Published in final edited form as:
*J Chem Theory Comput*. 2018 February 13; 14(2): 784–798. doi:10.1021/acs.jctc.7b01236.

# Development of CHARMM-compatible force-field parameters for cobalamin and related cofactors from quantum mechanical calculations

**Anna Pavlova**[†], **Jerry M. Parks**[‡], and **James C. Gumbart**[†,¶]

[†]School of Physics, Georgia Institute of Technology, Atlanta, GA 30332

[‡]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

[¶]School of Chemistry and Biochemistry, Parker H. Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332

## Abstract

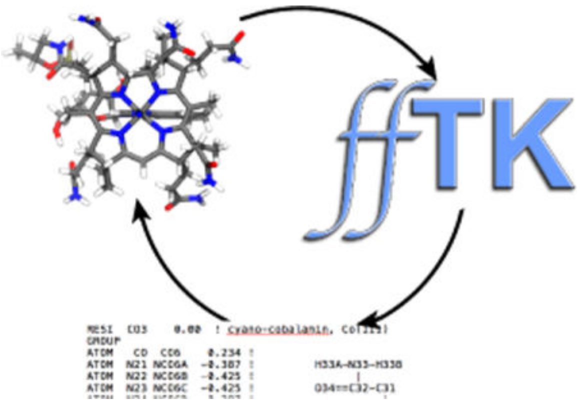
Corrinoid cofactors such as cobalamin are used by many enzymes and are essential for most living organisms. Therefore, there is broad interest in investigating cobalamin-protein interactions with molecular dynamics simulations. Previously developed parameters for cobalamins are based mainly on crystal structure data. Here, we report CHARMM-compatible force field parameters for several corrinoids developed from quantum mechanical calculations. We provide parameters for corrinoids in three oxidation states, $Co^{3+}$, $Co^{2+}$ and $Co^{1+}$, and with various axial ligands. Lennard-Jones parameters for the cobalt center in the Co(II) and Co(I) states were optimized using a helium atom probe, and partial atomic charges were obtained with a combination of natural population analysis (NPA) and restrained electrostatic potential (RESP) fitting approaches. The Force Field Toolkit was used to optimize all bonded terms. The resulting parameters, determined solely from calculations of cobalamin alone or in water, were then validated by assessing their agreement with DFT geometries and by analyzing molecular dynamics simulation trajectories of several corrinoid proteins for which X-ray crystal structures are available. In each case, we obtained excellent agreement with the reference data. In comparison to previous CHARMM-compatible parameters for cobalamin we observe a better agreement for the fold angle and lower RMSD in the cobalamin binding site. The approach described here is readily adaptable for developing CHARMM-compatible force-field parameters of other corrinoids or large biomolecules.

## Graphical abstract

---

Note to publisher. This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan(http://energy.gov/downloads/doe-public-access-plan).

**Supporting Information**
The Supporting Information is available free of charge on the ACS Publication website at http://pubs.acs.org/. It contains additional details of the fitting procedures, including assignment and/or optimization of parameters, as well as additional validation. CHARMM topology and parameter files along with instructions for their use are also included.



```
RESI  CO3     0.00  ! cyano-cobalamin, Co(III)
GROUP
ATOM   CO  CO6     0.234 !
ATOM  N21 NCO6A   -0.387 !            H33A-N33-H33B
ATOM  N22 NCO6B   -0.425 !                  |
ATOM  N23 NCO6C   -0.425 !            O34==C32-C31
```

---

## Introduction

Cobalamin, often called vitamin $B_{12}$, is a biological cofactor that is used by numerous enzymes to catalyze essential reactions in living organisms. Methylcobalamin (MeCbl), 5′-deoxyadenosylcobalamin (AdoCbl) and related corrinoids are used to perform diverse biochemical reactions including methyl transfer, isomerization and dehalogenation.[1,2] Among the most complex biological cofactors known, corrinoids consist of a tetrapyrrolic corrin ring with a central cobalt (Co) cation that can adopt the +1, +2 or +3 oxidation state (Scheme 1A). The corrin ring bears multiple methyl, acetamide and propionamide substituents, with one of the propionamide groups being fused through a propanolamine linkage to a long ribonucleotide tail. Microorganisms incorporate a wide range of compounds at the tail position, including substituted benzimidazoles, purines and phenolics,[3,4] but the most common tail substituent is the nucleotide 5,6-dimethylbenzimidazole (DMB) that is present in cobalamin. Besides the four equatorial N ligands from the corrin ring, Co can also bind up to two additional axial ligands, one each in the $\alpha$ and $\beta$ positions (see Scheme 1A for the complete structure of cobalamin and Scheme 1B for PDB atom numbering).[1,5–7] The nucleotide tail can serve as an intramolecular ligand at the $\alpha$ position, whereas methyl or adenosyl groups coordinate at the $\beta$ position to form unique organometallic Co–C bonds. The upper ($\beta$)-axial site can also be occupied by hydroxy (HO-), glutathionyl (GS-) or cyano (NC-) ligands.

Three main cofactor configurations with respect to the $\alpha$ position have been observed in corrinoid proteins. When the nucleotide tail is an intramolecular ligand to Co at the $\alpha$ position, the cofactor is said to be in a “base-on” configuration (or “DMB-on” when the nucleotide is DMB). In some corrinoid proteins, such as the corrinoid iron-sulfur protein (CFeSP) from the Wood-Ljungdahl pathway of acetogenic bacteria, DMB is dissociated from Co and no ligand is present at the $\alpha$ position.[8,9] In this case, the cofactor is said to bind to the protein in a “DMB-off” configuration. In other DMB-off corrinoids, a His residue from the protein serves as the $\alpha$ ligand to Co. This arrangement is referred to as “Hison” coordination and is exemplified by the cobalamin-dependent methionine synthase, MetH.[10,11] Recently, the corrinoid protein HgcA was predicted to bind its corrinoid cofactor in an unprecedented “base-off, Cys-on” configuration.[12]

The oxidation state of Co affects the metal coordination shell, with Co(III) complexes typically being hexacoordinate, Co(II) pentacoordinate and Co(I) tetracoordinate.[1,5] The Co–C bond in MeCbl and AdoCbl can be homolytically cleaved, resulting in a square-pyramidal Co(II) species, although square-planar Co(II) structures have also been reported.[9,13,14] Under strongly reducing conditions, Co(II) can be further reduced to Co(I), leading to the dissociation of the $\alpha$ ligand and a square-planar complex.

Because of their biological importance, numerous studies have been carried out to characterize the structure and function of corrinoids,[1,5,6,15] and several structures of $B_{12}$-dependent enzymes and cobalamin-binding proteins have been reported.[10,13,14,16–20] The three major classes of $B_{12}$-dependent enzymes are the AdoCbl-dependent eliminases and mutases, and the MeCbl-dependent methyltransferases, although corrinoids are also used by enzymes for many other reactions.[5,6,11,15] Recently, X-ray structures of two reductive halogenases have been determined. RdhA from *Nitratireductor pacificus* uses cobalamin to catalyze the reduction of halogenated phenolic compounds,[21] while PceA from *Sulfurospirillum multivorans* uses the corrinoid norpseudo-B12, which has an adenine tail instead of DMB and lacks one of the methyl groups that are present in cobalamin, to catalyzes the reduction of tetrachloroethene and trichloroethene.[22] The photoreceptor CarH from *Thermus thermophilus* uses base-off/His-on AdoCbl to sense light.[10] It has been shown that changes in the oxidation state of Co can induce structural changes in the protein to which it is bound.[13,14,18] However, it is not clear how these changes occur. Detailed knowledge of cobalamin-protein interactions, and how such interactions change under different Co oxidation states, is crucial for a better understanding of cobalamin-dependent enzymes.

