH_2)_2NC_2H_4]_2NCH_2COOH$, # Thiol = 0, $LN^2 = 36$ (the possible maximum), and Het:C = 1.3. Therefre, $\log K_{ZnL}' = (1.14 \times 0) + (0.363 \times 36) + (0.993 \times 1.3) = 14.359$.

Future Work- Development of a QSPR for Predicting Monomethylmercury Complexation by Molecules Similar in Ligand Group Composition to NOM.

Considering that complexation with natural organic matter (NOM) limits methylmercury availability for bioaccumulation (Ravichandran 2004), synthesizing molecules similar in ligand group composition to NOM for the purpose of binding to monomethylmercury (CH_3Hg^+) is important from the viewpoint of chelation therapy. Therefore, developing a QSPR that allows the prediction of CH_3Hg^+ binding to hypothetical chelates is greatly helpful.

4.3 References

Atalay, Y. B., Carbonaro, R. F. & Di Toro, D. M. 2009 Distribution of proton dissociation constants for model humic and fulvic acid molecules. *Environ. Sci. Technol.* **43**(10), 3626-3631.

Benedetti, M. F., Milne, C. J., Kinniburgh, D. G., Van Riemsdijk, W. H. & Koopal, L. K. 1995 Metal ion binding to humic substances: Application of the non-ideal competitive adsorption model. *Environ. Sci. Technol.* **29**(2), 446-457.

Cabaniss, S. E. 1987 Titrator: An interactive program for aquatic equilibrium calculations. *Environ. Sci. Technol.* **21**(2), 209-210.

Cabaniss, S. E. 2009 Forward modeling of metal complexation by NOM: I. *A priori* prediction of conditional constants and speciation. *Environ. Sci. Technol.* **43**(8), 2838-2844.

Cabaniss, S. E., Madey, G., Leff, L., Maurice, P. A. & Wetzel, R. 2005 A stochastic model for the synthesis and degradation of natural organic matter. Part I. Data structures and reaction kinetics. *Biogeochemistry.* **76**(2), 319-347.

Cai, L., Li, X.; Song, Y. & Cherian, M. G. 2005 Essentiality, toxicology and chelation therapy of zinc and copper. *Curr. Med. Chem.* **12**(23), 2753-2763.

Cheng, T. & Allen, H. E. 2006 Comparison of zinc complexation properties of dissolved natural organic matter from different surface waters. *J. Environ. Manage.* **80**(3), 222-229.

Diallo, M. S., Simpson, A., Gassman, P., Faulon, J. L., Johnson, J. H., Goddard, W. A. & Hatcher, P. G. 2003 3-D structural modeling of humic acids through experimental characterization, computer assisted structure elucidation and atomistic simulations. 1. Chelsea soil humic acid. *Environ. Sci. Technol.* **37**(9), 1783-1793.

Encyclopedia of Environmental Pollution and Cleanup 1999 Vol. 2 John Wiley & Sons, Inc., New York, NY, p. 1681.

Flora, S. J. S., Mittal, M. & Mehta, A. 2008 Heavy metal induced oxidative stress & its possible reversal by chelation therapy. *Indian J. Med. Res.* **128**(4), 501-523.

Greenwood, N. N. & Earnshaw, A. 1994 *Chemistry of the Elements*. Pergamon Press, Tarrytown, NY, pp 1395-1422.

Haitzer, M., Aiken, G. R. & Ryan, J. N. 2002 Binding of mercury(II) to dissolved organic matter: The role of the mercury-to-DOM concentration ratio. *Environ. Sci. Technol.* **36**(16), 3564-3570.

Humic Substances in Soil, Sediment, and Water: Geochemistry, Isolation, and Characterization (eds. G. R. Aiken, D. M. McKnight, R. L. Wershaw and P. MacCarthy) 1985. Wiley, New York, NY.

Hutchison, A., Atwood, D. & Santilliann-Jiminez, Q. E. 2008 The removal of mercury from water by open chain ligands containing multiple sulfurs. *J. Hazard Mater* **156**(1-3), 458-465.

Khwaja, A. R., Bloom, P. R. & Brezonik, P. L. 2006 Binding constants of divalent Mercury (Hg^{2+}) in soil humic acids and soil organic matter. *Environ. Sci. Technol.* **40**(3), 844-849.

Karlsson, T., Elgh-Dalgren, K., Björn, E. & Skyllberg, U. 2007 Complexation of cadmium to sulfur and oxygen functional groups in an organic soil. *Geochim. Cosmochim. Acta* **71**, 604-614.

Leenheer, J. A., McKnight, D. M., Thurman, E. M. & MacCarthy, P. 1995 Structural components and proposed structural models of fulvic acid from the Suwannee River. *US Geol. Surv. Water-Supply Pap.* 195-211.

McGraw-Hill Encyclopedia of Science and Technology 2007 Vol. 4 McGraw-Hill, New York, NY, p. 10.

Niederer, C. & Goss, K. U. 2007 Quantum chemical modeling of humic acid/air equilibrium partitioning of organic vapors. *Environ. Sci. Technol.* **41**(10), 3646-3652.

NIST Critically Selected Stability Constants of Metal Complexes Database, 2004 Version 8.0, US Dept. of Commerce, Gaithersburg, MD.

Perdue, E. M.; Ritchie, J. D. Dissolved organic matter in freshwaters. In *Surface and Ground Water, Weathering, and Soils* (ed. J.I. Drever) Vol. 5 *Treatise on Geochemistry* (eds. H.D. Holland and K.K. Turekian), Elsevier-Pergamon, Oxford, **2003**; pp 273-318.

Ravichandran, M. 2004 Interactions between mercury and dissolved organic matter-a review. *Chemosphere* **55**(3), 319-331.

Simian, H.; Robert F. & Blank, I. 2004 Identification and synthesis of 2-heptanethiol, a new flavor compound found in bell peppers. *J. Agric. Food Chem.* **52**(2), 306-310.

Sato, I. & Tsuda, S. 2008 Removal of ^{65}Zn from mouse body by isotopic dilution and by DTPA chelation. *J. Vet. Med. Sci.* **70**(3), 213-216.

Tipping E. 2007 Modeling the interactions of Hg(II) and methylmercury with humic substances using WHAM/Model VI. *Appl. Geochem.* **22**(8), 1624-1635.

Xia, K., Weesner, F., Bleam, W. F., Bloom, P. R., Skyllberg, U. L. & Helmke, P. A. 1998 XANES studies of oxidation states of sulfur in aquatic and soil humic substances. *Soil Sci. Soc. Am. J.* **62**(5), 1240-1246.

Xue, H. B. & Sigg, L. 1998 Cadmium speciation and complexation by natural organic ligands in fresh water. *Anal. Chim. Acta* **363**(2-3), 249-259.

CHAPTER 5

Summary

Group IIB metals in the environment are all of toxicological concern and consequently of environmental concern leading to the importance of their geochemistry. The pollution of natural waters by any of the three elements of mercury, cadmium, and zinc is most serious because natural waters are sources of life in nature. The toxicological and environmental concerns associated with Group IIB metals increase as we go from zinc to cadmium to mercury.

While all forms of mercury (Hg^0, Hg^{2+}, and mercury organometallic compounds) are variously toxic to aquatic biota and humans, monomethylmercury and dimethylmercury species are of special environmental importance. Monomethylmercury, to which people are exposed almost entirely through the consumption of contaminated fish products, fish, and wildlife that are at the top of aquatic food chains, is the principal mercury related human health concern of today because of its high neurotoxicity. Fish consumption being the dominant pathway of exposure to methylmercury for most human populations has given rise to a variety of environmental preventive regulations in the world. However, monomethylmercury in aquatic environments is produced through the methylation of Hg^{2+} predominantly through the action of aquatic microorganisms (primarily sulfate-reducing bacteria).

There are a number of chemical processes that govern the fate of Hg^{2+} and affect its methylation in natural waters. Hg^{2+} methylation is enhanced with deceasing pH. On the other hand, reverse processes such as demethylation and Hg^{2+} reduction lead to decrease in monomenthylmercury concentration. Also Hg^{2+} precipitation by hydrogen sulfide in natural

waters is believed to inhibit the availability of mercury for methylation. Another important chemical process that limits the availability of mercury for methylation is the binding of Hg(II) to natural organic matter (NOM). These associations are exceptionally strong, and, as NOM is ubiquitous in aquatic environments, knowing binding constants for Hg(II)-NOM complexes is of special importance.

Cadmium is moderately toxic to all organisms. It is a cumulative poison in mammals, and prolonged intake of it leads to dysfunction of the kidneys of humans and, if the exposure level is high, may cause skeletal damage. The first report of such skeletal damage is from Japan where itai-itai disease was caused by exposure to cadmium in cadmium-contaminated water. A major factor affecting cadmium concentration in natural waters is cadmium sorption on particulate matter and bottom sediments. Except in cases where the concentration of suspended particulate matter is high, the speciation of cadmium is generally considered to be dominated by dissolved forms. However, the free ion form of the metal is thought to be the most available and toxic. Therefore, considering the abundance of dissolved NOM in natural waters, Cd(II) complexation by NOM is considerably important from an environmental point of view.

Zinc is an essential element and micronutrient for normal growth by plants and animals; however, it can be detrimental to organisms at both high and low concentrations. High concentrations of zinc that are known to be toxic or even lethal to organisms have been observed in natural waters. Some of zinc species most harmful to aquatic life under certain environmental conditions are zinc aqua ions. Similar to Hg(II) and Cd(II), complexation with organic ligands in aquatic systems may occur significantly for Zn(II) and leaves only a small fraction of zinc as aqua ions. Therefore, in order to evaluate zinc toxicity in natural

waters, thermodynamic constants of zinc complexation by NOM in natural waters are of interest.

As important as equilibrium constants for Hg(II), Cd(II), and Zn(II) bindings to NOM in natural waters are, and despite the availability of several experimental methods for determining these constants, experimental determination of the constants is costly and time consuming. It is therefore desirable to estimate those equilibrium constants without the benefit of additional experimental data. This work uses QSPRs (Quantitative Structure-Property Relationships) to predict binding constants from hypothetical structures of NOM molecules. A QSPR useful for predicting Hg(II) binding equilibrium constants has been developed for the first time, to our knowledge, and two QSPRs useful for predicting Cd(II) and Zn(II) binding equilibrium constants have been improved.

The QSPRs are calibrated using literature stability constants, that were converted to conditional stability constants (at pH =7.0) for environmental advantages, for a variety of organic compounds, similar in ligand group composition to NOM. This variety included 44 compounds for Hg(II), 63 compounds for Cd(II), and 68 compounds for Zn(II). Most of the compounds contained some or all of carboxylate, amine, and thiol ligand groups. Each QSPR was calibrated by linear regression using all ligand molecules. The descriptor variables were numbers of certain individual ligand groups (for example carboxylic acids, amines, and thiols), size-dependent variables, or certain combinations of them. Calibration of the QSPRs gave a standard error of prediction (S_{pred}) of 1.60 log units and adjusted $R^2 = 0.965$ for Hg(II), an S_{pred} of 0.935 log units and adjusted $R^2 = 0.931$ for Cd(II), and an S_{pred} of 0.984 log units and adjusted $R^2 = 0.934$ for Zn(II).

The special standard of the earlier QSPRs for Cd(II) and Zn(II) (that did not include any thiol-containing molecule in their calibration data sets) developed by Cabaniss (2008), of an S_{pred} of ~1.0 log unit, was maintained in the Cd(II) and Zn(II) QSPRs presented in this work, and they describe 93.1% and 93.4% of the variance in the data for Cd(II) and Zn(II) respectively. In the case of the Hg(II) QSPR developed in this work, although the adjusted $R^2 = 0.965$ is higher than those of Cd(II) and Zn(II) QSPRs, the $S_{pred} = 1.60$ log units is also greater than ~1.0. However, this S_{pred} is acceptable considering two facts: First of all the range of experimental data used in the Hg(II) QSPR (from ~5 to ~35, that is ~30 orders of magnitude) is nearly double that of Cd(II) and Zn(II). Also the individual data points are less reliable for Hg(II) due to analytical difficulties.

Sulfur is typically ≤1% by weight of NOM, and only a portion of the sulfur is present as thiols. Still there is evidence for Hg(II), Cd(II), and Zn(II) binding to sulfur in NOM. Further, all the three metals interact with sulfur in biological systems. However, the affinity of thiols for Hg(II) is so high that there is significant inconsistency in thiol equilibrium binding constants reported for Hg(II) in different literature using different analytical methods for the determination of the constants. The high chemical significance of thiols for Hg(II), representing the exceptionally high affinity of thiols for Hg(II) (that is log K_f ~30-40), is clearly reflected in the positively large Hg(II) QSPR thiol coefficient.

