

# Data and code for the optimization of bond valence parameters in metalorganic binding sites

This document describes how  $R_0$  parameters from the bond valence model were semi-automatically derived for metal-organic binding sites from the Cambridge Structural Database (CSD). The data and results described here, as well as the code used can found and reference using the DOI [10.6084/m9.figshare.964285](https://doi.org/10.6084/m9.figshare.964285), and is also accessible at <https://github.com/langner/bvparam-metalorganics>.

## 1 Description of the data

There are two CSV files provided, which represent the input and output. The first, `initial.csv`, lists all the binding sites and distances input into the optimization procedure before filtering, and was extracted from the CSD. Specifically, all ligand atoms of the elements with R factors below 7.5%. The second, `optimized.csv`, lists all the distances and corresponding bond valence contributions after the optimization procedure, for binding sites that passed the validation phase.

## 2 Description of the code

The Python script `optimize.py` reads in `initial.csv` and writes `optimized.csv` after performing the optimization procedure. The code contains a number of comments that explain the optimization process and technical details related to validation and processing binding sites.

## 3 Theoretical background for optimization

The bond valence sum rule, the central tenet of the model, says that for a cationic atom  $i$  of element  $\alpha$  the sum of valences of the bonds it participates in should be equal to the oxidation state  $S_i$ ,

$$\sum_{j \in i} v_{ij} = S_i,$$

and the bond valences  $v_{ij}$  are proportional to the bond lengths  $d_{ij}$ , with the most popular fitting function being,

$$v_{ij} = e^{\frac{R_0 - d_{ij}}{b}}.$$

where  $R_0 = R_{\alpha\beta}$  is the distance for a bond of unit valence between atom types  $\alpha$  and  $\beta$ , and  $b$  is typically 0.37 (may also be optimized). Index  $j$  enumerates anions in the first coordination sphere of  $i$ . In practice there is always a deviation, but the calculated bond valence sum (BVS) should within acceptable limits be  $V_i \approx S_i$ .

For homoleptic sites, the  $e^{\frac{R_0}{b}}$  factor can be extracted and so there is one unique value of  $R_0$  satisfying the sum rule for any particular site,

$$R_0 = b \ln \left( \frac{S_i}{\Omega_i} \right),$$

where

$$\Omega_i = \sum_{j \in i} e^{-\frac{d_{ij}}{b}},$$

and a typical approach in the literature has been to calculate such parameters for a number of manually chosen sites and to report the average  $\bar{R}_0$  as an accurate value.

For heteroleptic sites, there are multiple  $R_0$  parameters for each BVS equation, so no unique values can be found for any one binding site. However, an optimal set of  $\{\bar{R}_{\alpha\beta}\}$  for a given cation type  $\alpha$  can be found by minimizing the summed square deviation from the expected oxidation state for many sites,

$$\{\bar{R}_{\alpha\beta}\} = \arg \min (\sigma_\alpha^2),$$

$$\sigma_\alpha^2 = \sum_{i \in \alpha} (V_i - S_i)^2,$$

and, in our case,  $i$  spans all suitable binding sites with cation  $\alpha$  as the central metal from the CSD.

Since  $\sigma_\alpha^2$  is a function of  $\{R_{\alpha\beta}\}$  through exponentials, it is convenient to use the gradient for optimization (I use the nonlinear conjugate gradient method),

$$\frac{\partial \sigma_\alpha^2}{\partial R_{\alpha\beta}} = \sum_{i \in \alpha} \frac{2}{b} e^{\frac{R_{\alpha\beta}}{b}} \Omega_i^\beta (V_i - S_i),$$

where  $\Omega_i^\beta$  is the  $\beta$ -selecting version of  $\Omega_i$ ,

$$\Omega_i^\beta = \sum_{j \in i} \delta(j \in \beta) e^{-\frac{d_{ij}}{b}}.$$

A disadvantage of seeking to optimize the variance of  $V_i$  for a wide range of heteroleptic sites is that there is no clear way to estimate a reliable error implied in any particular  $R_{\alpha\beta}$ . However an upper bound may be set by assuming that  $V_i - S_i$  is caused by the error in just one parameter and calculating the change  $\Delta R_{\alpha\beta}^i$  required in order to achieve  $V_i = S_i$ . There is one unique such value for each binding site,