Atomistic molecular dynamics (MD) simulations are frequently performed to obtain insight into protein-small-molecule interactions. However, accurate force-field parameters for both the protein and the small molecule are required for reliable results. Current additive force fields, e.g., AMBER[23] and CHARMM,[24] are highly accurate for reproducing protein and nucleic acid structures[25] and several general force fields such as CGenFF,[26] GAFF[27] and OPLS[28] can be used to describe ligands of interest. Unfortunately, some chemical groups, including metal cations, are typically not covered by general force fields. Thus, parameters must developed from quantum mechanical (QM) calculations, experimental data, or both. For example, parameters for heme,[29,30] photosystem II iron complexes,[31] chlorophyll ligands,[32] several small transition-metal complexes[33] and Zn-containing cofactors (Zinc AMBER Force Field, ZAFF)[34,35] have been developed from QM calculations. Recent software releases can facilitate fitting of force-field parameters from QM data. For example, visual force field derivation toolkit (VFFDT)[36] can be used to optimize bond and angle parameters for metal-containing systems, and the Force Field Toolkit (ffTK) plugin in VMD,[37] developed in our group, can be used to generate CHARMM-compatible parameters. Automatic optimization of vdW parameters is not presently available in any of the software packages. Instead, these parameters for the metal atom are typically taken from other work[29,30,32,34,35] or optimized from QM calculations.[33]

Several approaches are commonly used to describe ligand coordination to metal cations in atomistic force fields: the bonded model, non-bonded model, cationic dummy atom model and ligand field molecular mechanics (LFMM) model.[38–42] In the bonded model, explicit bond and angle terms are used to describe the interactions between the metal atom and the coordinated ligands. In the non-bonded model, these interactions are described using only Lennard-Jones (LJ) potential and electrostatic terms.[38] The cationic dummy model uses fractional charges around the metal center to mimic its valence electrons.[42] Although the non-bonded model reduces the number of bonded parameters that need to be determined and allows association and dissociation of ligands from the metal center, in practice, the metal often loses its coordination shell during simulations.[41,43] The cationic dummy model improves the stability of metal-ligand coordination compared to the non-bonded model. Non-bonded and cationic dummy models assume that LJ parameters and partial charges do not change significantly during ligand dissociation,[38,41] making them less suited for cases with changing metal oxidation states. Bonded models result in stable metal-ligand complexes, but ligand dissociation cannot be studied with such models.[41] The Ligand Field Molecular Mechanics (LFMM) model adds additional terms to the MM potential between the coordinating ligands and the metal atom that describe the ligand field stabilization energy. This model can be used to simulate ligand dissociation; however, it requires parametrization of the ligand field stabilization energy.[44] Since the general interest is in studying specific redox states of cobalamin with MD, the bonded model was the most suitable choice for the force-field parameters.

One of the challenges with optimizing force field parameters is that the resulting optimization problem is typically neither consistent nor determined.[45] Due to inconsistency it is not possible to obtain an exact fit to all of the target experimental and QM data in most cases. For example, fitting of the harmonic terms for bonds and angles (see Eq. 1 in Methods) aims to determine the equilibrium values and the force constants for each bond and angle in the molecule. Yet the MM equilibrium value and distortion energy for each bond or angle will depend not only on its parameters but also on the bonded parameters for adjacent atoms and the non-bonded interactions with other atoms in the molecule. Due to this interdependence it is not possible to exactly reproduce the target equilibrium values and distortion energies of all bonds and angles in the molecule simultaneously. Additionally, the arising optimization problem is often underdetermined with several solutions giving similar agreement with the target data. Because of the issues discussed above additional validation of the force field parameters is desired. Properties that are not used for fitting of the parameters should be used for validation.

Bonded force-field models of Co(III)-cobalamins for the MM2[46,47] and AMBER[48] force fields have been previously developed by Marques and coworkers. Crystal structures of isolated cobalamins were used to determine the parameters, and these structures were reproduced with high accuracy. Gumbart et al. adapted the AMBER parameters for cyanocobalamin for CHARMM by changing the partial atomic charges to the calculated natural population analysis (NPA) charges.[49] These parameter sets have been used in simulations of Co(III)-cobalamin in water[50,51] and bound to the protein BtuB.[49] Al though crystal structures are important for fitting and validating force-field parameters, comparisons to electronic structure calculations are presently lacking for cobalamins. Therefore, we applied the latest tools and procedures to develop CHARMM-compatible force-field parameters for cobalamin based on QM calculations. We used ffTK to construct bonded models of several important forms of cobalamin, including the octahedral adenosyl-, cyano- and methylcob(III)alamin, square-pyramidal cob(II)alamin and square-planar cob(I)alamin. Although previous work only considered cobalamins in the $Co^{3+}$ state, we have re-optimized several parameters for cob(II)alamin and cob(I)alamin to correctly capture the electrostatic and geometric changes caused by the decreased oxidation state of Co. The parameters were carefully validated by showing that they reproduce both DFT and crystal structure geometries. Comparison to simulations with previous CHARMM-compatible cobalamin parameters shows that our parameters increase the conservation of crystal-structure interactions and give better agreement with the cobalamin fold angle in those structures. The approaches described here and in our previous work[52] can be used to develop CHARMM-compatible force-field parameters for nonstandard corrinoids and other large biomolecules.

## Methods

### Development of force-field parameters

In the CHARMM force field, the potential energy function consists of bonded (Eq. 1) and non-bonded terms (Eq. 2). The input of the following force-field parameters is required: the force constants $K$; the equilibrium values of the bonded terms, $b_0$, $\theta_o$, and $\varphi_o$; dihedral phase $\delta$; partial atomic charges $q$; and LJ parameters $\varepsilon$ and $\sigma$.[24]

$$V_{\text{bonded}} = \sum_{\text{bonds}} K_b(b - b_o)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_o)^2 + \sum_{\text{dihedrals}} \sum_{n=1}^{n=6} K_{\phi,\text{n}}(1 + \cos(n\phi - \delta)) + \sum_{\text{impropers}} K_\varphi(\varphi - \varphi_o)^2 \quad (1)$$

$$V_{\text{nonbonded}} = \sum_{\text{pairs ij, charges}} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + \sum_{\text{pairs ij, Lennard-Jones}} \varepsilon_{ij}\left[\left(\frac{r_{\text{min, ij}}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{\text{min, ij}}}{r_{ij}}\right)^{6}\right] \quad (2)$$

Because parameters for several parts of the cobalamin molecule are already present in the CGenFF force field,[26] such as the amide and methyl side chains, the nucleotide loop, and the adenosyl axial ligand, parameters were not optimized for these groups. Instead, the corresponding CGenFF charges and atom types were used for these atoms, with the exception of the phosphate group and the ribose sugar of the nucleotide loop, for which CHARMM36 nucleic acid parameters[53] were used instead. QM calculations were used to determine the LJ and partial charge parameters for Co, the corrin ring and the Co-bound equatorial and axial ligands, as well as the bond, angle and dihedral terms arising from the side chains on the corrin ring. All LJ parameters were taken by analogy except for Co(II) and Co(I), which were optimized here (see the Supporting Information for the assignment of LJ parameters by analogy). Intially, all improper dihedral parameters for the corrin ring were taken from the previously developed heme parameters for the CHARMM force field.[29] The force constants for the impropers were gradually increased to better reproduce experimental corrin fold angles (see Validation and Supporting Information).

To determine the force field parameters for cobalamin, which contains nearly 200 atoms, we adopted a fragment-based approach in which we used representative models for the QM calculations. These models include the unsubstituted corrin ring, Co, and various axial ligands. DMB was replaced with imidazole, and the adenine base was removed from the Ado ligand (Scheme 2A). The smaller models were used to determine the partial atomic charges and the bonded terms for the corrin ring and the axial ligands. Five smaller models were used, one for each state and each $\beta$ ligand: MeCbl(III), AdoCbl(III), CN-Cbl(III), Cbl(II) and Cbl(I). Additionally, we used two slightly larger model systems, referred to as ExtAD and ExtBC, to determine the bonded parameters resulting from the attachment of the amide side chains and the methyl groups to the smaller model (Scheme 2A). New atom types were introduced for the heavy atoms of the corrin ring, Co, and the carbon in the Co–C bond (see the topology file in the Supporting Information). We aimed to keep the number of new atom types to a minimum, while still correctly reproducing the asymmetry of the corrin ring, which required nine different atom types for the heavy atoms.