The developed QSPRs have various potential applications examples of which are presented. The most obvious application is predicting M(II) (M = Hg, Cd, Zn) binding to hypothetical NOM solutions. All three QSPRs developed were applied on the components of a hypothetical NOM mixture with a dissolved organic carbon (DOC) concentration of 5 mg/L at pH = 7.0. Two other possible applications of the developed QSPRs were addressed

both of which are in designing small M(II)-binding molecules similar in ligand group composition to NOM for environmental remediation or chelation therapy. In designing such molecules, it is desirable to estimate the binding of the would-be synthesized compound without the benefit of experimental data. The developed QSPRs can be applied for a prediction as such. Two examples are given: 1- The Cd(II) QSPR is used for predicting Cd(II) binding by 2-heptanethiol, an odorant in red bell pepper extracts, as a potential molecule for Cd(II) environmental remediation. 2- The Zn(II) QSPR is used for predicting Zn(II) binding by diethylenetriaminepentaacetic acid (DTPA), a chelating agent that increases the urinary excretion of zinc and inhibits its lethal effect.

APPENDIX 1

Additional Details of the QSPRs Development

A1.1 The Step-by-Step Process for Developing a QSPR for M(II) (M = Hg, Cd, Zn)

For each QSPR developed and accepted in this dissertation, the following step-by-step process was followed:

1- Using NIST Standard Reference Database 46 Version 8.0 (NIST 2004), a modeling-sufficient number (40 being the minimum acceptable number) of ligand molecules for which 1:1 stability constants for the cation M(II) (M = Hg, Cd, Zn) were available at ionic strength 0.1 and 25 ^{0}C or under similar conditions (if no stability constant was available at I = 0.1 and T = 25 ^{0}C) were selected such that one or more of the functional groups of alcohol, carboxylic acid, amine, thiother, or one thiol was present on each molecule. Also each of these functional groups was included in the data set with the requirement of appearing on at least three molecules (for example the Hg(II) data set included five thiol-containing molecules).

2- For M(II)-L complexes of which values of log K_f > ~15 were reported in the NIST database (NIST 2004), the references were checked for the analytical methods used to determine the stability constants, and a log K_f value was used only if it was determined accurately enough (see Chapter 2 for details on Hg(II)-thiol complexes and pH titrations). 3- The values of K_f in the finally chosen log K_f values are the equilibrium constants of the reactions which have the general equation of $M^{2+} + L^{n-} \rightleftarrows ML^{2-n}$ (M = Hg, Cd, Zn). In other words, $K_f = K_{ML} = [ML^{2-n}] / [M^{2+}][L^{n-}]$. To facilitate integration with models of NOM and to make predictions at neutral pH, it is useful to know the pH-adjusted conditional constant.

For each organic molecule in the data set, therefore, K_{ML} is used to calculate a corresponding conditional equilibrium constant, $K_{ML}' = [ML^{2-n}] / [M^{2+}][L']$, where $[L']$ is the sum of the concentrations of all ligand species which are not complexed by M(II). Combining K_{ML} and K_{ML}' equations may relate the two constants to each other as $K_{ML}' = K_{ML} \cdot [L^{n-}]/[L'] = K_{ML}$ α_n, where α_n is the fraction of the uncomplexed ligand which exists in the $-n$ charge state (see Chapters 2 and 3). The use of the word "conditional" in naming K_{ML}' reflects the fact that it varies with pH, ionic strength, and temperature. In this study, pH 7.0 is chosen, but as α_n can be calculated for other pH values, so can K_{ML}'. This enables us to use the model for any environmental pH of interest provided acid pKa's are known.

4- Several structure-property-related variables were chosen as candidate descriptors based on the possibility of being quantitatively contributing to the log K_f' values of the finalized set of ligand molecules.

5- On an Excel spreadsheet, the full set of the candidate descriptors was subjected to multiple linear regression as the set of the independent variables describing the log K_f' values as the dependent variables.

6- The candidate for which the p value was the highest was eliminated if the p value was $\geq$ 0.05, and the regression was then repeated with the reduced set of descriptors until all variables had satisfactory p values ($p < 0.05$) as a result of this stepwise elimination.

7- The final QSPRs were accepted only if a) the standard error of the prediction (S_{pred}) was $S_{pred} < 2.0$ for the Hg(II) QSPR and $S_{pred} \sim 1.0$ for the Cd(II) and Zn(II) QSPRs and b) the adjusted R^2 of the prediction was > 0.95 for the Hg(II) QSPR and > 0.90 for the Cd(II) and Zn(II) QSPRs.

8- Every QSPR that met the S_{pred} and adjusted R^2 requirements was subjected to an examination of residuals. This examination was useful in two ways. First, outlier data points were identified and were, in some extreme cases, studied as for the analytical methods which had led to their presentation. Second, the randomness of the pattern of the residual scatter was checked to make sure that the pattern was random (meaning that the descriptor variables did not miss any repeatedly effective chemical factor).

A1.2 Which log K_f Values Were Rejected and Why?

The log K_f values which were accepted in the NIST database (NIST 2004) and rejected in this dissertation were all determined using indirect pH titrations. For most rejection cases the values were rejected because the measured pH is relatively insensitive to differences in log K_{ML} (M = Hg, Cd, Zn) above ~15. The details are discussed in Chapter 2. However, for a couple of rejection cases, being log K_f values for ZnL and CdL where L is the fully deprotonated form of glutathione, the values were rejected for a different reason. Krężel and Bal (Krężel & Bal 2004) performed studies of zinc(II)-glutathione binary complexes coordination equilibria. They showed the proposed structures of the complexes, and the structure of ZnL where L is the fully deprotonated form of glutathione was not proposed (Krężel & Bal 2004). With this evidence existing, glutathione was excluded from the QSPR calibration datasets of both Zn(II) and Cd(II) as zinc and cadmium are rather similar (Greenwood & Earnshaw 1998). Table A1.1 shows the rejected stability constants and the ligand molecules associated with them for Hg(II), Cd(II), and Zn(II).

Table A1.1 Rejected Stability Constants and the Ligand Molecules Associated with them for Hg(II), Cd(II), and Zn(II).

Ligand Molecule	log K_{HgL}	log K_{CdL}	log K_{ZnL}
D-Penicillamine	18.8		
L-Cysteine	14.4		
Glutathione	26.0	10.18	7.98
N-(2-Mercaptoethyl)iminodiacetic acid		16.72	15.92

A1.3 The Sets of All Candidate Descriptors and the Sets of Rejected Descriptors

A1.3.1 Hg(II)

The initial candidate descriptors for the Hg(II) QSPR were constitutional variables, including molecular weight, numbers of sulfur atoms, hydrogen atoms, carbon atoms, oxygen atoms, nitrogen atoms, thiol groups, hydroxyl groups, carboxylic acid groups, primary amine groups, secondary amine groups, tertiary amine groups, thioether sulfur atoms, and many of their numerous combinations (for example the product of the number of nitrogen atoms and molecular weight). However, the descriptor variables were gradually limited to the final Hg(II) QSPR variables addressed in Chapter 2 (see A1.1).

A1.3.2 Cd(II)

The initial candidate descriptors for the Cd(II) QSPR were all the final descriptors Cabaniss (Cabaniss 2008) had used for his Cd(II) QSPR; however, the presence of thiol was also taken into account. Therefore, the full set of the initial candidate descriptors for the Cd(II) QSPR was the set of # Amine, # Thiol, # Ether, LN^2, Z_{dens}, and AA. However, # Ether was eliminated later (see A1.1).

A1.3.3 Zn(II)

The initial candidate descriptors for the Zn(II) QSPR were some of the final descriptors Cabaniss (Cabaniss 2008) had used for his Zn(II) QSPR; however, the presence

of thiol was also taken into account. Therefore, the full set of the initial candidate descriptors for the Zn(II) QSPR was the set of # COOH, # Amine, # Thiol, LN^2, and Het:C. However, # COOH and # Amine were eliminated later (see A1.1).

A1.4 References

NIST Critically Selected Stability Constants of Metal Complexes Database, 2004 Version 8.0, US Dept. of Commerce, Gaithersburg, MD.

Krężel, A. & Bal, W. 2004 Studies of zinc(II) and nickel(II) complexes of GSH, GSSG and their analogs shed more light on their biological relevance. *Bioinorg. Chem. Appl.* **2**(3-4), 293-305.

Greenwood, N. N. & Earnshaw, A. 1994 *Chemistry of the Elements.* Pergamon Press, Tarrytown, NY.

Cabaniss, S. E. 2008 Quantitative structure-property relationships for predicting metal binding by organic ligands. *Environ. Sci. Technol.* **42**(14), 5210-5216.

APPENDIX 2

Geometry of Zn(II) Complexes with Dissolved Organic Matter: X-ray Studies at Variable pH

This appendix is the main draft of a manuscript titled *Geometry of Zn(II) Complexes with Dissolved Organic Matter: X-ray Studies at Variable pH* in preparation by David Tierney, Aliyar Mousavi, and Stephen E. Cabaniss for submission to **Aquatic Geochemistry** for publication.

A2.1 Introduction

Zn(II) is both an essential micronutrient and a potential toxicant in both marine and fresh waters (Cheng & Allen 2006). Because the aquo ion, Zn^{2+}, is believed to be the principal bioavailable form of dissolved Zn(II), zinc speciation in natural waters has been studied using a variety of methods (van den Berg 1984; Jansen *et al.* 1998; Christensen & Christensen 1999; Meylan *et al.* 2004; Jakuba *et al.* 2009). The fraction of total Zn(II) present as the aquo ion varies as a function of water chemistry and environment.

Most dissolved Zn(II) is complexed by dissolved organic matter (DOM), a ubiquitous and heterogeneous mixture of ligands derived from natural products (see Chapter 3). However, experimental measurements of DOM complexation at environmental levels of dissolved Zn(II) are complicated by the heterogenous nature of the organic ligands and are time-consuming and expensive. Consequently, predictive models of metal speciation are desirable, but must be carefully formulated and calibrated for accuracy.

Thermodynamic models like WHAM and NICA-Donnan have been extensively calibrated using metal-DOM complexation data for Zn(II) and other metals, and are able to reproduce calibration data on humic and fulvic isolates at variable pH and ionic strength

(Tipping 2005; Koopal *et al.* 2006). However, they have been less uniformly successful at predicting Zn(II) complexation by DOM not included in the calibration, perhaps due to differing isolation methods or to variations in DOM composition between ecosystems (Christensen & Christensen 1999; Cheng & Allen 2006; Balistrieri *et al.* 2008).

An alternative, forward modeling approach accounts for molecular heterogeneity and variations between ecosystems by using agent-based simulations of the evolution of DOM in the environment (Cabaniss *et al.* 2005, 2007). However, unlike the thermodynamic models above, this approach requires assumptions about site structure which will affect its accuracy. The quantitative structure-property relationship (QSPR) used to predict zinc binding constants assumed an octahedral geometry for complexes with O- and N-ligand groups, and was able to describe both small molecule and DOM binding of Zn(II) (Cabaniss 2008, 2009).

Zn(II) is well-known to form tetrahedral complexes as well as octahedral, depending on ligand structure (Greenwood and Earnshaw 1994; Cotton *et al.* 1999), and both structures have been reported for Zn(II)-DOM complexes. Xia and co-workers (Xia *et al.* 1997) used XANES and EXAFS to examine the Zn(II) and other metal complexes with aquatic DOM and soil extracts at pH 4.0. They concluded that although the soil and aquatic extracts differed in ligand composition, all Zn(II) complexes had octahedral geometry. More recently, however, Ramalho and co-workers (Ramalho *et al.* 2007) published density functional computations which indicated that tetrahedral geometry would be favored for a model DOM molecule. That same year, Karlsson and Skyllberg (Karlsson & Skyllberg 2007) examined Zn(II) complexes with organic soil material at pH 5.6-7.3 using EXAFS and concluded that four-coordinate (tetrahedral) geometry predominated.

In light of the inconsistencies between these studies and the importance of the binding geometry in the structure-based modeling, it seems appropriate to ask whether the geometry of Zn(II)-DOM complexes varies with pH.

A2.2 Materials and Methods

A2.2.1 Sample Preparation

A stock solution of 0.50 mM Zn(II), 50 mg/L organic matter and 17.5% glycerol (v/v added as a glassing agent) was made from reagent grade $ZnCl_2$ (99.999%, ALDRICH), reverse-osmosis isolated DOM from McDonald's Branch, NJ (Maurice, et al. 2002), glycerol (OmniPur, EM SCIENCE) and 18-MΩ deionized water. 50.0 mL aliquots of the stock solution were placed into 100-mL beakers and the pH adjusted to approximately 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, and 9.0 with small volumes of 0.10 M NaOH. Final pH readings were as follows: 4.94, 5.38, 6.13, 6.48, 6.87, 7.55, 7.83, 8.47, and 8.86. These solutions were transferred to glass vials and sealed.