$$\Delta R_{\alpha\beta}^i = b \ln \left( \frac{S_i}{V_i} \right).$$

This value will be the same for each anion for any particular binding site  $i$ , but will depend on the distribution of  $V_i - S_i$  values for sites that contain anions of type  $\beta$ . This distribution will confer the quality of the set of binding sites, in terms of the dispersion of calculated BVS sums. The related variance  $|\Delta R_{\alpha\beta}^2|$  will be an upper bound on the variance of  $R_{\alpha\beta}$ . This is what was reported as the error in the manuscript.

## 4 Listing of initial and optimized bond valence sum distributions

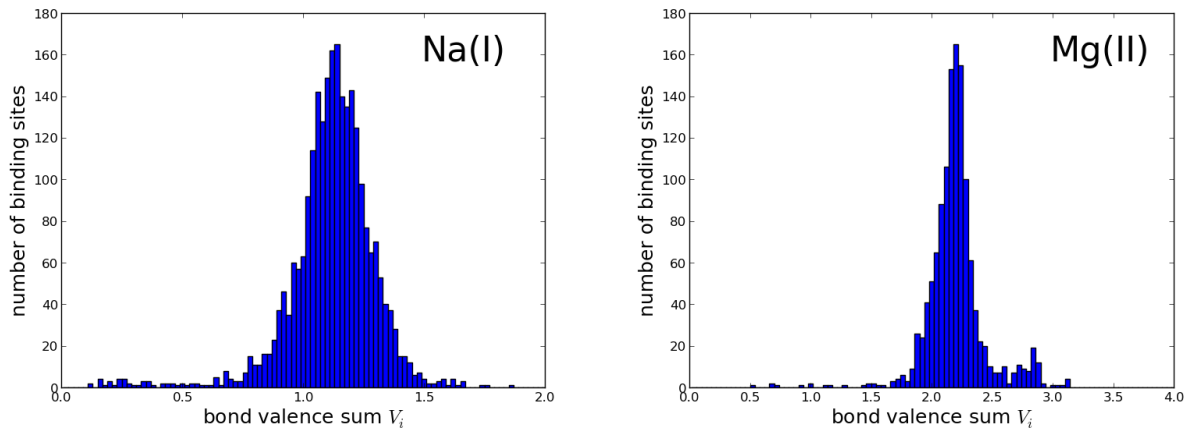


Figure 1: Initial bond valence sum distributions around expected oxidation state for the various central cations studied.