## Quantum mechanical calculations

Single-reference correlated wave function methods such as MP2 are known to perform poorly for describing transition metal-containing systems.[54] Multireference methods are often highly accurate, but are computationally extremely costly for large systems. Fortunately, density functional theory provides a good balance between accuracy and computational efficiency for these systems. The BP86[55–57] density functional has been shown to reproduce the experimentally observed geometries and various properties of corrinoids accurately.[58–62] Thus, we used the BP86 density functional along with the Def2-SVP basis set[63,64] to optimize the geometries of the model complexes. Recently, however, it was shown that PBE energies (computed at BP86 geometries) provide superior accuracy compared to BP86 and other density functionals for computing reduction potentials, p$K_a$s and Co-ligand binding free energies for aquacobalamin.[65] Here, we followed a similar approach in which we computed interaction energies with the PBE density functional.

Default convergence criteria were used for the geometry optimizations, and no constraints or restraints were applied during any of the geometry optimizations. All DFT calculations were performed with Gaussian 09.[66] With the exception of the small AdoCbl(III) model, no imaginary frequencies were found for the optimized geometries. In the AdoCbl(III) model, a single imaginary frequency of $-18$ cm$^{-1}$ was found, but it involved several angles of the deoxyribose ring that were not parameterized.

## Lennard-Jones parameters

To determine the LJ parameters for Co in the Cbl(I) and Cbl(II) cobalamins, we used the He atom probe approach developed by MacKerell and coworkers.[67,68] Simplified, corrin-based models were used for these calculations. In each case, the total charge of the system was zero. First, the geometry of each complex was optimized without the He probe at the BP86-D3/Def2-SVP level of theory. For the Cbl(II) model, HO-Co(II)-corrin, a He atom was then placed along the Co–O axis on the corrin face opposite the hydroxy group (i.e., He ⋯ Co–O angle = 180). For the Cbl(I) model, Co(I)-corrin, the He atom was placed along a vector normal to the plane defined by atoms in the corrin ring corresponding to C6, N22, C9, C10, C11, N23 and C14 in cobalamin (Scheme 1B). In each case, the He ⋯ Co distance was varied in increments of 0.1 Å and single-point energies were computed with three density functionals: BP86,[55–57] PBE[69] and B3LYP,[70,71] and three basis sets: Def2-SVP, Def2-TZVP, and Def2-TZVPP.[63,64] D3 dispersion corrections[72] with Becke-Johnson damping[73] were included in all cases. The resolution-of-the-identity (RI) approximation with the corresponding Coulomb density fitting basis sets.[54] was used to accelerate the evaluation of two-electron integrals. The corresponding non-bonded molecular mechanical (MM) interaction energies were calculated with the NAMDenergy plugin in VMD.[74]

## Partial atomic charges

We considered various approaches for determining partial atomic charges. We examined the possibility to fit the charges based on either QM interactions with water molecules[26] or the restricted electrostatic potential (RESP method),[75] which are the standard approaches for the CHARMM and AMBER force fields, respectively.[23,24] Additionally, we looked into assigning charges based on either Mulliken[76] or natural population analysis (NPA).[77] Ultimately, we decided to use a combination of NPA and RESP for the assignment of partial charges (see Results). RESP could not be used on all atoms because it is often problematic for buried atoms.[75,78–81] For example, when we used only RESP for the MeCbl(III) model, a physically unrealistic negative partial charge was obtained for Co.

NPA was used to determine the charges for the Co, N21-N24 and N3B atoms. Standard CHARMM charges of +0.09 and +0.015 were used for aliphatic and aromatic hydrogens, respectively, for consistency and compatibility with the CHARMM force field. For the rest of the atoms in the corrin ring, the charges were determined with the RESP approach using AmberTools 14[82] with the Co and N charges constrained to their NPA values. Twofold symmetry was imposed on the charges along the axis defined by Co and C10 (see Scheme 1B for atom numbering) in all RESP calculations because these two parts of the molecule are nearly identical. For the NPA charges, the average charge of N21 and N24 or N22 and N23 were used to maintain symmetry. When we added the side chains to the small corrin model or the adenosyl group to the MeCbl(III) model, the charges of the hydrogens that had been removed were absorbed into the adjacent carbons, ensuring an integer charge of the new molecule.

## Bonded parameters

ffTK was used to optimize all bond, angle and dihedral parameters. QM geometries were used to determine the target bond and angle equilibrium values, and the Hessian in internal coordinates was used to optimize the bond and angle force constants. The Hessian is the matrix of secondary derivatives of the potential energy as a function of the input atom-pair coordinates, and it was used to construct the QM potential energy surface (PES) arising from the displacement of each internal coordinate. The optimized MM geometry and the molecular mechanics (MM) PES were obtained for trial parameters and compared to the corresponding QM ones. The objective function described by Mayne et al.[37] was used to measure the quality of the fit of the MM equilibrium bond and angle values and energy changes due to geometric displacements. The fitting procedure consisted of using either the simplex or simulating annealing method, implemented in ffTK, to minimize the objective function. Several refinement runs of this procedure were performed until the objective function stopped decreasing with additional iterations.

Relaxed dihedral scans were used to determine the target QM PES for fitting the dihedral constants. Corrin ring dihedrals were scanned 60° in each direction in 10° increments. The dihedral scan around the Co–N3B bond, which describes the rotation of the imidazole with respect to the corrin ring, was scanned 180° in each direction in 15° increments. All other dihedrals were scanned 90° in each direction in 15° increments. The optimization of the dihedral force constants was done with the simulated annealing protocol developed by Guvench and MacKerell[83] and implemented in ffTK.[37] The fitting of dihedral terms is often highly undetermined, and several sets of dihedral parameters may fit the same PES.[45,52] Additionally, the optimization problem is ill-conditioned and the algorithm can produce unrealistically high force constants when it attempts to fit a small peak in the PES.[45] These issues can lead to overfitting and unrealistic high-energy geometries, e.g., nonplanar aromatic structures or eclipsed hydrogen conformations in the methyl groups during MD simulations.[52] Thus, to avoid overfitting we sought to reduce both the number of fitted dihedrals and the multiplicity of each dihedral. The bonded model of cobalamin results in numerous dihedral terms arising from the creation of five Co–N bonds, which increases the likelihood of overfitting. Hence, the force constants were set to zero for most of these dihedrals (see the Results section for details), and higher multiplicities were only used for a dihedral if they significantly improved the fit to the QM PES. Additionally, dihedrals containing hydrogen atoms were not optimized during the dihedral scans. Instead, they were either set to zero or taken from the CGenFF force field if a good analogue was found. Only QM geometries less than 10 kcal $mol^{-1}$ above the energy minimum were used for the QM and MM energy comparison.

## Molecular dynamics simulations

All MD simulations were performed with NAMD 2.9.[84] In addition to our optimized cobalamin parameters, we used the CHARMM36 protein,[85] nucleic acid[53] and lipid[86] force fields and the CHARMM TIP3P model for water.[87] The temperature and pressure were controlled with the Langevin thermostat and piston, respectively. The target temperature was 310 K and the target pressure was 1 bar. The Langevin piston period was 200 fs and the piston decay was 100 fs. To allow integration of the equations of motion with a 2-fs time step, we used rigid bonds for all covalently bound hydrogens. A 12-Å cutoff was used for the van der Waals interactions and a switching function was applied from 10 to 12 Å to ensure a smooth decay to zero. The particle-mesh Ewald method[88] was used to calculate long-range electrostatic interactions. We validated our parameters for cobalamin-protein and cobalamin-RNA interactions starting with the corresponding crystal structures, 1N4A, 3GAI, 5C8A, 4GMA and 1NQH. The first four structures were solvated in bulk water, while the 1NQH BtuB structure was embedded in DMPE membrane using CHARMM-GUI.[89] The Solvate and Autoionize plugins in VMD were used to add bulk water and ions to the systems, respectively. The systems for 1N4A, 3GAI, 5C8A, and 4GMA were ionized with NaCl, while the system for 1NQH was ionized with $CaCl_2$; a salt concentration of 0.15 M was used in all cases. For all systems, the energy was minimized before the MD simulations, and a two-step equilibration was applied. First, water and ions were equilibrated for 500 ps, after which cobalamin and the protein side chains were equilibrated for an additional 1.5 ns. Additional equilibration steps were added for the membrane system. After equilibrating water and ions, lipid tails and then the whole membrane were equilibrated for 1.5 ns each before proceeding with the equilibration of the protein side chains. All production runs were 10 ns long. Because the positions of multiple adjacent residues were missing in the 4GMA structure, we did not include these residues in the simulations. Instead, we included only those residues that were present in the PDB structure and applied harmonic restraints with a force constant of 10 kcal $mol^{-1}$ $Å^{-2}$ to those residues farther than 15 Å from cobalamin in the crystal structure.