A2.2.2 Instrumental Analysis

X-Ray Absorption Spectroscopy. Samples of Zn(II)-DOM solution were loaded in Lucite cuvettes with 6 μm polypropylene windows and frozen rapidly in liquid nitrogen. X-ray absorption spectra were measured at the National Synchrotron Light Source (NSLS), beamline X3B, with a Si(111) double crystal monochromator; harmonic rejection was accomplished using a Ni focusing mirror. Fluorescence excitation spectra for all samples were measured with a 13-element solid-state Ge detector array. Samples were held at ~ 15 K in a Displex cryostat during XAS measurements. X-ray energies were calibrated by reference to the absorption spectrum of a zinc metal foil, measured concurrently with the Zn-DOM spectra. All of the data shown represent the average of 6 scans. Data collection and reduction

were performed according to published procedures (1) with E_0 set to 9680 eV for Zn. The Fourier-filtered EXAFS were fit to Equation 1 using the nonlinear least-squares engine of IFEFFIT that is distributed with SixPack (2, 3).

$$\chi(k) = \sum \frac{N_{as} A_s(k) S_c}{kR_{as}^{\ 2}} \exp(-2k^2 \sigma_{as}^{\ 2}) \exp(-2R_{as}/\lambda) \sin[2kR_{as} + \phi_{as}(k)] \qquad (1)$$

In Eq. 1, N_{as} is the number of scatterers within a given radius (R_{as}, ± σ_{as}), $A_s(k)$ is the backscattering amplitude of the absorber-scatterer (*as*) pair, S_c is a scale factor, $\phi_{as}(k)$ is the phase shift experienced by the photoelectron, λ is the photoelectron mean free-path, and the sum is taken over all shells of scattering atoms included in the fit. Theoretical amplitude and phase functions, $A_s(k)exp(-2R_{as}/\lambda)$ and $\phi_{as}(k)$, were calculated using FEFF v. 8.00 (4). The scale factor (S_c) and ΔE_0 for Zn-N (S_c = 0.78, ΔE_0 = -21 eV), Zn-S (0.85, -21 eV), Co-N (0.74, -26 eV) and Co-S (0.85, -26 eV) scattering were determined previously and held fixed throughout this analysis (1). Fits to the current data were obtained for all reasonable integer or half-integer coordination numbers, refining only R_{as} and $\sigma_{as}^{\ 2}$.

A2.3 Results

Although the ratio of Zn(II) to DOM is high, the XANES spectra of the Zn(II)-DOM solutions are quite different from spectra aquo Zn^{2+}, indicating complex formation predominates.

The average Zn-ligand environment in this system is best described as 6-coordinate with all low-Z (N or O) donors at an average distance of 2.07 Å, regardless of pH (Figure Appendix.1). While minor variations are apparent in the EXAFS Fourier transforms, these correspond to minimal perturbations in the first shell disorder (σ^2, see Table Appendix.1). The existence of 1 narrow peak in the EXAFS spectra indicates one type of ligand atom, and

the distance from the fit (R) being 2.07 angstroms is indicative of six-coordination. As O and N are similar in size, the fits cannot distinguish one versus the other, but considering the stoichiometry of O versus N sites in the DOM, the binding site must be overwhelmingly O.

More subtle changes can be observed by comparing the XANES spectra (Figure Appendix.2, left), which are best visualized by examination of the absorption derivatives (Figure Appendix.2, right). The data indicate three distinct spectral types, each showing a slight shift in the derivative maximum (marking to the half-height position of the rising absorption edge). These apparent breaks are annotated by the vertical lines in the right panel of Figure 2, and correspond to shifts of approximately 0.3 eV in the edge energy. As edge energy can be associated with the ease of ionization of the central metal atom, these data suggest that core electrons of the Zn(II) ions are more easily excited at lower pH, consistent with a decrease in the average donor ability of ligating atoms.

A2.4 Discussion

A2.4.1 Site Geometry and Coordination Number

Both EXAFS and XANES spectra are consistent with the results of Xia and co-workers (Xia *et al.* 1997) that the Zn(II)-aquatic DOM binding environment has 6 O or N ligand atoms at all pH values examined. The contrary conclusion of Ramalho and co-workers (Ramalho *et al.* 2007) may be due to their selection of an inappropriate 'model' ligand, a phthalic acid analogue. Phthalic acid has a weaker affinity for Zn(II) than small aliphatic poly-acids (malonic, tartaric, citric, etc.) (NIST); in addition, all of these ligands have a higher affinity for Zn(II) than for Cd(II), in contrast to phthalic acid. The EXAFS results of Karlsson and Skylberg (Karlsson & Skylberg 2007) point to tetrahedral binding in organic soils which are rich in reduced sulfur. In general, the presence of thio-ligands favors

tetrahedral geometry, while O and N ligands favor octahedral (Glusker *et al.* 1999). The discrepancy between the present data, which agree with the Xia and co-workers (Xia *et al.* 1997) and the Karlsson and Skylberg (Karlsson & Skylberg 2007) results may be due to the difference in thiol content of the ligands.

A2.4.2 pH Effects

Although the overall binding geometry remains octahedral as pH increases, the XANES spectra (Fig. Appendix.2) indicate changes in some other aspect of the complexes. Two possible explanations are:

1) changes in the average protonation state of the Zn donors. In this case, the edge shifts seen in Figure 2 correspond approximately to two different pK_a values of the ligating DOM molecules, the first occurring between pH 6.87 and 7.55, and the second occurring between pH 5.38 and 6.13.

2) changes in the binding groups from predominantly carboxylate at low pH to predominantly phenolic at high pH. For example, in an equimolar mixture of tartaric acid (dicarboxylic acid) and catechol (di-phenolic), Zn(II) binding to tartaric acid would be favored at $pH < 7$ while binding to catechol would be favored at $pH > 8$.

While the data are consistent with both explanations, the heterogeneity of the DOM ligand mixture favors the latter explanation.

A2.5 Conclusion

Zn(II) binding to aquatic DOM is octahedral (6-coordinate) with predominantly O and N ligands across the range of environmental pH 5-9. Additional experiments are required to resolve whether the presence of thiol groups favors tetrahedral (Karlsson & Skyllberg, 2007) or octahedral (Xia *et al.* 1997) binding.

A2.6 References

Balistrieri, L. S. & Blank, R. G. 2008 Dissolved and labile concentrations of Cd, Cu, Pb, and Zn in the South Fork Coeur d'Alene River, Idaho: Comparisons among chemical equilibrium models and implications for biotic ligand models. *Appl. Geochem.* **23**(12), 3355-3371.

Botelho, C. M. S., Boaventura, R. A. R. & Goncalves, M. D. S. 2007 Metal complexation with different types of soluble and adsorbed freshwater ligands followed by DPASV. *Aquat. Geochem.* **13**(2),173-186.

Cabaniss, S. E., Madey, G., Leff, L., Maurice, P. A. & Wetzel, R. 2005 A stochastic model for the synthesis and degradation of natural organic matter. Part I. Data structures and reaction kinetics. *Biogeochemistry* **76**(2), 319-347.

Cabaniss, S. E., Madey, G., Leff, L., Maurice, P. A. & Wetzel, R. 2007 A stochastic model for the synthesis and degradation of natural organic matter. Part II. Molecular property distributions. *Biogeochemistry* **86**(3), 269-286.

Cabaniss, S. E. 2008 Quantitative structure-property relationships for predicting metal binding by organic ligands. *Environ. Sci. Technol.* **42**(14), 5210-5216.

Cabaniss, S. E. 2009 Forward modeling of metal complexation by NOM: I. A priori prediction of conditional constants and speciation. *Environ. Sci. Technol.* **43**(8), 2838-2844.

Cheng, T. & Allen, H. E. 2006 Comparison of zinc complexation properties of dissolved natural organic matter from different surface waters. *J. Environ. Manag.* **80**(3), 222-229.

Christensen, J. B. & Christensen, T. H. 1999 Complexation of Cd, Ni and Zn by DOC in polluted groundwater: A comparison of approaches using resin exchange, aquifer material, sorption and computer speciation models (WHAM and MINTEQA2). *Environ. Sci. Technol.* **33**(21), 3857-3863.

Donat, J. & Bruland, K. 1990 A comparison of two voltammetric techniques for determining zinc speciation in Northeast Pacific Ocean waters. *Mar Chem.* **28**(4), 301–323.

Greenwood, N. N. & Earnshaw, A. 1994 *Chemistry of the Elements*. Pergamon Press, Tarrytown, NY, pp 1395-1422.

Jakuba, R. W., Moffett, J. W. & Saito, M. A. 2008 Use of a modified, high-sensitivity, anodic stripping voltammetry method for determination of zinc speciation in the North Atlantic Ocean. *Anal. Chim. Acta* **614**(2), 143-152.

Glusker, J. P.; Katz, A. K. & Bock, C. W. 1999 Metal ions in biological Systems. *Rigaku J.* **16**(2), 8:16.

Jouvin, D.; Louvat, P.; Juillot, F.; Marechal, C. N. & Benedetti, M. F. 2009 Zinc isotopic fractionation: why organic matters. *Environ. Sci. Technol.* **43**(15), 5747-5754.

Karlsson, T. & Skyllberg, U. 2007 Complexation of zinc in organic soils - EXAFS evidence for sulfur associations. *Environ. Sci. Technol.* **41**(1), 119-124.

Koopal, L. K.; Saito, T.; Pinheiro, J. P.; Van Riemsdijk, W. H. 2005 Ion binding to natural organic matter: general considerations and the NICA-Donnan model. *Coll. Surf. A-Physicochemical Engr. Aspects* **265**(1-3), 40-54.

Martinez, C. E.; Bazilevskaya, K. A. & Lanzirotti, A. 2006 Zinc coordination to multiple ligand atoms in organic-rich surface soils. *Environ. Sci. Technol.* **40**(18), 5688-5695.

Maurice, P. A.; Pullin, M. J.; Cabaniss, S. E.; Zhou, Q. H.; Namjesnik-Dejanovic, K. & Aiken, G. R. A comparison of surface water natural organic matter in raw filtered water samples, XAD, and reverse osmosis isolates. *Water Res.* **36**(9), 2357-2371.

Meylan, S.; Odzak, N.; Behra, R. & Sigg, L. 2004 Speciation of copper and zinc in natural freshwater: comparison of voltammetric measurements, diffusive gradients in thin films (DGT) and chemical equilibrium models. *Anal. Chim. Acta* **510**(1), 91-100.

Milne, C. J.; Kinniburgh, D. G.; Van Riemsdijk, W. H. & Tipping, E. 2003 Generic NICA-Donnan model parameters for metal-ion binding by humic substances. *Environ. Sci. Technol.* **37**(5), 958-971.

Morel, F. M. M. & Hering, J. G. 1993 Principles *and Applications of Aquatic Chemistry.* John Wiley & Sons, Inc., New York, NY, pp 319-420.

NIST Critically Selected Stability Constants of Metal Complexes Database, 2004 Version 8.0, US Dept. of Commerce, Gaithersburg, MD.

Ramalho, T. C.; da Cunha, E. F. F.; de Alencastro, R. B.; *et al* 2007 Differential complexation between Zn2+ and Cd2+ with fulvic acid: a computational chemistry study. *Wat. Air Soil Pollut.* **183**(1-4), 467-472.

Shi, Z. Q.; Di Toro, D. M.; Allen, H. E. & Sparks, D. L. 2008 A WHAM-based kinetics model for Zn adsorption and desorption to soils. *Environ. Sci. Technol.* **42**(15), 5630-5636.

Tipping, E. 2005 *Cation Binding by Humic Substances,* Cambridge University Press, Cambridge, 444 pp.

Van Riemsdijk, W. H.; Koopal, L. K.; Kinniburgh, D. G.; Benedetti, M. F. & Weng, L. 2006 Modeling the interactions between humics, ions, and mineral surfaces. *Environ. Sci. Technol.* **40**(24), 7473–7480.

Xia, K.; Bleam, W.; Helmke, P. A. 1997 Studies of the nature of binding sites of first row transition elements bound to aquatic and soil humic substances using X-ray absorption spectroscopy. *Geochim Cosmochim Acta* **61**(11), 2223-2235.

Yang, R.J. & van den Berg, C. M. G. 2009 Metal complexation by humic substances in seawater. *Environ. Sci. Technol.* **43**(19), 7192-7197.