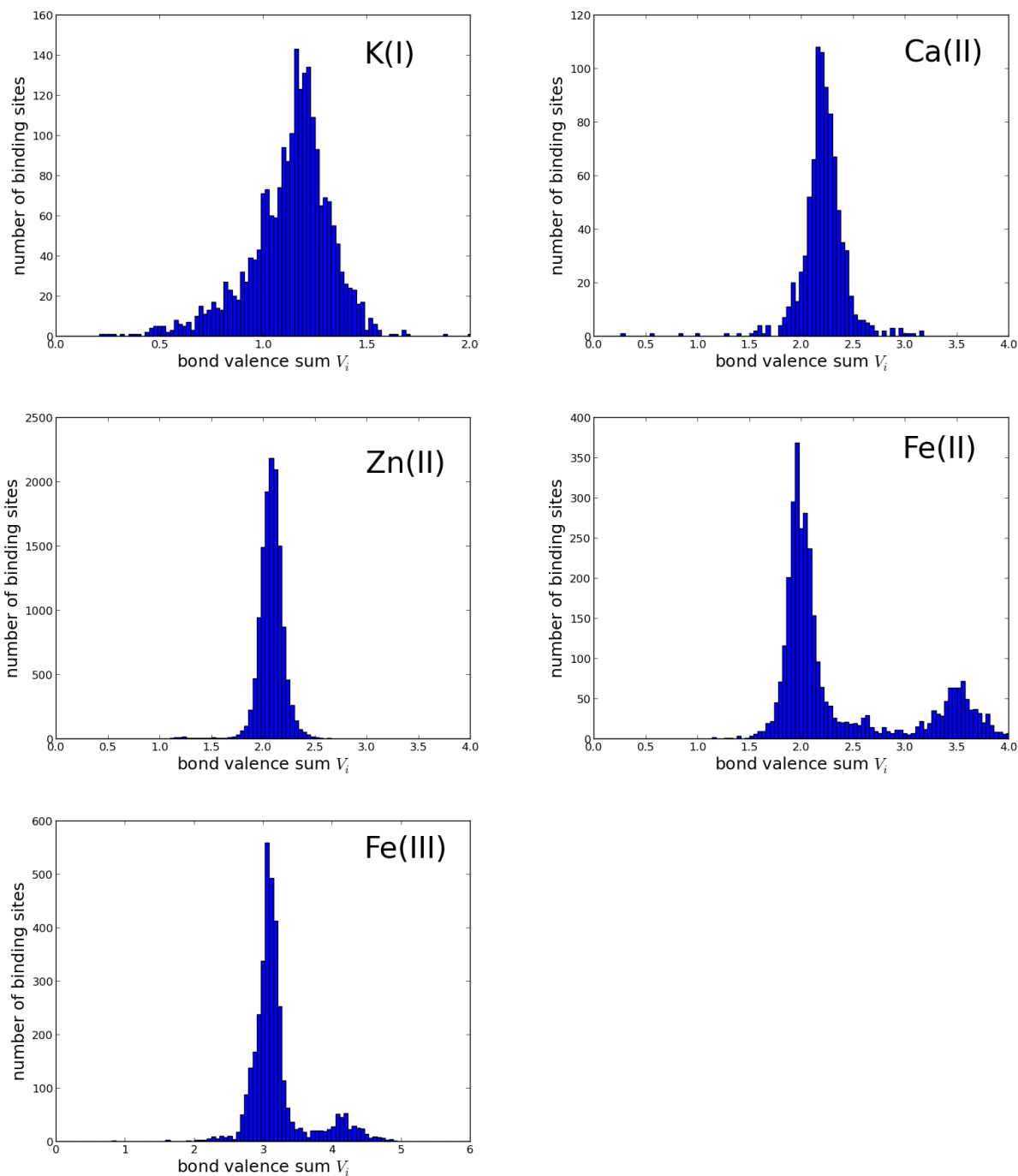


Figure 1: (continued) Initial bond valence sum distributions around expected oxidation state for the various central cations studied.

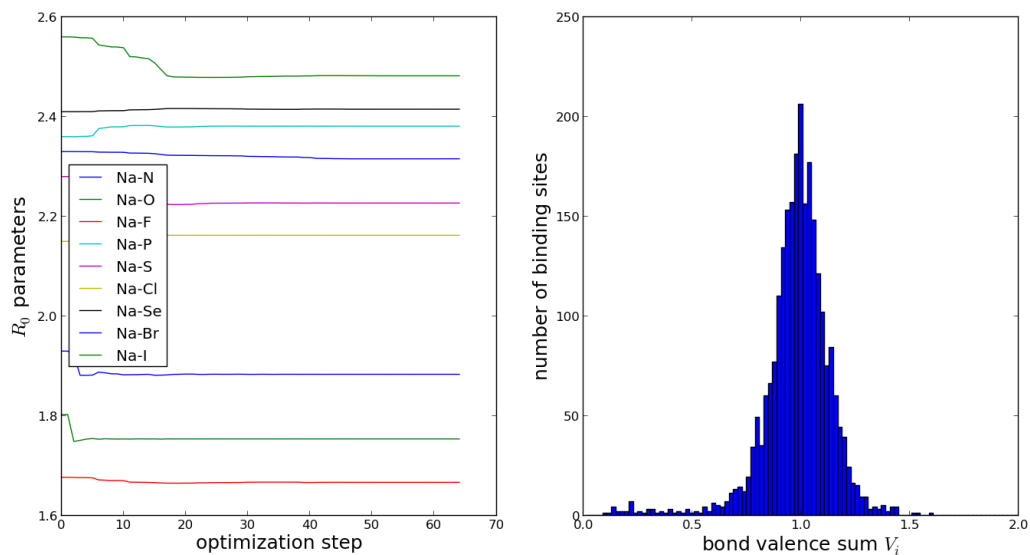


Figure 2: Optimization history (left) and final distribution of bond valence sums (right) for binding sites with Na as the central cation (expected oxidation state of 1).