To calculate the corrin fold angle, we first fit two planes to groups of atoms C4 C5 C6 C9 C10 N21 N22 and C10 C11 C14 C15 C16 N23 N24, respectively (Scheme 1B).[2] To fit the plane, the covariance matrix of each set of atoms with respect to their centroid is first calculated; the normal vector of the fit plane is then the cross product of the first two (largest) eigenvectors of this matrix. Finally, the fold angle is calculated between the two fit planes' normal vectors.

# Results

## Lennard-Jones (LJ) parameters

The LJ parameters for Cbl(II) and Cbl(I) were determined using the He probe approach developed by MacKerell and coworkers.[67,68] The non-bonded interactions between He and the cobalamin models were calculated for a range of Co-He distances, using three different DFT functionals: PBE, BP86 and B3LYP. For the small cobalamin models (Scheme 2A) we used a $\beta$-hydroxy ligand for the $Co^{2+}$ state and no axial ligands for the $Co^{1+}$ state. The He probe approach was not used for the hexacoordinated Cbl(III) models because all Co coordination sites are occupied. Hexacoordination of Co greatly limits intermolecular interactions of Co, and the optimization of bonded parameters was expected to correct for any potential intramolecular interaction errors. Therefore, the LJ parameters for all Cbl(III) models were set to zero.

Basis set convergence for the QM interaction energies between a He atom and the cobalamin models Cbl(I) and HO-Cbl(II) was tested in the Supporting Information for the three considered DFT functionals. Figure S1 shows that the energy differences between Def2-TZVP and Def2-TZVPP basis sets are very small for all cases. Figures 1A and 1D display QM interaction energies between a He atom and the two cobalamin models calculated with the three DFT functionals using Def2-TZVPP basis set and D3 correction. We also calculated the corresponding non-bonded molecular mechanical (MM) interaction energies for both cobalamin models with the $\varepsilon$ value for Co set to zero (MM Start). The difference between the QM and MM interaction ($En_{Diff}$) with this potential was used to determine the LJ parameters. Because it is usually not possible to reproduce the QM PES with the functional form of the LJ potential over all distance ranges,[67,68] we aimed to reproduce the minimum QM interaction energy and distance for the Co–He interaction by adjusting $\varepsilon$ and $r_{min}$ from Eq. 2. The differences between QM energies at different levels of theory and MM Start energies are plotted in Figures 1B and 1E. Improving the agreement between the QM and MM energies requires fitting an LJ potential to these differences. The shape of the LJ potential is determined by $\varepsilon$, which governs the well depth, and $r_{min}$, which governs the location of the minimum. All atoms have attractive vdW interactions, so the LJ potential should always have a minimum below zero. As shown in Figures 1B and 1E, the difference in the energy curves between QM and MM Start has a minimum below zero for both Cobl(I) and Cbl(II) states for only the PBE functional. Thus, this level of theory was selected to determine the LJ parameters for Co.

We chose the $\varepsilon$ and $r_{min}$ for the He ··· Co LJ potential based on the minimum energy and distance for the calculated $En_{Diff}$ potential. The Co-specific $\varepsilon$ and $r_{min}$ parameters were then derived from the Lorentz-Berthelot combination rules used in the CHARMM force field[24]

$$\varepsilon_{Co} = \frac{\varepsilon^2_{He-Co}}{\varepsilon_{He}} \text{ and } \frac{r_{min,\,Co}}{2} = r_{min,\,He\,-\,Co} - \frac{r_{min,\,He}}{2}. \quad (3)$$

Table 1 compares the minimum interaction energies and distances from QM calculations to the corresponding MM values before and after the fit, and shows the final LJ parameters for Co. AMBER-compatible LJ parameters of similar magnitude were previously reported for other $Co^{2+}$-containing complexes.[33] Figures 1C and 1F compare the QM and MM energies at all computed ranges and show that MM Start potential energies were too favorable and the energy minimum distances were too short compared to the corresponding QM energies and distances. The addition of LJ parameters to the Co atom in MM Fit significantly improved the agreement between QM and MM energies, although the new MM potential energies became somewhat too repulsive at distances lower than the QM minimum.

## Partial Atomic Charges

Several approaches for determining partial atomic charges were considered. Although optimization of partial atomic charges based on reproducing QM interaction energies with water molecules is recommended for CHARMM-compatible parameters,[26] such an approach was not possible for the octahedral cobalamin complex because steric clashes prevent the water molecules from interacting with Co and the nitrogens N21-N24. An alternative approach for determining partial atomic charges is the RESP method, which is used in the AMBER force field and is based on fitting atom-centered charges to the QM electrostatic potential.[23,75] However, this method often struggles with determining accurate charges of buried atoms,[75,78,79,81] such as Co and N21-N24 in cobalamin. We initially used the RESP method to fit all the charges for the MeCbl(III) model (Table S2), but we obtained unrealistic charges for several atoms. Specifically, Co was assigned a negative charge and N21-N24 were assigned positive charges. Although we obtained the expected signs for the charges of these atoms in the Cbl(II) model (positive for Co and negative for the nitrogens), the magnitude of these charges was closer to zero than expected. We concluded that the presence on the axial ligands made the charges of the Co atom and N21-N24 highly underdetermined in the RESP approach.

Mulliken population analysis has also been used to obtain partial atomic charges.[76] Although Mulliken charges are easily determined even for buried atoms, they are highly sensitive to the chosen basis set.[76] For example, the Mulliken charges for Co in the MeCbl(III) model calculated at the BP86 level of theory with the Def2-SVP and 6-31G(d) basis sets are 0.22 and 0.87, respectively. Natural population analysis (NPA), which is based on the electron density distribution obtained by transforming molecular orbitals into localized, one-center natural atomic orbitals,[90] yields more consistent atomic charges with regard to the basis set than Mulliken population analysis.[77,91] The natural atomic charge for atom A in a molecule, $Q_A$, is obtained by subtracting the sum of the natural populations for that atom, $p_k^{(A)}$, from its nuclear charge, $Z_A$, i.e. $Q_A = Z_A - \sum_k p_k^{(A)}$.

In the present work, we decided to use NPA in combination with the RESP method to determine the partial charges. First, we restrained only the charge of Co to its NPA charge and used RESP for all other heavy atoms, including N21-N24. The resulting charges had the correct sign for Co and N21-N24 (Table S3), but the magnitude of the charges for the nitrogens differed significantly in response to changes in the Co oxidation state. These large differences were not observed for the NPA charges, which suggests that the nitrogens were buried too deeply in the Cbl(II) and Cbl(III) models for a reliable RESP treatment of their charges. Therefore, in all the cobalamin models we used NPA for Co and the nitrogens N21-N24 and N3B; RESP was used for all other heavy atoms in the corrin ring.

The charges were re-optimized for all heavy atoms for Cbl(II) and Cbl(I) models. For the CN-Cbl(III) model we started with the MeCbl(III) charges and re-optimized only the charges on Co, the cyano group and the nitrogens N21-N24 and N3B. Because the NPA charges for Co and N21-N24 in the AdoCbl(III) model were nearly identical to those in the MeCbl(III) model (Table S4), we concluded that no re-optimization of charges was necessary for the analogous atoms in the AdoCbl(III) model. The final charges for Co and the Co-bound nitrogens for all cobalamin models are displayed in Table 2.