Table A2.1 EXAFS curve fitting results for Zn-DOM as a function of pH[a]

pH	Model	$R_{Zn\text{-}N/O}$	σ^2	R_f [b]	R_u
4.94	6 N/O	2.07	5.3	15	105
5.38	6 N/O	2.07	5.7	21	123
6.13	6 N/O	2.07	5.2	10	70
6.48	6 N/O	2.07	7.1	12	77
6.87	6 N/O	2.07	6.7	22	126
7.55	6 N/O	2.07	7.2	12	78
7.83	6 N/O	2.07	6.3	8	45
8.47	6 N/O	2.07	6.6	24	133
8.86	6 N/O	2.07	6.4	25	126

[a] Distances ($R_{Zn\text{-}N/O}$ in Å) and disorder parameters (σ^2 in10^{-3} Å^2) shown derive from integer coordination number fits to Fourier filtered EXAFS data (Δk = 2 - 12 Å^{-1}; ΔR = 0.8 - 2.0 Å). Fits to unfiltered data gave similar results.

[b] Goodness of fit (R_f for fits to filtered data, R_u for fits to unfiltered data) defined as

$$1000 \times \frac{\sum_{i=1}^{N}\left\{\left[\mathrm{Re}(\chi_{i_{calc}})\right]^2 + \left[\mathrm{Im}(\chi_{i_{calc}})\right]^2\right\}}{\sum_{i=1}^{N}\left\{\left[\mathrm{Re}(\chi_{i_{obs}})\right]^2 + \left[\mathrm{Im}(\chi_{i_{obs}})\right]^2\right\}}$$

, where N is the number of data points.

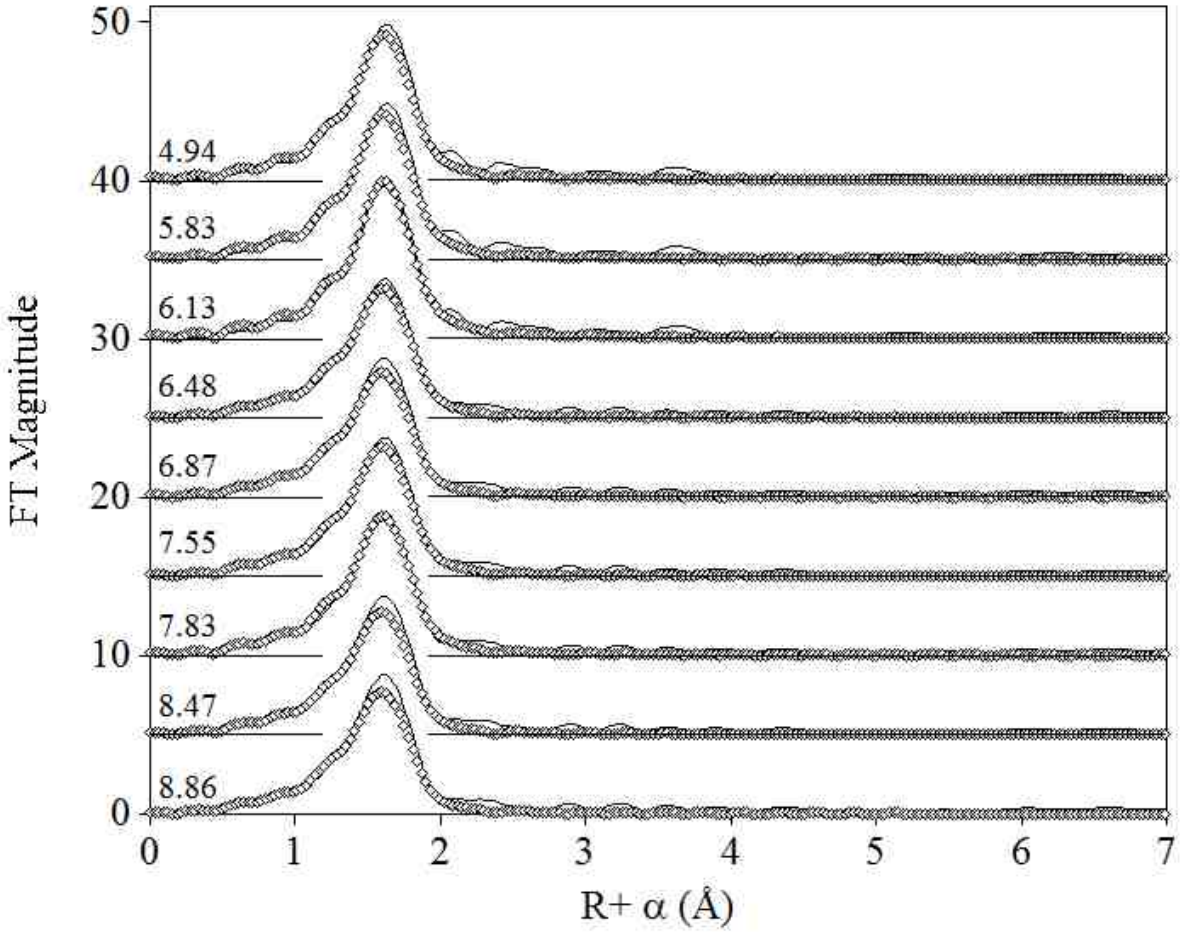


Figure A2.1 Fourier transforms of k^3-weighted EXAFS for Zn-DOM as a function of pH (values given on the plot). All curves have been offset vertically for clarity. The fitted data are presented as black lines; the fits are presented as open diamonds.

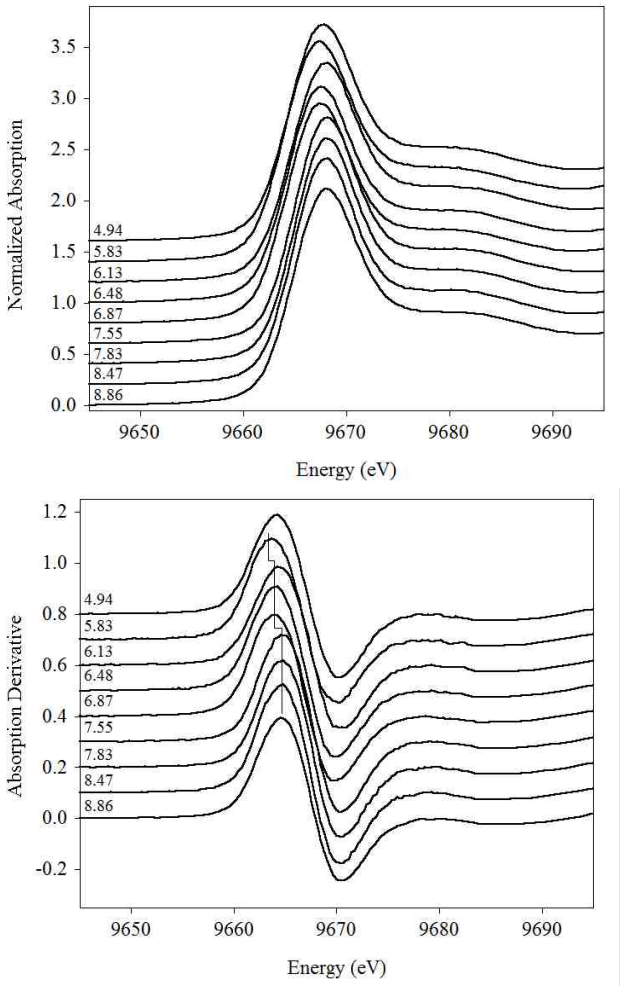


Figure A2.2 Normalized XANES and corresponding derivatives for Zn-DOM as a function of pH (values given on the plots). The vertical lines in the derivative plot are meant to guide the reader; all spectra have been offset vertically for clarity.

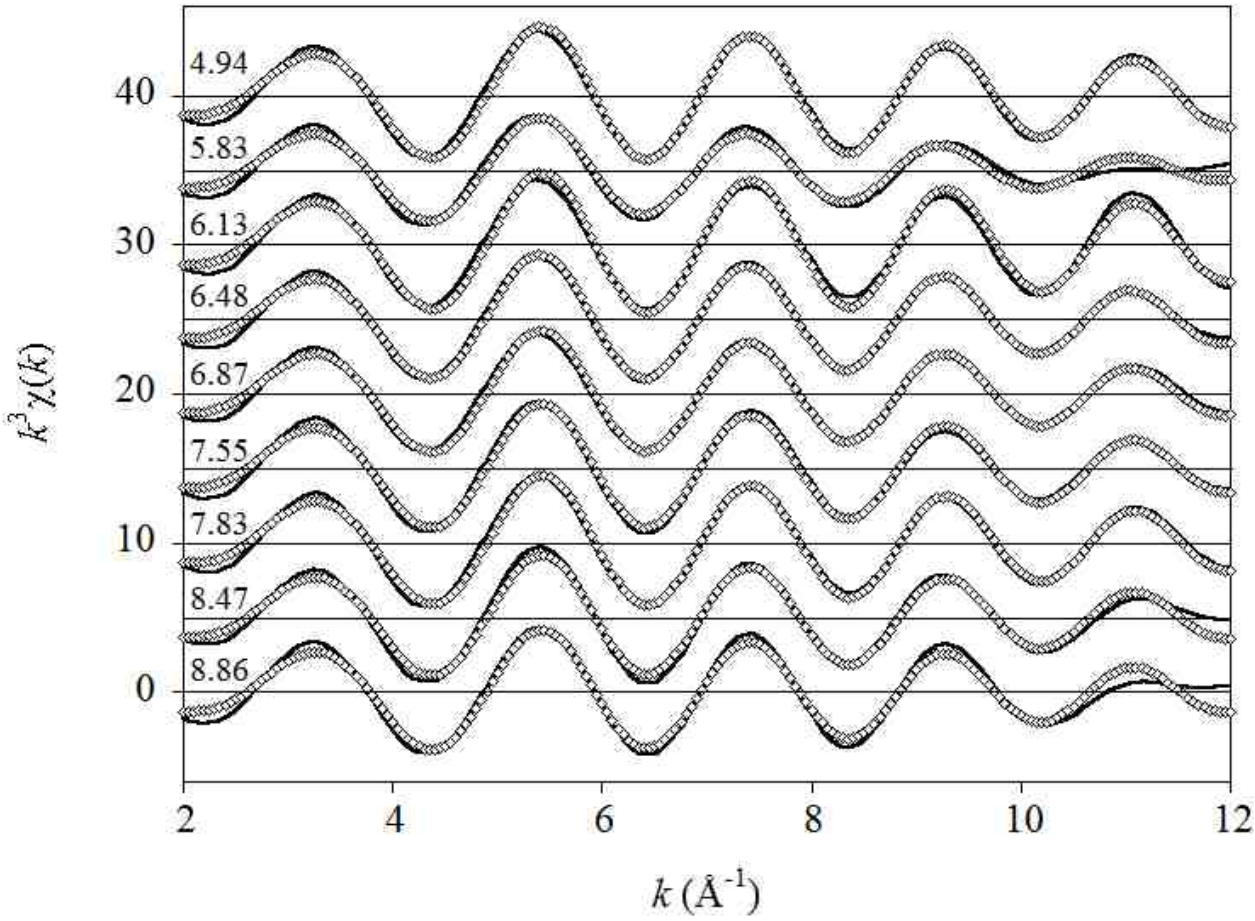


Figure A2.3 Best fits (open diamonds) to $k^3\chi(k)$ EXAFS data (solid lines) for Zn-DOM as a function of pH. The fits correspond to those reported in Table A.1 and shown in Figure A.1.

APPENDIX 3

The Effect of Group IIB Periodicity on Descriptor Significance

From the viewpoint of descriptive chemistry, mercury is somewhat distinct in Group IIB, and cadmium and zinc are rather similar. Mercury has a positive electrode potential, while cadmium and zinc both have negative electrode potentials. Mercury has a much greater tendency to covalency than cadmium and zinc do. Hg(II) forms complexes with N-, P-, and S-donor ligands whose stability is rarely exceeded by those of any other divalent cation, although cadmium and rather similarly zinc do form stable complexes not only with O-donor ligands but with N- and S-donor ligands as well (Greenwood & Earnshaw 1998). On the other hand, Hg^{2+}, Cd^{2+}, and Zn^{2+} all have d^{10} electronic configuration. Chapters 3 and 4 showed the development of QSPRs for predicting Hg(II), Cd(II), and Zn(II) binding by organic ligands, and the developed models have satisfactory statistics. However, how selective and specific are the variables in each QSPR for the cation associated with that QSPR? The answer to this question is given by a case-by-case statistical examination of a) how well the descriptor variables for each of the three cations describe the binding for the other two cations and b) how well the cumulative set of all the descriptor variables used in the three QSPRs describes the binding for Hg(II), Cd(II), and Zn(II).