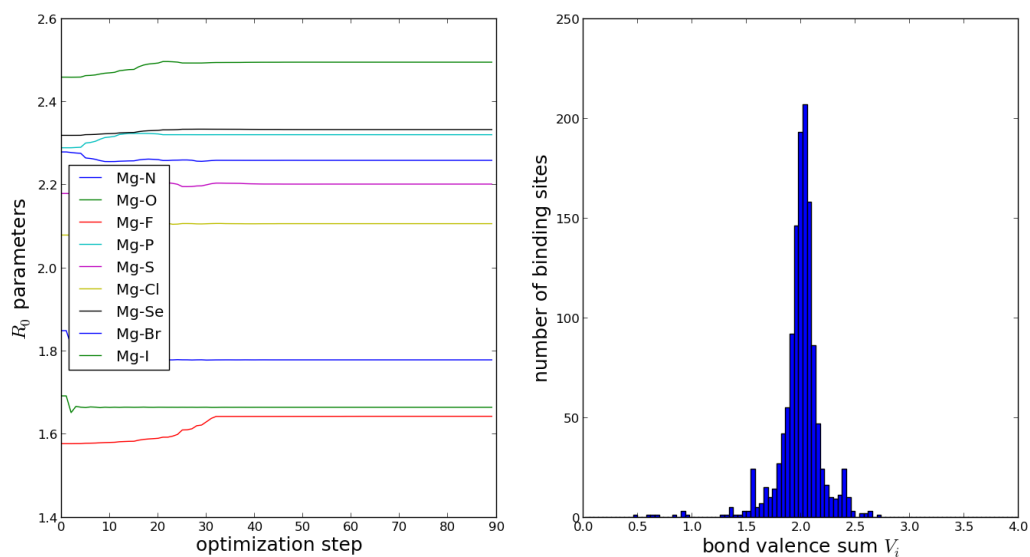


Figure 3: Optimization history (left) and final distribution of bond valence sums (right) for binding sites with Mg as the central cation (expected oxidation state of 2).

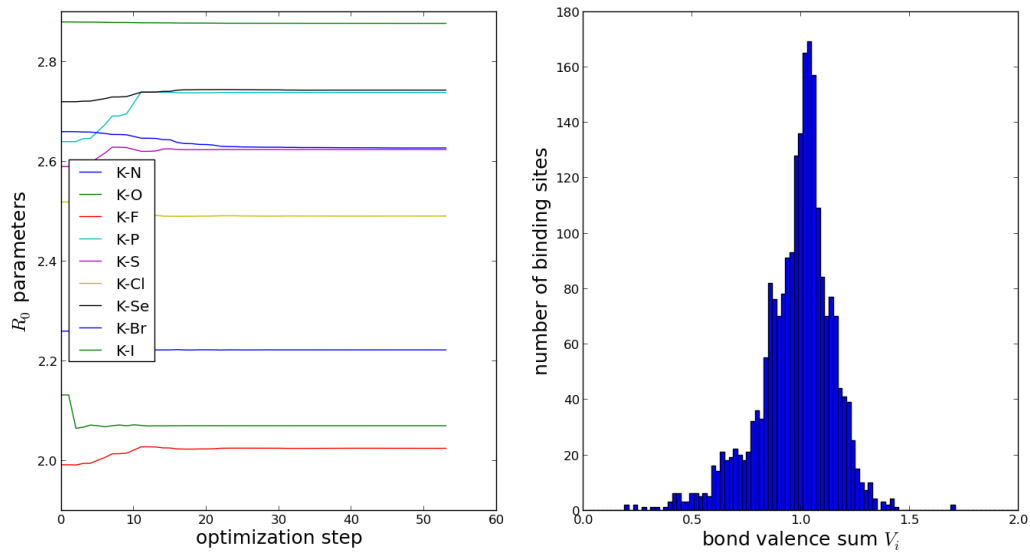


Figure 4: Optimization history (left) and final distribution of bond valence sums (right) for binding sites with K as the central cation (expected oxidation state of 1).