## Bond and angle parameters

Bond and angle parameters were fitted simultaneously as described in the Methods section. To reduce the number of fitted parameters, we initially attempted to impose the same molecular symmetry for the bonds and angles as for the partial atomic charges (see Methods). We refer to this fitting approach as Fit 1. However, Fit 1 did not correctly reproduce the geometric asymmetry of the QM model of MeCbl(III) (Figure S2A). Therefore, we introduced four different atom types (instead of two) for the four corrin ring nitrogens (N21-N24), allowing for differences in bonds and angles parameters between N21 and N24, as well as N22 and N23 (Fit 2). Fit 2 resulted in better agreement with QM geometries (Figure S2A). All target QM bond and angle values were reproduced to within 0.04 Å and 5°, respectively, which is close to the recommended maximum deviations of 0.02 Å and 3° for the CGenFF force field.[26]

The MeCbl(III) model was used to fit the initial bond and angle parameters. Only the bonds and angles containing Co or N21-N24 were refit for the Cbl(II) and Cbl(I) models. The structural differences between the MeCbl(III) and Cbl(II) QM geometries are concentrated at these atoms (Figure S2B). Although the Cbl(I) model also shows structural differences compared with MeCbl(III) at the edges of the corrin ring (Figure S2B), these differences were addressed by refitting the relevant dihedrals. For the CN-Cbl(III) and AdoCbl(III) models only the parameters involving the two axial ligands and the Co-bound carbon were refitted, respectively. The parameters correctly reproduce the changes in bonds and angles for the three different Co oxidation states (Tables 3 and 4).

Selected bond and angle parameters for the different Co states are shown in Tables S5 and S6, respectively. The bond force constants are relatively weak for the bonds between the cobalt atom the corrin ring (ranging from 30 to 70 kcal mol$^{-1}$ Å$^{-2}$), whereas the bond between Co and the imidazole N (i.e., Co–N3B) is significantly stronger ($K_b$ = 225.8 and 122.7 kcal mol$^{-1}$ Å$^{-2}$ for Cbl(III) and Cbl(II) states, respectively). The force constant for the N23–Co–N3B angle is 222.4 kcal mol$^{-1}$ Å$^{-2}$, while the corresponding force constants for the other three corrin ring nitrogens (N21, N22 and N24) are significantly smaller, ranging from 90 to 98 kcal mol$^{-1}$ Å$^{-2}$ (Table S6). These differences in force constants for chemically similar atoms suggest that some overfitting was introduced due to the use of four atom types for the corrin nitrogens instead of two. Nevertheless, it would not be possible to reproduce the asymmetry of the QM geometry without the introduction of additional atom types. Both bond and angle constants tend to decrease with reduced Co oxidation state, in agreement with previous results by Šebesta et al. for other Co-containing complexes.[33]

## Dihedral parameters

As discussed in the Methods section, many dihedral terms of the corrin ring were set to zero to avoid overfitting. The following dihedrals were optimized from QM dihedral scans: rotation of the imidazole ligand along the Co–N3B bond and bending of this ligand relative to the plane of the corrin ring (imidazole dihedrals), the dihedrals of the five-membered rings containing N21 and N24 (ring dihedrals), the rotation of the Ado ligand along the Co–C bond (Ado dihedrals) and the dihedrals arising from the connection of the side chains to the corrin ring (connection dihedrals) (Figure S3A). Figures 2, S4 and S5 compare the QM and MM PESs for the dihedral scans and Table 5 shows the root mean-square error (RMSE) for the MM PES. Although the RMSE was relatively high in some cases (> 1 kcal mol$^{-1}$), the higher deviations from the QM PES were mainly for the high-energy configurations, whereas the low-energy configurations were reproduced to within ~0.5 kcal mol$^{-1}$, as recommended for the CGenFF force field.[26]

There are eight dihedral terms that are centered on the Co–N3B bond. Initially, we fitted all of these dihedral terms, which resulted in a distortion of the planar imidazole ring during MM geometry optimization. Therefore, the number of terms fitted for imidazole dihedrals was reduced to avoid overfitting. We found that fitting only two dihedrals (i.e., N21–Co–N3B–C8B and N22–Co–N3B–C8) using multiplicities of $n = 4$ and 2, respectively, with the constants for the other six dihedrals set to zero, was sufficient to obtain the desired agreement between the QM and MM PES without distorting the imidazole ring (Figure 2).

To minimize overfitting, we initially did not fit the dihedrals of the corrin ring. However, MD simulations of the MeCbl(III) model in the gas phase showed significant fluctuations for some of these dihedrals, and comparison of the QM and MM PESs for these dihedrals revealed some discrepancies (Figures 2 and S6). Fitting of these dihedrals improved the agreement between the QM and MM PES (Figure 2) and reduced the fluctuations of these dihedrals during MD simulations (Figure S6). Agreement between QM- and MM-optimized geometries also increased. Ring, imidazole, and connection dihedrals were fit using the MeCbl(III) model and were only refitted for other models to improve the agreement with the QM geometries (see Supporting Information). Connection dihedrals were optimized using models ExtAD and ExtBC, and Ado dihedrals were fitted using the AdoCbl(III) model (Figures S4 and S5).

## Validation

The overall agreement between optimized MM and QM geometries for the small cobalamin models described in the Methods section was evaluated. Also, the optimized parameters were carefully validated according to two metrics: comparison of corrin fold angles in MD simulations of free and bound cobalamins to the angles reported experimentally[92] and the stability of bound cobalamins in their binding sites.

Comparison of minimized MM and QM geometries of the small cobalamin models showed that our force-field parameters reproduced the overall shape of the QM geometries very well (Figures 3 and S7). Almost no deviations were found for the Cbl(I) model, while the deviations were very small for the other four models and were not expected to affect the quality of the simulations. The RMSDs between the QM and MM geometries were less than 0.12 Å in all cases (Table 6).

Crystal structures of four proteins and one RNA co-crystallized with cobalamin were selected to validate the force field parameters. These structures included the periplasmic cobalamin-binding protein BtuF from *Escherichia coli* with base-on CN-Cbl(III) (PDB entry 1N4A),[17] a cobalamin-dependent riboswitch with base-on AdoCbl(III) (PDB entry 4GMA),[19] an adenosyltransferase from *Lactobacillus reuteri* complexed with Cbl(II) and ATP (PDB entry 3GAI),[14] the $B_{12}$-dependent photoreceptor CarH from *Thermus thermophilus* with base-off/His-on (PDB entry 5C8A)[10] and the $B_{12}$ transporter protein in *Escherichia coli* BtuB (PDB entry 1NQH).[93] Two 10-ns simulations were performed for each crystal structure. Additionally, one 10-ns-long simulation of free, solvated cobalamin was done for each of the five considered states: MeCbl(III), CN-Cbl(III), AdoCbl(III), Cbl(II) and Cbl(I).

Averaged cobalamin fold angles were calculated for simulations of both bound and free cobalamins and compared to values from the initial crystal structures and experiments,[92] respectively. The corrin fold angle was determined by dividing the conjugated part of the corrin ring in two along the Co-C10 line and calculating the angle between the two resulting planes (see Methods for details). As shown in Tables 7 and 8, the fold angles were too large compared to the target values during the initial MD simulations ($MD_{In}$); in several cases, the difference was larger than 10°. Hence, we decided to adjust our parameters to decrease the fold angles. It was found that increasing the force constants for all impropers that involved the nitrogens N21-N24 was an effective way to decrease the fold angle with minimal changes to the minimized MM geometries. Therefore, the force constants for these impropers were gradually increased to obtain the best agreement with target fold angles for free cobalamins (see Supporting Information for details). We used the simulations of cobalamins in water to fit the improper constants to target experimental data in Table 7. Next, the simulations of the crystal structures with the new improper parameters were used to validate the transferability of the parameters to a protein environment. After fitting the improper constants, the fold angles were significantly smaller in MD simulations ($MD_{Fit}$) and in better agreement with target values for both aqueous cobalamins and those in protein and RNA environments (see Tables 7 and 8). Some deviations of the fold angle were still observed for simulations of structures 1N4A, 1NQH and 5C8A. For all other cases the angle was within 3° of the experimental value. Adjustments of the improper constants resulted in very small changes to the overall MM geometries and MM bonds and angles. RMSD for the geometries before and after the changes was less than 0.02 Å, while the changes in bonds and angles were less than 0.01 Å and 3°, respectively. The MM-minimized geometries in Figures 3 and S7, as well as MM bonds and angles in Tables S5 and S6, are all after refitting the improper force constants.