This appendix shows the application of the descriptor variables of each and every one of the three QSPRs on the other two cations as well as the application of the cumulative set of the descriptor variables used in the three QSPRs on each and every one of the three cations. First, the set of the descriptor variables for each M(II) (M = Hg, Cd, Zn) QSPR is applied on the calibration data sets of the other two M(II) QSPRs. Second, the cumulative set of the three M(II) QSPRs descriptor variables is applied on each of the three calibration

data sets to predict Hg(II), Cd(II), or Zn(II) binding by organic ligands. The numerical results of these applications are presented in Tables A3.1-A3.9. Sections A3.1-A3.9 provide semi-qualitative descriptions of the results of each regression attempt.

Table A3.1 Hg(II) QSPRs Coefficients and their Standard Deviations.

Cd(II) Descriptors	Coefficient	Standard Deviation	p Value
# Amine	2.56	0.79	0.00255
# Thiol	28.8	1.86	3.13×10^{-18}
LN^2	0.422	0.092	4.89×10^{-5}
Z_{dens}	18.9	67.2	0.780
AA	196.	179.	0.282
Zn(II) Descriptors	Coefficient	Standard Deviation	p Value
# Thiol	24.3	1.7	6.44×10^{-18}
LN^2	0.523	0.050	4.56×10^{-13}
Het:C	5.36	0.81	5.78×10^{-8}
ALL Descriptors	Coefficient	Standard Deviation	p Value
# Amine	0.475	1.614	0.770
# Thiol	26.0	1.2	7.18×10^{-22}
LN^2	0.0404	0.0986	0.685
# N	5.45	1.52	0.00106
# N * # C	-0.266	0.058	5.51×10^{-5}
# OH + 8.3 # COOH	0.380	0.092	0.000211
Z_{dens}	17.5	39.7	0.661
Het:C	0.856	0.853	0.322
AA	-270.	81.	0.00203

Table A3.2 Cd(II) QSPRs Coefficients and their Standard Deviations.

Hg(II) Descriptors	Coefficient	Standard Deviation	p Value
# N	0.130	0.51	0.799
# N * # C	0.162	0.088	0.0707
# OH + 8.3 # COOH	0.157	0.014	7.75×10^{-17}
# Thiol	3.75	0.57	1.17×10^{-8}
Zn(II) Descriptors	Coefficient	Standard Deviation	p Value
# Thiol	3.18	0.36	2.63×10^{-12}
LN^2	0.344	0.022	3.12×10^{-23}
Het:C	0.490	0.190	0.0124
ALL Descriptors	Coefficient	Standard Deviation	p Value
# Amine	1.51	0.61	0.0156
# Thiol	2.62	0.46	4.95×10^{-7}
LN^2	0.327	0.049	1.64×10^{-8}
# N	-0.110	0.587	0.851
# N * # C	0.00149	0.07406	0.984
# OH + 8.3 # COOH	-0.0433	0.0349	0.219
Z_{dens}	74.0	29.5	0.0151
Het:C	0.415	0.415	0.322
AA	-144.	54.	0.00993

Table A3.3 Zn(II) QSPRs Coefficients and their Standard Deviations.

Hg(II) Descriptors	Coefficient	Standard Deviation	p Value
# N	0.192	0.462	0.678
# N * # C	0.222	0.077	0.00543
# OH + 8.3 # COOH	0.180	0.012	2.83×10^{-23}
# Thiol	1.81	0.53	0.00107
Cd(II) Descriptors	Coefficient	Standard Deviation	p Value
# Amine	2.04	0.36	3.32×10^{-7}
# Thiol	1.05	0.36	0.00514
LN^2	0.282	0.029	7.28×10^{-14}
Z_{dens}	98.1	16.5	1.32×10^{-7}
AA	-134.	46.	0.00462
ALL Descriptors	Coefficient	Standard Deviation	p Value
# Amine	1.11	0.52	0.0372
# Thiol	1.53	0.45	0.00109
LN^2	0.263	0.043	1.02×10^{-7}
# N	-1.19	0.552	0.0352
# N * # C	0.173	0.068	0.0142
# OH + 8.3 # COOH	0.0320	0.0314	0.312
Z_{dens}	-12.5	33.0	0.706
Het:C	1.12	0.402	0.00742
AA	-83.1	47.3	0.0841

Table A3.4 Hg(II) Binding Data (N = 44).

Set of Descriptors	Hg(II)	Cd(II)	Zn(II)	ALL
Adjusted R^2	0.965	0.918	0.931	0.963
S_{pred}	1.60	3.78	3.36	1.46
Are all $p < 0.05$?	Yes	No	Yes	No

Table A3.5 Cd(II) Binding Data (N = 63).

Set of Descriptors	Hg(II)	Cd(II)	Zn(II)	ALL
Adjusted R^2	0.838	0.931	0.922	0.928
S_{pred}	1.57	0.935	1.02	0.953
Are all $p < 0.05$?	No	Yes	Yes	No

Table A3.6 Zn(II) Binding Data (N = 68).

Set of Descriptors	Hg(II)	Cd(II)	Zn(II)	ALL
Adjusted R^2	0.883	0.937	0.934	0.943
S_{pred}	1.39	0.948	0.984	0.876
Are all $p < 0.05$?	No	Yes	Yes	No

Table A3.7 Hg(II) QSPRs Descriptor Correlations.

	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9
D_1	1								
D_2	-0.306	1							
D_3	0.572	-0.127	1						
D_4	0.988	-0.217	0.569	1					
D_5	0.881	-0.207	0.675	0.890	1				
D_6	-0.256	-0.014	0.609	-0.264	-0.047	1			
D_7	-0.657	0.376	0.018	-0.640	-0.340	0.653	1		
D_8	-0.298	0.225	0.118	-0.304	-0.227	0.519	0.564	1	
D_9	-0.124	0.120	0.114	-0.149	-0.155	0.304	0.272	0.289	1

Note: D_1 = # Amine, D_2 = # Thiol, $D_3 = LN^2$, D_4 = # N, D_5 = # N * # C, D_6 = # OH + 8.3 # COOH, $D_7 = Z_{dens}$, D_8 = Het:C, D_9 = AA

Table A3.8 Cd(II) QSPRs Descriptor Correlations.

	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9
D_1	1								
D_2	0.042	1							
D_3	0.428	-0.064	1						
D_4	0.856	0.183	0.294	1					
D_5	0.735	0.099	0.405	0.870	1				
D_6	-0.166	-0.366	0.742	-0.235	-0.058	1			
D_7	-0.654	0.009	0.176	-0.572	-0.339	0.584	1		
D_8	0.052	0.014	0.332	0.096	-0.042	0.369	0.266	1	
D_9	0.657	-0.009	0.246	0.643	0.503	-0.060	-0.254	0.162	1

Note: D_1 = # Amine, D_2 = # Thiol, $D_3 = LN^2$, D_4 = # N, D_5 = # N * # C,
D_6 = # OH + 8.3 # COOH, $D_7 = Z_{dens}$, D_8 = Het:C, D_9 = AA

Table A3.9 Zn(II) QSPRs Descriptor Correlations.

	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9
D_1	1								
D_2	0.000	1							
D_3	0.522	-0.035	1						
D_4	0.884	0.154	0.394	1					
D_5	0.812	0.073	0.483	0.899	1				
D_6	-0.143	-0.307	0.695	-0.227	-0.084	1			
D_7	-0.721	0.065	0.033	-0.669	-0.505	0.552	1		
D_8	0.025	0.087	0.310	0.073	-0.047	0.370	0.323	1	
D_9	0.671	0.022	0.286	0.670	0.528	-0.106	-0.392	0.155	1

Note: D_1 = # Amine, D_2 = # Thiol, $D_3 = LN^2$, D_4 = # N, D_5 = # N * # C, D_6 = # OH + 8.3 # COOH, $D_7 = Z_{dens}$, D_8 = Het:C, D_9 = AA

A3.1 Hg(II) QSPR Descriptor Variables for Predicting Cd(II) Binding by Organic Ligands

Applying Hg(II) QSPR descriptor variables for modeling Cd(II) binding by organic ligands results in a model that is both of lower quality compared to the Cd(II) QSPR developed in Chapter 3 and unsatisfactory in general. The standard error of prediction (S_{pred}) is 1.57 log units, a number higher than both 0.935 log units and ~1 log unit (that was set as the highest acceptable S_{pred} for a Cd(II) QSPR). The adjusted R^2 value is 0.838, which is smaller than both 0.931 and the minimum desired value of ~0.90. Another problem with the model is that # N and # N * # C are both statistically insignificant variables in the model.

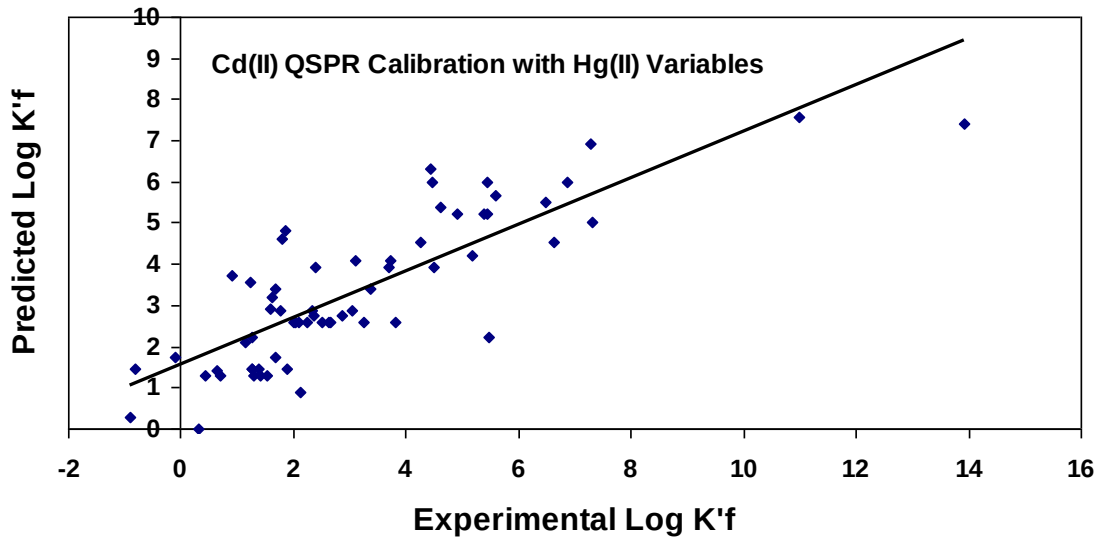


Figure A3.1 QSPR predictions versus experimental log K'_f values

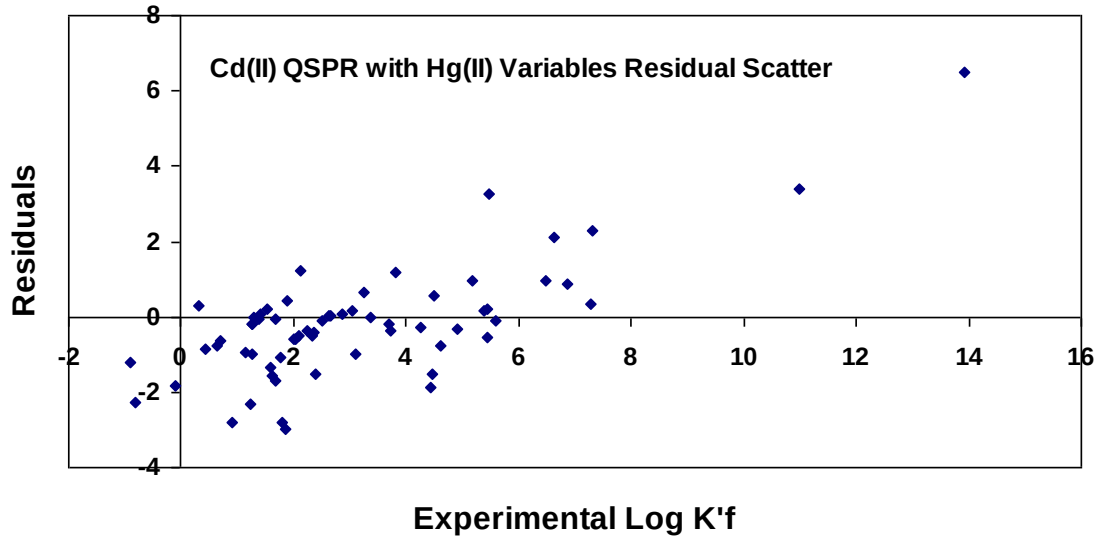


Figure A3.2 The QSPR calibration residuals versus experimental log K'_f

A3.2 Hg(II) QSPR Descriptor Variables for Predicting Zn(II) Binding by Organic Ligands

Applying Hg(II) QSPR descriptor variables for modeling Zn(II) binding by organic ligands results in a model that is both of lower quality compared to the Zn(II) QSPR developed in Chapter 3 and unsatisfactory in general. The standard error of prediction (S_{pred}) is 1.39 log units, a number higher than both 0.984 log units and ~1 log unit (that was set as the highest acceptable S_{pred} for a Zn(II) QSPR). The adjusted R^2 value is 0.883, which is smaller than both 0.934 and the minimum desired value of ~0.90. Another problem with the model is that # N is a statistically insignificant variable in the model.