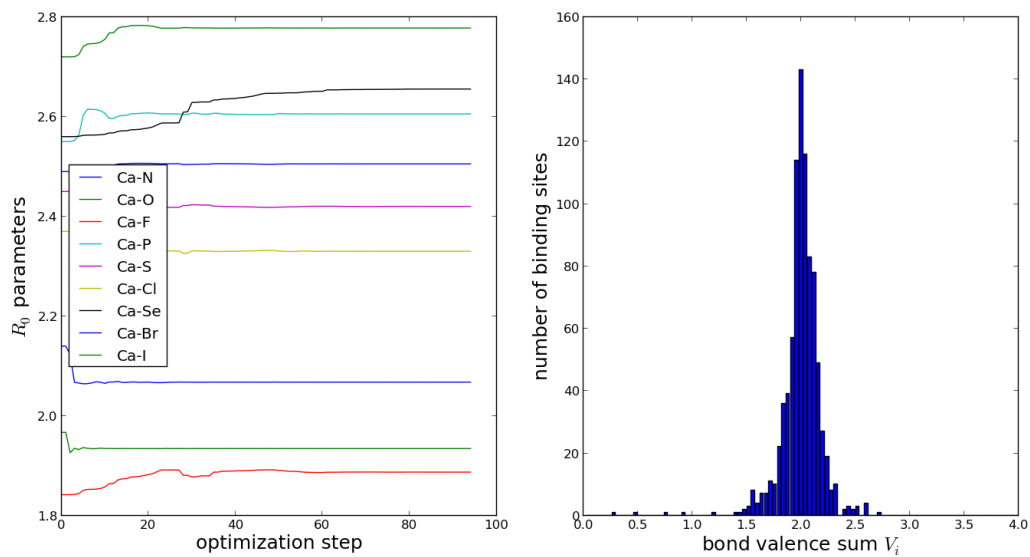


Figure 5: Optimization history (left) and final distribution of bond valence sums (right) for binding sites with Ca as the central cation (expected oxidation state of 2).

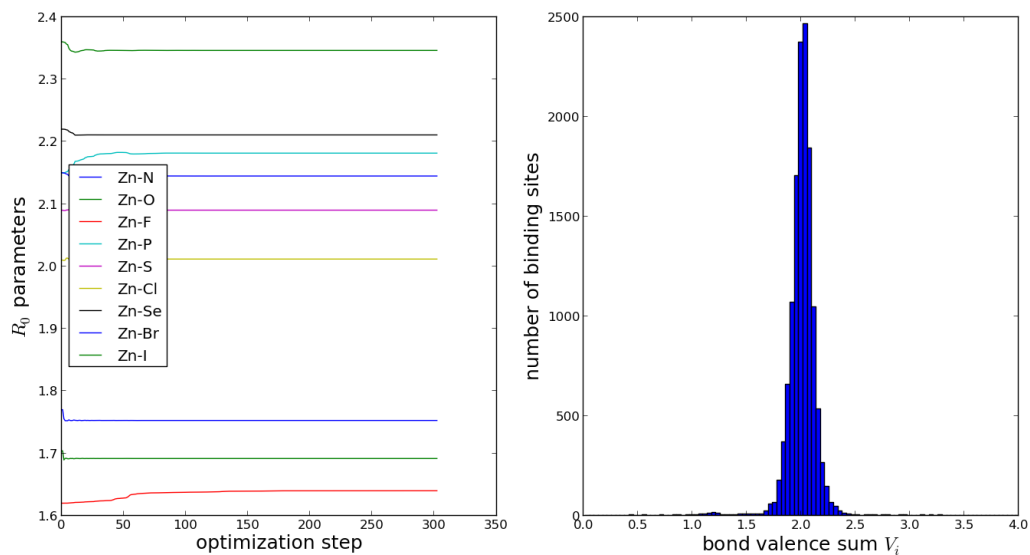


Figure 6: Optimization history (left) and final distribution of bond valence sums (right) for binding sites with An as the central cation (expected oxidation state of 2).