The stability of cobalamin in bound structures was evaluated based on RMSD from the crystal structure position and preservation of important interactions between cobalamin and the protein or RNA. For each system, two key interactions, either within the cobalamin or between it and the protein or RNA bound to it, were chosen and monitored over time (Table 9). For 4GMA and 1NQH, we selected two hydrogen bonds between cobalamin and the RNA or protein. For 1N4A and 3GAI, we selected one hydrogen bond and one vdW interaction. To validate the cobalamin vdW interactions, we monitored the distance between DMB in cobalamin and W63 in 1N4A, which is important for cobalamin binding in BtuF.[17] To assess our Cbl(II) LJ parameters, we monitored the distance between Co and ATP in 3GAI. For 5C8A, we selected one hydrogen bond between the protein and cobalamin and an internal cobalamin hydrogen bond. Only chains A and B from the CarH tetramer in 5C8A were included in the simulation.

The fluctuations of the monitored distances and RMSD are shown in Figures 4 and 5. Overall, the RMSD is low (~ 3 Å or less after alignment of the protein/RNA), and the monitored distances are mostly conserved in all simulations, indicating that cobalamin remained stably bound in its binding site. Only temporary breaking of the monitored hydrogen bonds was observed in most cases, with the exception of distance 2 for 5C8A. This distance is broken and reformed very often for Chain B in both simulations (see Figure S8) and is broken for Chain A in simulation 2 after 4 ns. This interaction may be less stable because only two of the homodimers of the protein tetramer were included in the simulation. The measured distances for the vdW interactions (distance 2 for 1N4A and 3GAI) showed very small fluctuations (~ 1 Å).

Previously, only cobalamin parameters for CN-Cbl(III) have been adapted for simulations with the CHARMM force field.[49] The two crystal structures containing CN-Cbl(III), 1N4A and 1NQH, were also simulated using these older CHARMM-compatible parameters. Figure 5 compares RMSD and conservation of Distances 1 and 2 between the old and the new force field parameters. For 1N4A one of the two simulations with the old parameters resulted in a high RMSD and significant increases in the monitored distances, suggesting that cobalamin was not stable in its binding site. The other simulation with the old parameters looked very similar those performed with new parameters, yielding low RMSD and conservation of both interactions. Simulations of 1NQH showed less dramatic differences for the two sets of parameters, although fluctuations in Distance 2 were more frequent with the older parameters. The fold angle was also larger and more distinct from the target values for the old parameters than the new ones for both structures. This angle increased from 10.7 to 13.6 (deg.) and from 12.3 to 20.9 (deg.) with the old parameters for 1N4A and 1NQH, respectively.

The validation results show that the optimized cobalamin parameters are suitable for simulations of cobalamin in Cbl(III) and square-pyramidal Cbl(II) states with both base-on and base-off conformations (excluding the states with an extended Co-histidine bond). Comparison to earlier force field parameters for cobalamin reveals that the new parameters lead to smaller RMSD, less frequent fluctuations in interactions with the protein, and smaller deviations of the fold angle compared to the crystal structures. Although it was not possible to validate the Cbl(I) parameters with MD simulations of full cobalamin in a crystal structure, the small Cbl(I) model showed the best agreement with the QM geometry (Figure 3 and Table 6), suggesting that these parameters are also well-suited for MD simulations.

## Discussion

We have used DFT calculations and ffTK[37] to develop force field parameters for several forms of cobalamin: square-planar Cbl(I), trigonal-bipyramidal Cbl(II) and octahedral Cbl(III), as well as with various axial ligands. The parameters for the amide side chains were taken from the CGenFF force field,[26] whereas the parameters for the Co-corrin core and the corrin side chains were fitted explicitly using smaller models of cobalamin (Scheme 2A). In all cases, parameters were optimized from QM data and simulations of cobalamin alone; simulations of cobalamin-binding proteins were only used for subsequent validation.

The He probe approach[67,68] was used to determine the LJ parameters for the Cb(I) and Cbl(II) states. Because noble gases lack dipole moments and are neutral, they simplify the estimation of vdW interactions from QM calculations. Our calculations of the He ⋯ Co interactions suggest that the strength of these interactions may be underestimated without the D3 correction. By fitting the $\varepsilon$ and $r_{min}$ parameters (Table 1), we obtained good agreement between the QM and MM interaction energies for our models.

Cobalamin contains several buried atoms, which makes it difficult to determine the partial atomic charges for these atoms using the standard CHARMM method of computing water interaction energies[26] or the RESP method.[75] Therefore, we used a combination of the NPA[77] and RESP methods to determine the atomic partial charges. Although NPA charges tend to be more polarized than the corresponding charges from RESP,[91] some over-polarization is desirable to correct for the charges being calculated in gas phase. Usually, over-polarization of partial atomic charges is achieved by computing partial charges at the HF/6-31G(d) level of theory.[23,24,26] However, inclusion of HF exchange is highly detrimental for describing the electronic structure of corrinoids.[94]

The optimized parameters were validated by comparing the MM and QM geometries for the smaller cobalamin models, by examining the corrin fold angles in our simulations, and by evaluating the stability of $B_{12}$ in the binding site during MD simulation trajectories of several corrinoid proteins. Excellent agreement was found for the geometries (Figures 3 and

S7) and the corrin fold angle was reproduced for most cases (Tables 7 and 8) after the adjustment of several improper force constants. Finally, cobalamin remained stably bound in the cofactor binding site in all MD simulations of complexes with protein or RNA (Figure 4). Our parameters reduced the deviations of the cobalamin fold angle, RMSD of cobalamin from its binding-site crystal structure and fluctuations of interactions between cobalamin and the protein when compared to earlier CHARMM-compatible parameters.[49] Thus, the force field parameters are applicable for MD simulations of these and related systems with the CHARMM force field.

Accurate description of interactions with water molecules is an important aspect of force-field parameters. In the case of cobalamin, the metal center was not accessible for water interactions in any of our protein or RNA systems. Our aqueous simulations showed only limited interaction with water, where hydrogen bonds were formed between the corrin-ring nitrogens and water molecules (see Figure S9).

The parametrization approaches described here and in our previous work on macrolide antibiotics[52] can be used as a guide to develop force field parameters for other corrinoids or large biomolecules. In particular, our approaches for determining LJ parameters and partial atomic charges for the metal atom in different oxidation states are applicable to other metal cofactors. Similar challenges for both RESP and water interaction approaches for the buried atoms are expected for other octahedral metal complexes. Here, we show that the use of NPA charges provides a good description of electrostatics for cases in which standard methods are inadequate.

Topology and parameter files for cobalamin, as well as instructions for system preparation, are provided in the Supporting Information. The parameters developed here can be used to study cobalamin-protein interactions in $B_{12}$-dependent enzymes and $B_{12}$-binding transport proteins for which structures are available. The parameters can also be used to simulate cobalamin redox states for which structures are presently not available, e.g. Cbl(I) and MeCbl(III). However, because the bond between DMB or histidine with the cobalt atom is described by a harmonic spring, it is not possible to simulate ligand dissociation with our force field. Certain forms of corrinoids, such as cobalamin with long Co–N(His) coordination distances, base-off Cbl(II) and Cys-on corrinoids were not parametrized here. However, our molecular topologies allow straightforward modification of the Co oxidation state and the axial ligands, facilitating the investigation of non-standard corrinoids and Co-ligand binding configurations with MD simulations. Additionally, patches for norpseudo-$B_{12}$ and Factor $III_m$ forms of cobalamin are provided.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by awards from NSF (MCB-1452464) and NIH (R01-GM123169) to J.C.G. It was also supported by the U.S. Department of Energy (DOE) Office of Science, Biological and Environmental Research, Subsurface Biogeochemical Research (SBR) Program through the Mercury Science Focus Area Program (SFA) at Oak Ridge National Laboratory (ORNL), which is managed by UT-Battelle LLC for the U.S. DOE under contract number DE-AC05-00OR22725. This research used resources at the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. DOE under Contract No. DE-AC02-05CH11231, and the Compute and Data Environment for Science (CADES) at ORNL. Additional computational resources were provided via the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant number OCI-1053575. We thank Sarah J. Cooper for assistance with figures.