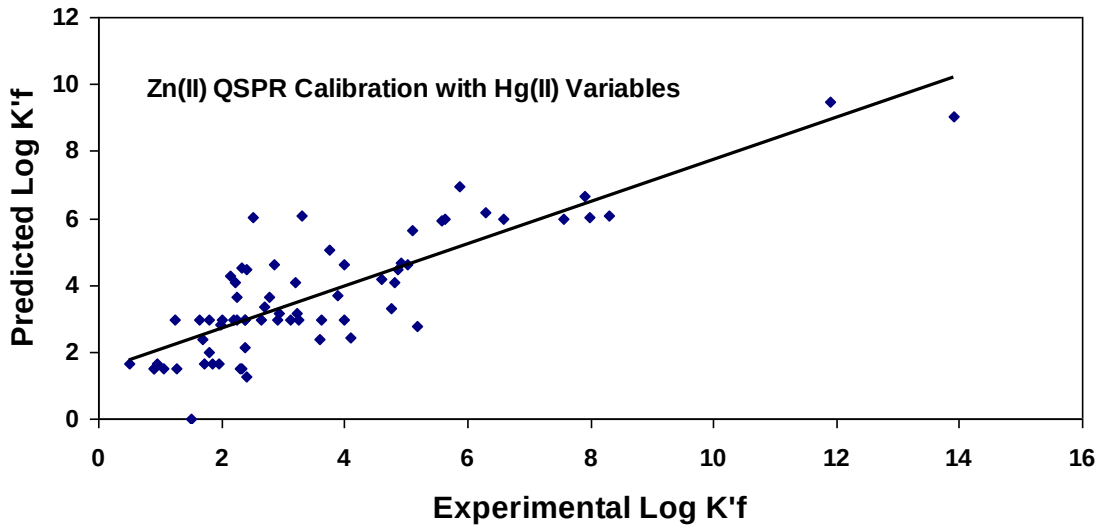


Figure A3.3 QSPR predictions versus experimental log *K'*$_f$ values

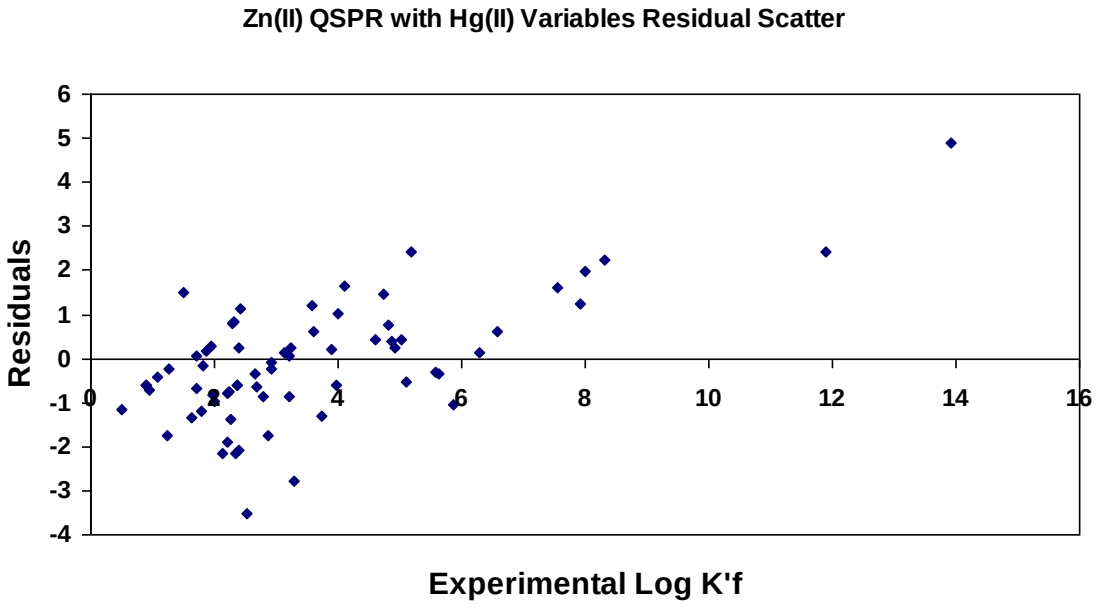


Figure A3.2 The QSPR calibration residuals versus experimental log K'_f

A3.3 Cd(II) QSPR Descriptor Variables for Predicting Hg(II) Binding by Organic Ligands

Applying Cd(II) QSPR descriptor variables for modeling Hg(II) binding by organic ligands results in a model that is both of lower quality compared to the Hg(II) QSPR developed in Chapter 2 and unsatisfactory in general. The standard error of prediction (S_{pred}) is 3.78 log units, a number both higher than 1.60 log units and not < 2.0 log units (that was set as the criteria of S_{pred} for a Hg(II) QSPR). The adjusted R^2 value is 0.918, which is smaller than 0.965 although greater than the minimum desired value of ~0.90. Another problem with the model is that both Z_{dens} and AA variables are statistically insignificant variables in the model.

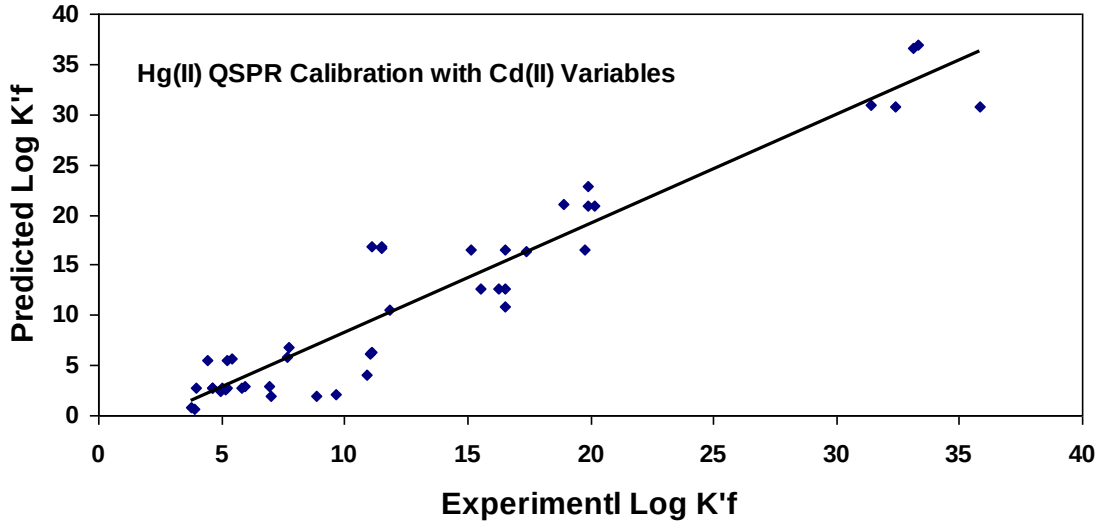


Figure A3.5 QSPR predictions versus experimental log K'_f values

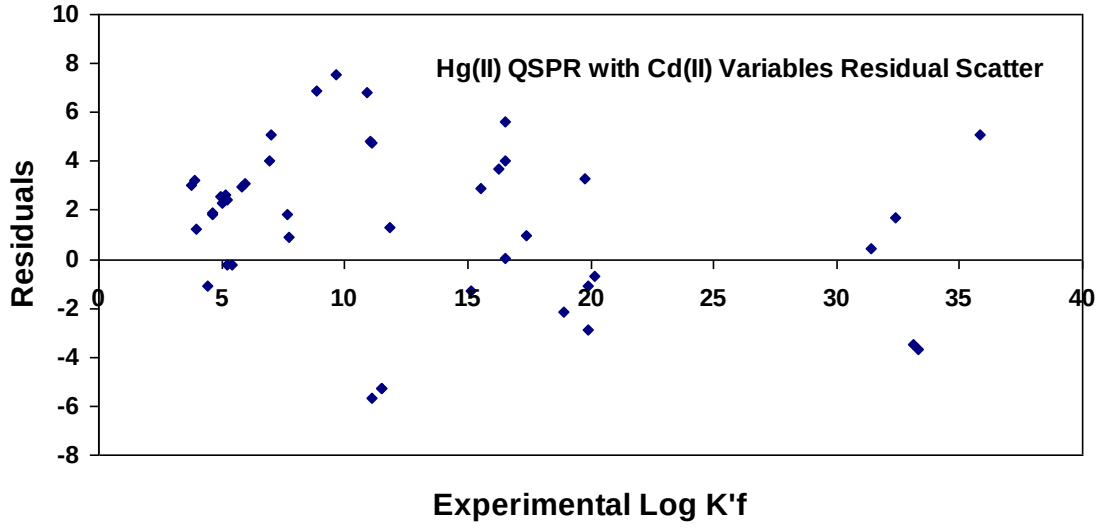


Figure A3.6 The QSPR calibration residuals versus experimental log K'_f

A3.4 Cd(II) QSPR Descriptor Variables for Predicting Zn(II) Binding by Organic Ligands

Applying Cd(II) QSPR descriptor variables for modeling Zn(II) binding by organic ligands results in a model that is both slightly better as for predictive ability compared to the Zn(II) QSPR developed in Chapter 3 and satisfactory in general. The standard error of prediction (S_{pred}) is 0.948 log units, a number smaller than both 0.984 log units and ~1 log unit (that was set as the highest acceptable S_{pred} for a Zn(II) QSPR). The adjusted R^2 value is 0.937, which is both slightly greater than 0.934 and greater than the minimum desired value of ~0.90. Also all the variables are statistically significant variables in the model. It has, however, one disadvantage compared to the Zn(II) QSPR developed in Chapter 3. That is requiring more descriptor variables.

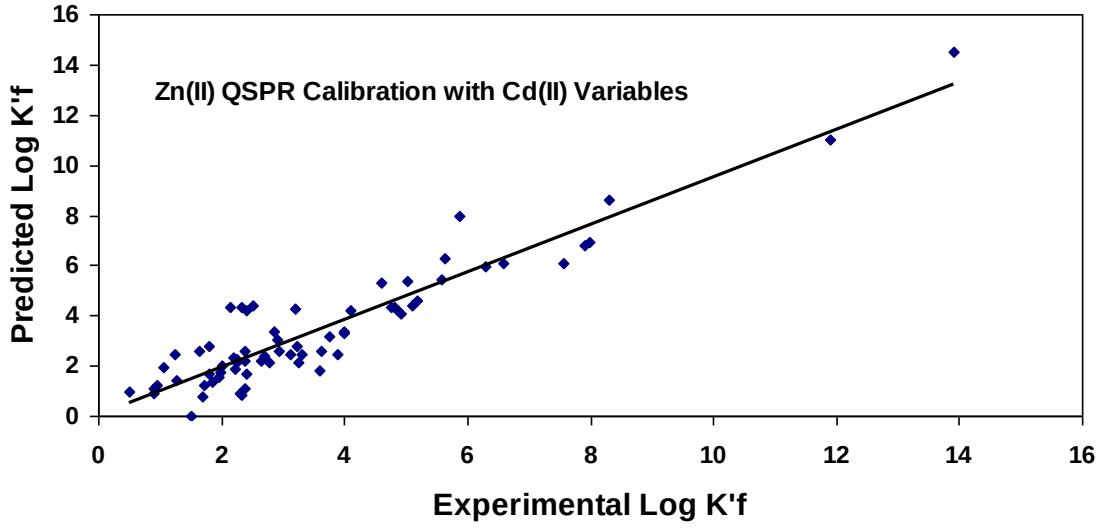


Figure A3.7 QSPR predictions versus experimental log K'_f values

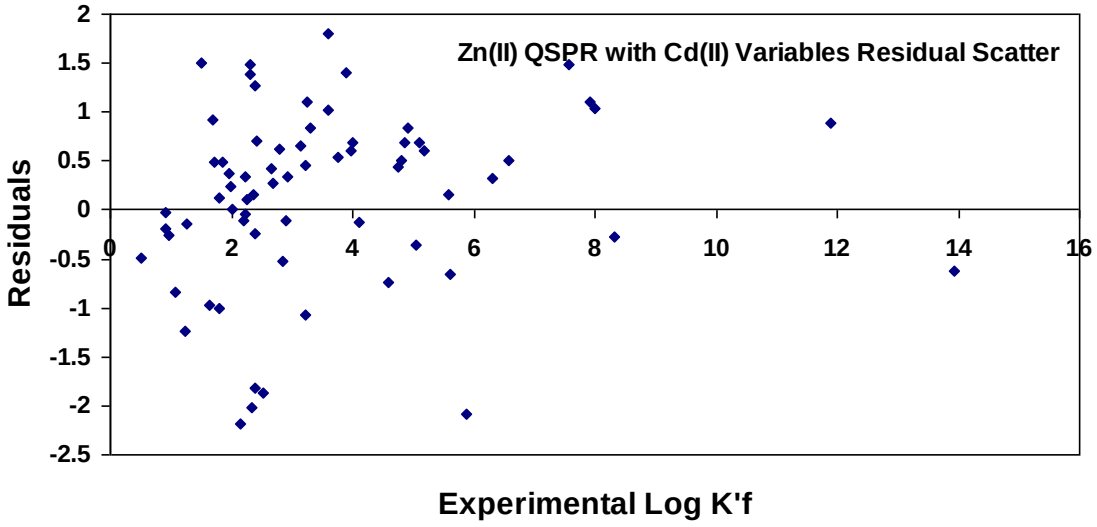


Figure A3.8 The QSPR calibration residuals versus experimental log K'_f

A3.5 Zn(II) QSPR Descriptor Variables for Predicting Hg(II) Binding by Organic Ligands

Applying Zn(II) QSPR descriptor variables for modeling Hg(II) binding by organic ligands results in a model that is both of lower quality compared to the Hg(II) QSPR developed in Chapter 2 and unsatisfactory in general. The standard error of prediction (S_{pred}) is 3.36 log units, a number both higher than 1.60 log units and not < 2.0 log units (that was set as the criteria of S_{pred} for a Hg(II) QSPR). The adjusted R^2 value is 0.931, which is smaller than 0.965 although greater than the minimum desired value of ~0.90. However, all variables are statistically significant in the model.