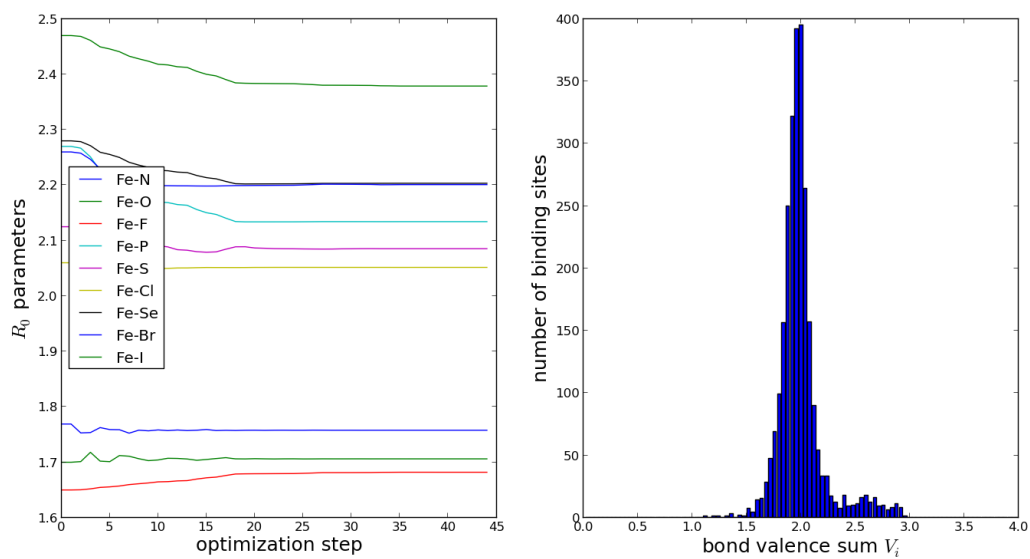


Figure 7: Optimization history (left) and final distribution of bond valence sums (right) for binding sites with Fe(II) as the central cation (expected oxidation state of 2).

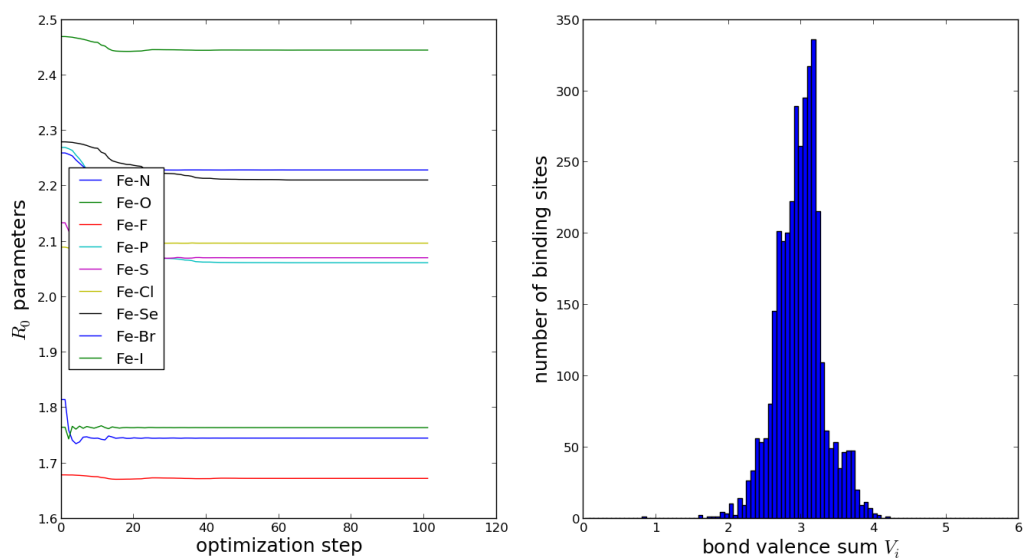


Figure 8: Optimization history (left) and final distribution of bond valence sums (right) for binding sites with Fe(III) as the central cation (expected oxidation state of 3).