## References

1. Banerjee R, Ragsdale SW. The Many Faces of Vitamin $B_{12}$: Catalysis by Cobalamin-Dependent Enzymes. Ann Rev Biochem. 2003; 72:209–247. [PubMed: 14527323]
2. Brown KL. Chemistry and enzymology of vitamin B12. Chem Rev. 2005; 105:2075–2149. [PubMed: 15941210]
3. Kraütler B, Moll J, Thauer RK. The corrinoid from *Methanobacterium thermoautotrophicum* (Marburg strain) Spectroscopic structure-analysis and identification as Co-$\beta$-cyano-5′-hydroxybenzimidazolyl-cobamide (Factor III). Eur J Biochem. 1987; 162:275–278. [PubMed: 3803385]
4. Men Y, Seth EC, Yi S, Crofts TS, Allen RH, Taga ME, Alvarez-Cohen L. Identification of specific corrinoids reveals corrinoid modification in dechlorinating microbial communities. Environ Microbiol. 2015; 17:4873–4884. [PubMed: 24803319]
5. Gruber K, Puffer B, Kräutler B. Vitamin B12-derivatives - Enzyme cofactors and ligands of proteins and nucleic acids. Chem Soc Rev. 2011; 40:4346–4363. [PubMed: 21687905]
6. Randaccio L, Geremia S, Nardin G, Wuerges J. X-ray structural chemistry of cobalamins. Coord Chem Rev. 2006; 250:1332–1350.
7. Marques HM, Brown KL. Molecular mechanics and molecular dynamics simulations of porphyrins, metalloporphyrins, heme proteins and cobalt corrinoids. Coord Chem Rev. 2002; 225:123–158.
8. Ragsdale S, Lindahl P, Münck E. Mössbauer, EPR, and optical studies of the corrinoid/iron-sulfur protein involved in the synthesis of acetyl coenzyme A by *Clostridium thermoaceticum*. J Biol Chem. 1987; 262:14289–14297. [PubMed: 2821001]
9. Svetlitchnaia T, Svetlitchnyi V, Meyer O, Dobbek H. Structural insights into methyltransfer reactions of a corrinoid iron-sulfur protein involved in acetyl-CoA synthesis. Proc Natl Acad Sci USA. 2006; 103:14331–14336. [PubMed: 16983091]
10. Jost M, Fernández-Zapata J, Polanco MC, Ortiz-Guerrero JM, Chen PY-T, Kang G, Padmanabhan S, Elías-Arnanz M, Drennan CL. Structural basis for gene regulation by a $B_{12}$-dependent photoreceptor. Nature. 2015; 526:536–541. [PubMed: 26416754]
11. Bridwell-Rabb J, Drennan CL. Vitamin $B_{12}$ in the spotlight again. Curr Opin Chem Biol. 2017; 37:63–70. [PubMed: 28167430]
12. Parks JM, Johs A, Podar M, Bridou R, Hurt RA, Smith SD, Tomanicek SJ, Qian Y, Brown SD, Brandt CC, Palumbo AV, Smith JC, Wall JD, Elias DA, Liang L. The genetic basis for bacterial mercury methylation. Science. 2013; 339:1332–1335. [PubMed: 23393089]
13. Moore TC, Newmister SA, Rayment I, Escalante-Semerena JC. Structural Insights into the Mechanism of Four-Coordinate Cob(II)alamin Formation in the Active Site of the *Salmonella enterica* ATP: Co(I)rrinoid Adenosyltransferase Enzyme: Critical Role of Residues Phe91 and Trp93. Biochemistry. 2012; 51:9647–9657. [PubMed: 23148601]
14. St Maurice M, Mera P, Park K, Brunold TC, Escalante-Semerena JC, Rayment I. Structural Characterization of a Human-Type Corrinoid Adenosyltransferase Confirms That Coenzyme B12 Is Synthesized through a Four-Coordinate Intermediate. Biochemistry. 2008; 47:5755–5766. [PubMed: 18452306]
15. Randaccio L, Geremia S, Demitri N, Wuerges J. Vitamin $B_{12}$: Unique Metalorganic Compounds and the Most Complex Vitamins. Molecules. 2010; 15:3228–3259. [PubMed: 20657474]
16. Mancia F, Keep NH, Nakagawa A, Leadlay PF, McSweeney S, Rasmussen B, Secke PB, Diat O, Evans PR. How coenzyme $B_{12}$ radicals are generated: The crystal structure of methylmalonyl-coenzyme A mutase at 2 Å resolution. Structure. 1996; 4:339–350. [PubMed: 8805541]

17. Karpowich NK, Huang HH, Smith PC, Hunt JF. Crystal Structures of the BtuF Periplasmic-binding Protein for Vitamin $B_{12}$ Suggest a Functionally Important Reduction in Protein Mobility upon Ligand Binding. J Biol Chem. 2003; 278:8429–8434. [PubMed: 12468528]

18. Kung Y, Ando N, Doukov TI, Blasiak LC, Bender G, Seravalli J, Ragsdale SW, Drennan CL. Visualizing molecular juggling within a $B_{12}$-dependent methyltransferase complex. Nature. 2012; 484:265–269. [PubMed: 22419154]

19. Johnson JE Jr, Reyes FE, Polaski JT, Batey RT. $B_{12}$ cofactors directly stabilize an mRNA regulatory switch. Nature. 2012; 492:133–137. [PubMed: 23064232]

20. Hou J, Townson SA, Kovalchin JT, Masci A, Kiner O, Shu Y, King BM, Schirmer E, Golden K, Thomas C, Garcia KC, Zarbis-Papastoitsis G, Furfine ES, Barnes TM. Design of a superior cytokine antagonist for topical ophthalmic use. Proc Natl Acad Sci USA. 2013; 110:3913–3918. [PubMed: 23431173]

21. Payne KA, Quezada CP, Fisher K, Dunstan MS, Collins FA, Sjuts H, Levy C, Hay S, Rigby SE, Leys D. Reductive dehalogenase structure suggests a mechanism for $B_{12}$-dependent dehalogenation. Nature. 2015; 517:513–516. [PubMed: 25327251]

22. Bommer M, Kunze C, Fesseler J, Schubert T, Diekert G, Dobbek H. Structural basis for organohalide respiration. Science. 2014; 346:455–458. [PubMed: 25278505]

23. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J Am Chem Soc. 1995; 117:5179–5197.

24. MacKerell AD Jr, Wiorkiewicz-Kuczera J, Karplus M. An all-atom empirical energy function for the simulation of nucleic acids. J Am Chem Soc. 1995; 117:11946–11975.

25. Lopes PE, Guvench O, MacKerell AD Jr. Current Status of Protein Force Fields for Molecular Dynamics. Mol Model Prot. 2015; 1215:47–71.

26. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem. 2010; 31:671–690. [PubMed: 19575467]

27. Mukherjee G, Patra N, Barua P, Jayaram B. A fast empirical GAFF compatible partial atomic charge assignment scheme for modeling interactions of small molecules with biomolecular targets. J Comput Chem. 2011; 32:893907.

28. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J Am Chem Soc. 1996; 118:11225–11236.

29. Autenrieth F, Tajkhorshid E, Baudry J, Luthey-Schulten Z. Classical force field parameters for the heme prosthetic group of cytochrome C. J Comput Chem. 2004; 25:1613–1622. [PubMed: 15264255]

30. Oda A, Yamaotsu N, Hirono S. New AMBER force field parameters of heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. J Comput Chem. 2005; 26:818–826. [PubMed: 15812779]

31. Adam S, Knapp-Mohammady M, Yi J, Bondar A-N. Revised CHARMM force field parameters for iron-containing cofactors of photosystem II. J Comp Chem. 2017 in press.