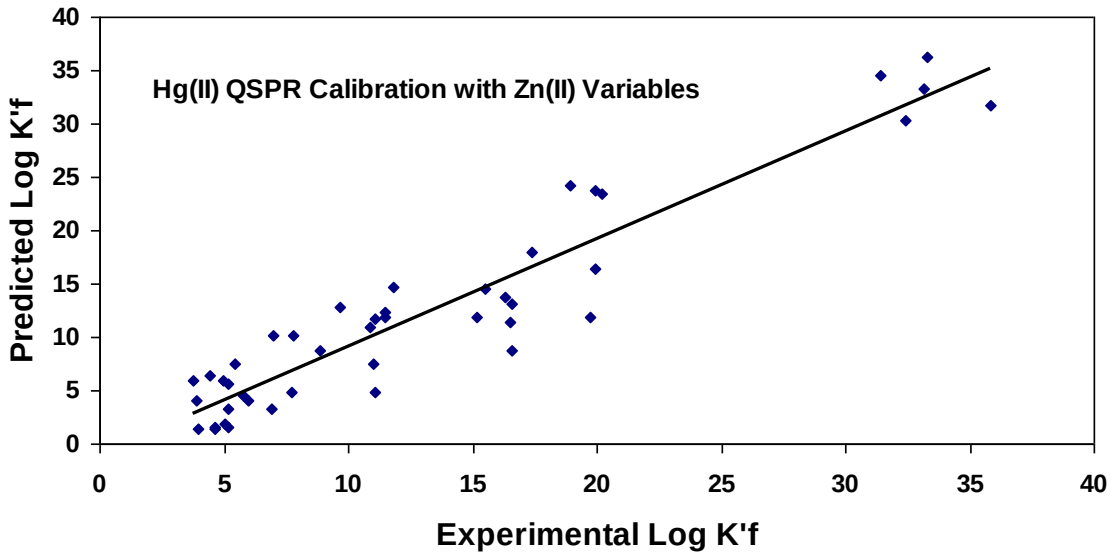


Figure A3.9 QSPR predictions versus experimental log K'_f values

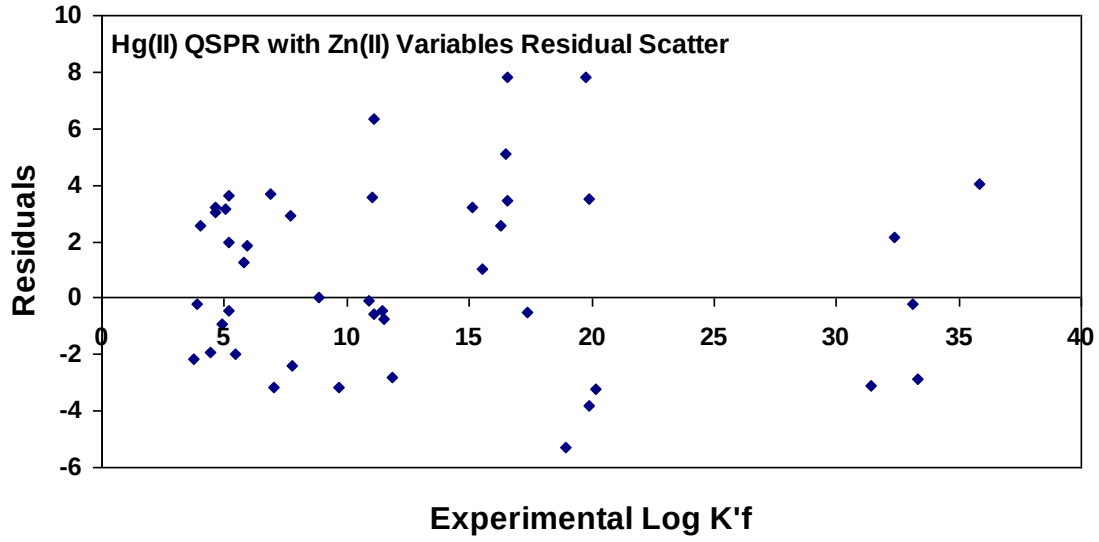


Figure A3.10 The QSPR calibration residuals versus experimental log K'_f

A3.6 Zn(II) QSPR Descriptor Variables for Predicting Cd(II) Binding by Organic Ligands

Applying Zn(II) QSPR descriptor variables for modeling Cd(II) binding by organic ligands results in a model that is of slightly lower quality compared to the Cd(II) QSPR developed in Chapter 3 yet satisfactory in general. The standard error of prediction (S_{pred}) is 1.02 log units, a number that is higher than 0.935 log units but is still ~1 log unit (that was set as the highest acceptable S_{pred} for a Cd(II) QSPR). The adjusted R^2 value is 0.922, which is slightly smaller than 0.931 yet higher than the minimum desired value of ~0.90. Also all the variables of the model are statistically significant in the model.

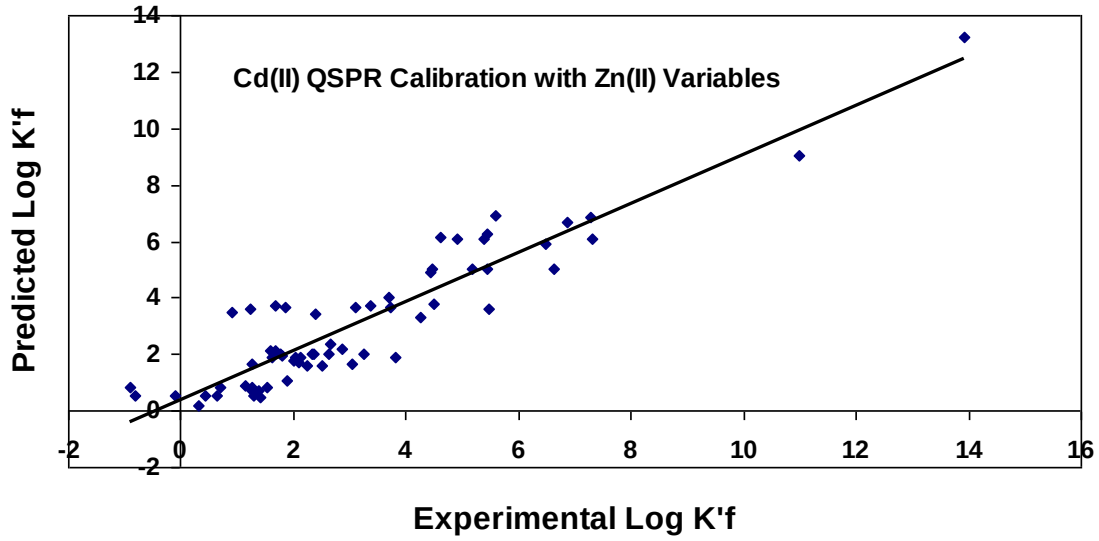


Figure A3.11 QSPR predictions versus experimental log K'_f values

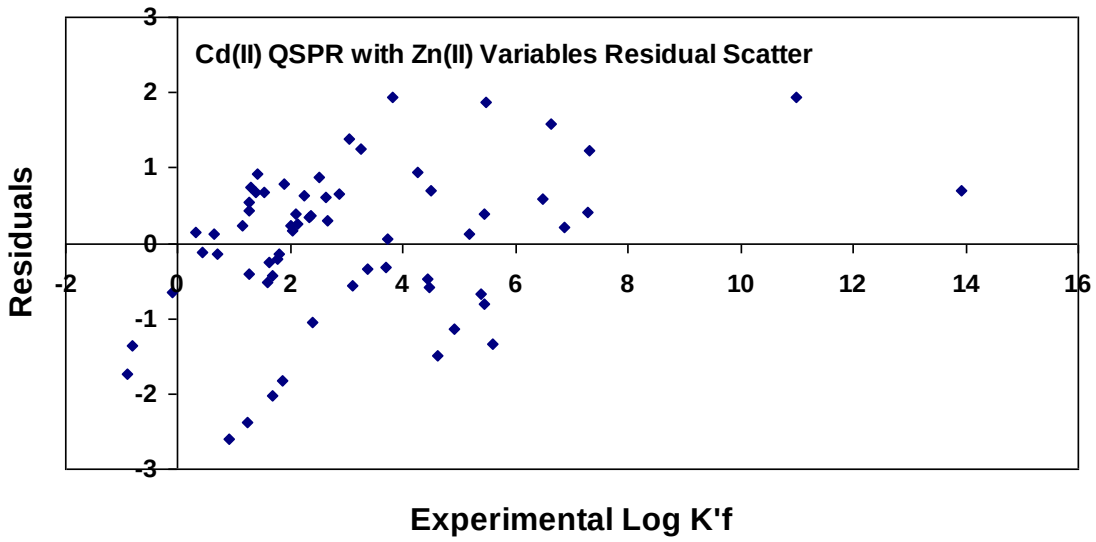


Figure A3.12 The QSPR calibration residuals versus experimental log K'_f

A3.7 The Cumulative Set of the Three QSPRs Descriptor Variables for Predicting Hg(II) Binding by Organic Ligands

Applying the cumulative set of the three QSPRs descriptor variables for modeling Hg(II) binding by organic ligands results in a model that is both of lower quality compared to the Hg(II) QSPR developed in Chapter 2 and unsatisfactory in general. The standard error of prediction (S_{pred}) is 1.46 log units, a number that is both smaller than 1.60 log units and < 2.0 log unit (that was set as the criteria of S_{pred} for a Hg(II) QSPR). The adjusted R^2 value is 0.963, which is slightly smaller than 0.965 but still higher than the minimum desired value of ~0.90. However, both an acceptable S_{pred} and an acceptable adjusted R^2 value should always be statistically expected when some external variables are added to the original variables. The determining factor in such cases is the statistical significance or insignificance of the

cumulative variables. This model has the problem that many of its variables, being # Amine, LN^2, Z_{dens}, and Het:C, are statistically insignificant variables in the model.

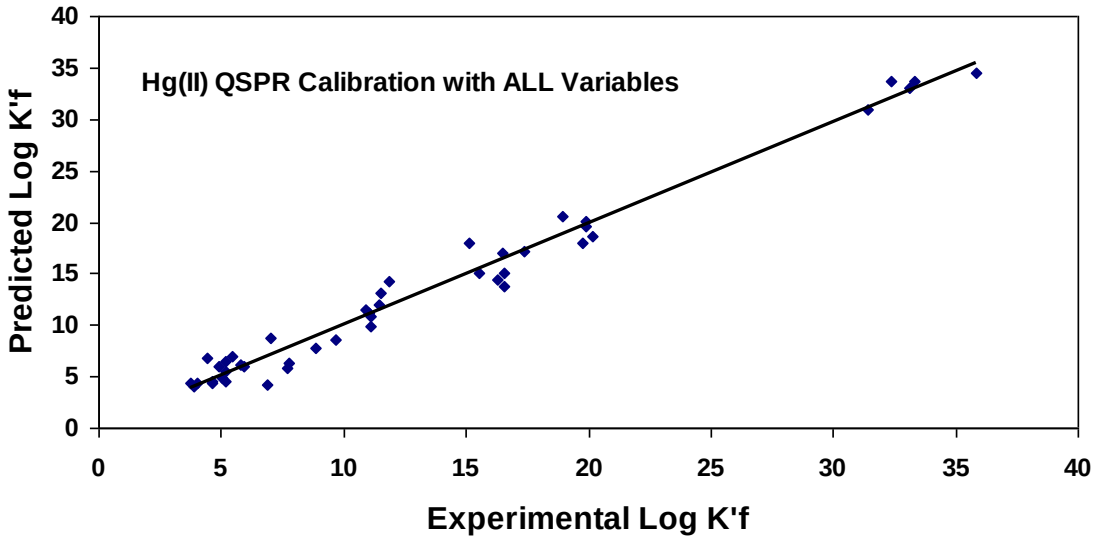


Figure A3.13 QSPR predictions versus experimental log K'_f values

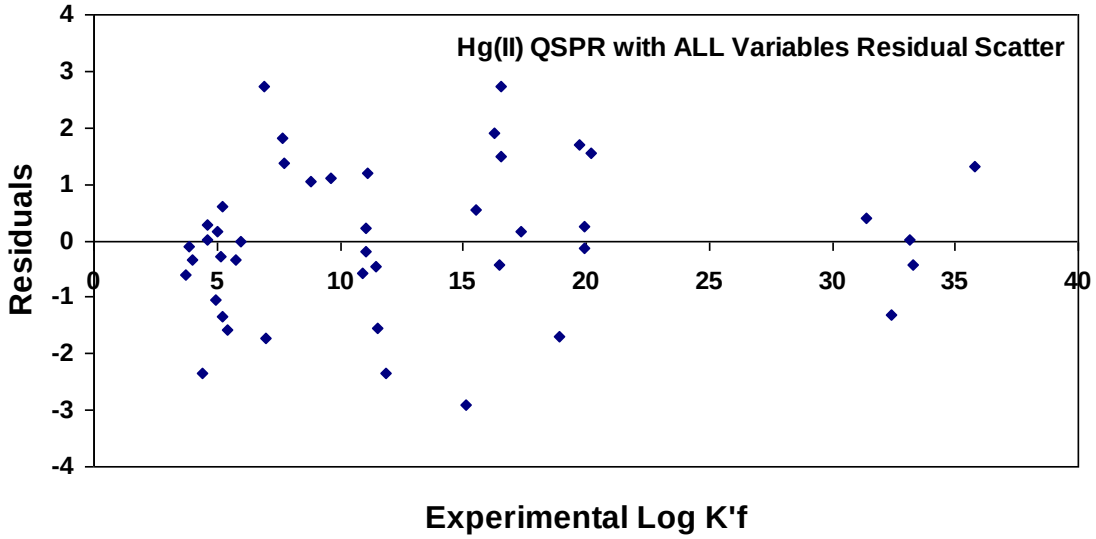


Figure A3.14 The QSPR calibration residuals versus experimental log K'_f

A3.8 The Cumulative Set of the Three QSPRs Descriptor Variables for Predicting Cd(II) Binding by Organic Ligands

Applying the cumulative set of the three QSPRs descriptor variables for modeling Cd(II) binding by organic ligands results in a model that is both of lower quality compared to the Cd(II) QSPR developed in Chapter 3 and unsatisfactory in general. The standard error of prediction (S_{pred}) is 0.953 log units, a number that is greater than 0.935 log units but is still ~1 log unit (that was set as the highest acceptable S_{pred} for a Cd(II) QSPR). The adjusted R^2 value is 0.928, which is slightly smaller than 0.931 but still higher than the minimum desired value of ~0.90. However, both an acceptable S_{pred} and an acceptable adjusted R^2 value should always be statistically expected when some external variables are added to the original variables. The determining factor in such cases is the statistical significance or insignificance of the cumulative variables. This model has the problem that many of its variables, being # N, # N * # C, # OH + 8.3 # COOH, and Het:C, are statistically insignificant variables in the model.

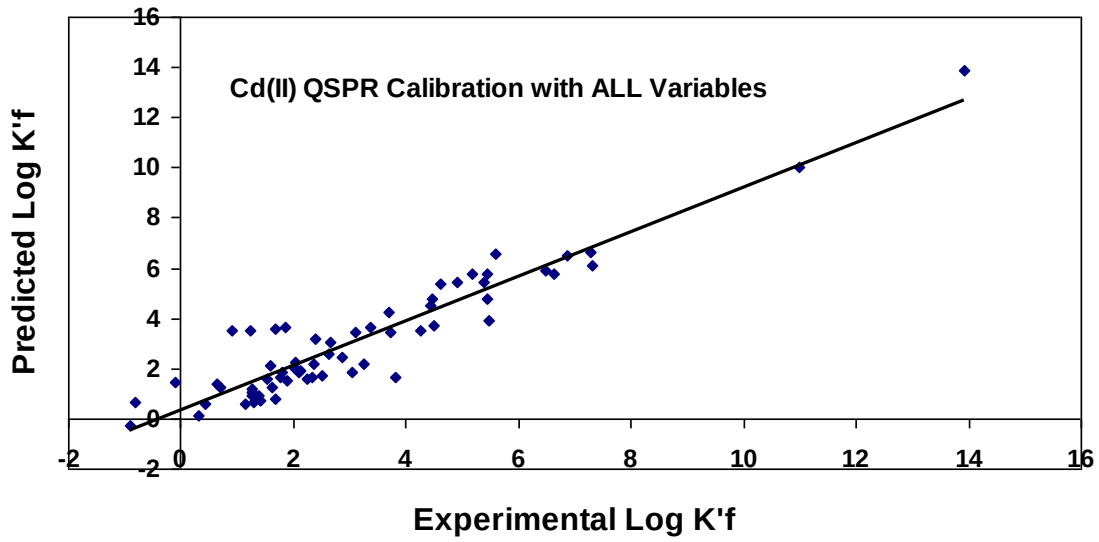


Figure A3.15 QSPR predictions versus experimental log K'_f values

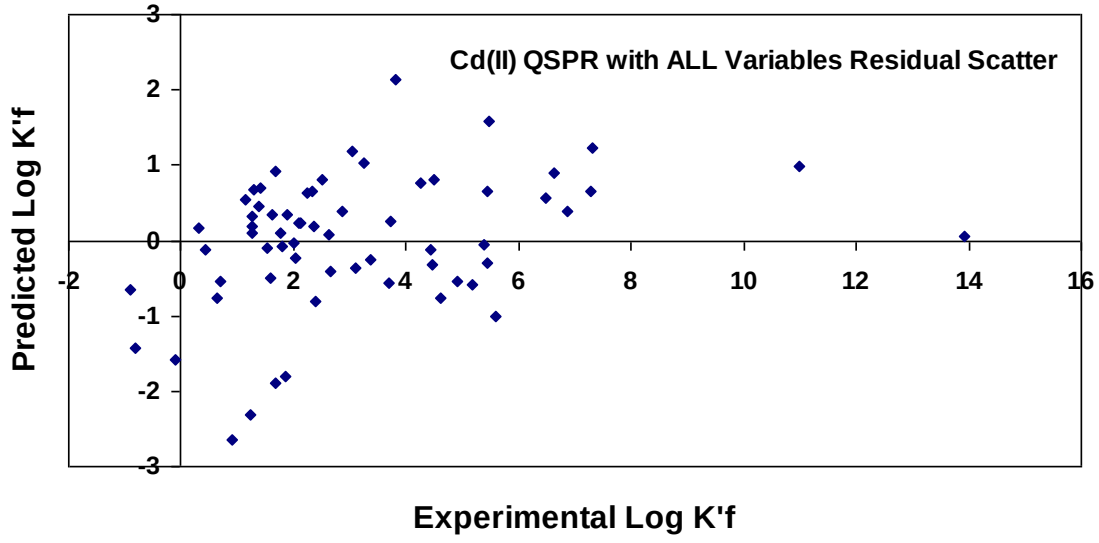


Figure A3.16 The QSPR calibration residuals versus experimental log K'_f

A3.9 The Cumulative Set of the Three QSPRs Descriptor Variables for Predicting Zn(II) Binding by Organic Ligands

Applying the cumulative set of the three QSPRs descriptor variables for modeling Zn(II) binding by organic ligands results in a model that is both of lower quality compared to the Zn(II) QSPR developed in Chapter 3 and unsatisfactory in general. The standard error of prediction (S_{pred}) is 0.876 log units, a number that is smaller than both 0.984 log units and ~1 log unit (that was set as the highest acceptable S_{pred} for a Zn(II) QSPR). The adjusted R^2 value is 0.943, which is both slightly greater than 0.934 and higher than the minimum desired value of ~0.90. However, both an acceptable S_{pred} and an acceptable adjusted R^2 value should always be statistically expected when some external variables are added to the original variables. The determining factor in such cases is the statistical significance or insignificance of the cumulative variables. This model has the problem that many of its variables, being # OH + 8.3 # COOH, Z_{dens}, and AA, are statistically insignificant variables in the model.

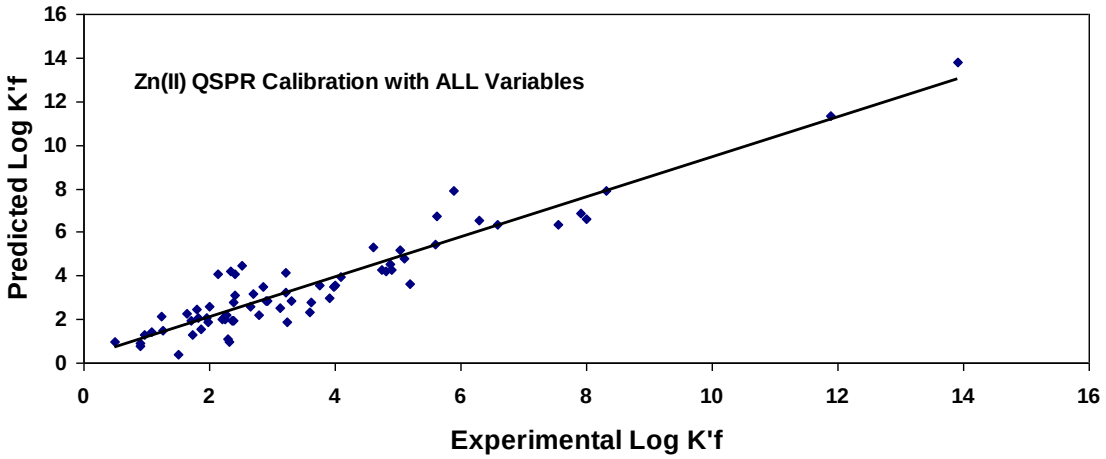


Figure A3.17 QSPR predictions versus experimental log K'_f values

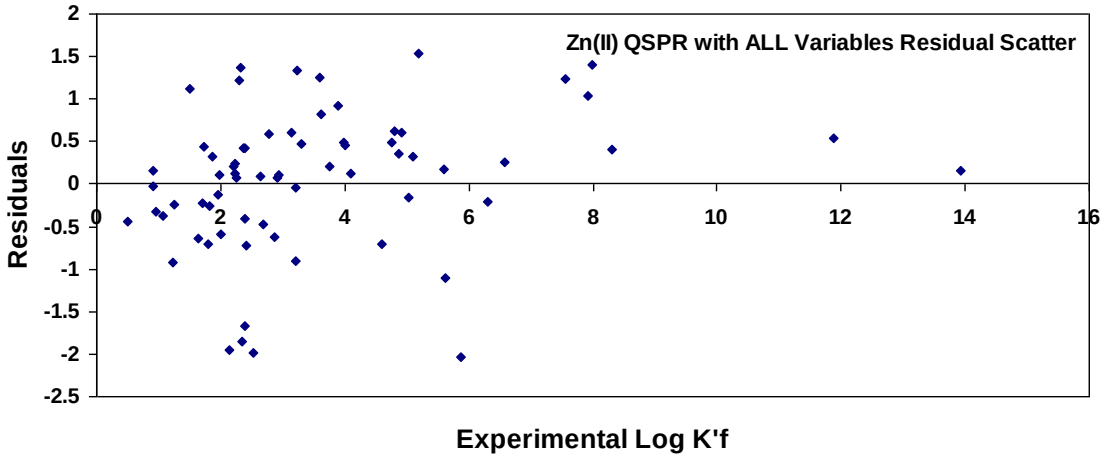


Figure A3.18 The QSPR calibration residuals versus experimental log K'_f

A3.10 References

Greenwood, N. N. & Earnshaw, A. 1994 *Chemistry of the Elements.* Pergamon Press, Tarrytown, NY.