32. Ceccarelli M, Procacci P, Marchi M. An ab initio force field for the cofactors of bacterial photosynthesis. J Comput Chem. 2003; 24:129–142. [PubMed: 12497594]

33. Šebesta F, Sláma V, Melcr J, Futera Z, Burda JV. Estimation of Transition-Metal Empirical Parameters for Molecular Mechanical Force Fields. J Chem Theory Comput. 2016; 12:3681–3688. [PubMed: 27337427]

34. Peters MB, Yang Y, Wang B, Füsti-Molnár L, Weaver MN, Merz KM. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). J Chem Theory Comput. 2010; 6:2935–2947. [PubMed: 20856692]

35. Burger SK, Lacasse M, Verstraelen T, Drewry J, Gunning P, Ayers PW. Automated Parametrization of AMBER Force Field Terms from Vibrational Analysis with a Focus on Functionalizing Dinuclear Zinc(II) Scaffolds. J Chem Theory Comput. 2012; 8:554–562. [PubMed: 26596604]

36. Zheng S, Tang Q, He J, Du S, Xu S, Wang C, Xu Y, Lin F. VFFDT: A New Software for Preparing AMBER Force Field Parameters for Metal-Containing Molecular Systems. J Chem Inf Model. 2016; 56:811–818. [PubMed: 26998926]
37. Mayne CG, Saam J, Schulten K, Tajkhorshid E, Gumbart JC. Rapid parameterization of small molecules using the Force Field Toolkit. J Comput Chem. 2013; 34:2757–2770. [PubMed: 24000174]
38. Hoops SC, Anderson KW, Merz KM. Force field design for metalloproteins. J Am Chem Soc. 1991; 113:8262–8270.
39. Stote RH, Karplus M. Zinc binding in proteins and solution: A simple but accurate nonbonded representation. Prot Struct, Funct Bioinf. 1995; 23:12–31.
40. Pang YP, Xu K, Yazal JE, Prendergas FG. Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. Prot Sci. 2000; 9:1857–1865.
41. Hu L, Ryde U. Comparison of Methods to Obtain Force-Field Parameters for Metal Sites. J Chem Theory Comput. 2011; 7:2452–2463. [PubMed: 26606619]
42. Pang Y-P. Novel Zinc Protein Molecular Dynamics Simulations: Steps Toward Antiangiogenesis for Cancer Treatment. J Mol Model. 1999; 5:196–202.
43. Singh N, Warshel A. Absolute binding free energy calculations: On the accuracy of computational scoring of protein-ligand interactions. Prot Struct Funct Bioinf. 2010; 78:1705–1723.
44. Comba P, Kerscher M. Computation of structures and properties of transition metal compounds. Coord Chem Rev. 2009; 253:564–574.
45. Vanommeslaeghe K, Yang M, MacKerell AD Jr. Robustness in the fitting of molecular mechanics parameters. J Comp Chem. 2015; 36:1083–1101. [PubMed: 25826578]
46. Allinger NL. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. J Am Chem Soc. 1977; 99:8127–8134.
47. Marques HM, Brown KL. A molecular mechanics force field for the cobalt corrinoids. J Mol Struct THEOCHEM. 1995; 340:97–124.
48. Marques HM, Ngoma B, Egan TJ, Brown KL. Parameters for the AMBER force field for the molecular mechanics modeling of the cobalt corrinoids. J Mol Struct. 2001; 561:71–91.
49. Gumbart J, Wiener MC, Tajkhorshid E. Coupling of Calcium and Substrate Binding through Loop Alignment in the Outer-Membrane Transporter BtuB. J Mol Biol. 2009; 393:1129–1142. [PubMed: 19747487]
50. Marques HM, Hicks RP, Brown KL. Solution structure of cyanocobalamin (vitamin $B_{12}$) by NMR-restrained molecular dynamics and simulated annealing calculations. Chem Comm. 1996:1427–1428.
51. Perry CB, Brown KL, Zou X, Marques HM. The solution structure of some cobalamins determined by NMR-restrained molecular modelling. J Mol Struct. 2005; 737:245–258.
52. Pavlova A, Gumbart JC. Parametrization of macrolide antibiotics using the Force Field Toolkit. J Comp Chem. 2015; 36:2052–2063. [PubMed: 26280362]
53. Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, MacKerell AD Jr. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. J Chem Theory Comput. 2012; 8:348–362. [PubMed: 22368531]
54. Neese F, Liakos DG, Ye S. Correlated wavefunction methods in bioinorganic chemistry. J Biol Inorg Chem. 2011; 16:821–829. [PubMed: 21541855]
55. Becke AD. Density-functional exchange-energy approximation with correct asymptotic behavior. Phys Rev A. 1988; 38:3098–3100.
56. Perdew JP. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. Phys Rev B. 1986; 33:8822–8824.
57. Perdew JP, Zunger A. Self-interaction correction to density-functional approximations for many-electron systems. Phys Rev B. 1981; 23:5048–5079.
58. Jensen KP, Ryde U. Theoretical prediction of the Co–C bond strength in cobalamins. J Phys Chem A. 2003; 107:7539–7545.
59. Rovira C, Biarnés X, Kunc K. Structure-energy relations in methylcobalamin with and without bound axial base. Inorg Chem. 2004; 43:6628–6632. [PubMed: 15476360]

60. Kuta J, Wuerges J, Randaccio L, Kozlowski PM. Axial bonding in alkylcobalamins: DFT analysis of the inverse versus normal trans influence. J Phys Chem A. 2009; 113:11604–11612. [PubMed: 19848426]

61. Kozlowski PM, Kumar M, Piecuch P, Li W, Bauman NP, Hansen JA, Lodowski P, Jaworska M. The cobalt–methyl bond dissociation in methylcobalamin: New benchmark analysis based on density functional theory and completely renormalized coupled-cluster calculations. J Chem Theory Comput. 2012; 8:1870–1894. [PubMed: 26593822]

62. Kepp KP. Co–C Dissociation of Adenosylcobalamin (Coenzyme B12): Role of Dispersion, Induction Effects, Solvent Polarity, and Relativistic and Thermal Corrections. J Phys Chem A. 2014; 118:7104–7117. [PubMed: 25116644]

63. Weigend F, Ahlrichs R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. Phys Chem Chem Phys. 2005; 7:3297–3305. [PubMed: 16240044]

64. Weigend F. Accurate Coulomb-fitting basis sets for H to Rn. Phys Chem Chem Phys. 2006; 8:1057–1065. [PubMed: 16633586]

65. Johnston RC, Zhou J, Smith JC, Parks JM. Toward quantitatively accurate calculation of the redox-associated acid-base and ligand binding equilibria of aquacobalamin. J Phys Chem B. 2016; 120:7307–7318. [PubMed: 27391132]

66. Frisch, MJ., Trucks, GW., Schlegel, HB., Scuseria, GE., Robb, MA., Cheeseman, JR., Scalmani, G., Barone, V., Petersson, GA., Nakatsuji, H., Li, X., Caricato, M., Marenich, A., Bloino, J., Janesko, BG., Gomperts, R., Mennucci, B., Hratchian, HP., Ortiz, JV., Izmaylov, AF., Sonnenberg, JL., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, VG., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, JA., Jr, Peralta, JE., Ogliaro, F., Bearpark, M., Heyd, JJ., Brothers, E., Kudin, KN., Staroverov, VN., Keith, T., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, JC., Iyengar, SS., Tomasi, J., Cossi, M., Millam, JM., Klene, M., Adamo, C., Cammi, R., Ochterski, JW., Martin, RL., Morokuma, K., Farkas, O., Foresman, JB., Fox, DJ. Gaussian 09, Revision D01. Gaussian, Inc.; Wallingford CT: 2016.

67. Yin D, MacKerell AD Jr. Combined ab initio/empirical approach for optimization of Lennard-Jones parameters. J Comput Chem. 1998; 19:334–348.

68. Chen IJ, Yin D, MacKerell AD Jr. Combined ab initio/empirical approach for optimization of Lennard-Jones parameters for polar-neutral compounds. J Comput Chem. 2002; 23:199–213. [PubMed: 11924734]

69. Perdew JP, Burke K, Ernzerhof M. Generalized gradient approximation made simple. Phys Rev Lett. 1996; 77:3865. [PubMed: 10062328]

70. Becke AD. Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys. 1993; 98:5648–5652.

71. Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B. 1988; 37:785.

72. Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys. 2010; 132:154104. [PubMed: 20423165]

73. Grimme S, Ehrlich S, Goerigk L. Effect of the damping function in dispersion-corrected density functional theory. J Comput Chem. 2011; 32:1456–1465. [PubMed: 21370243]

74. Humphrey W, Dalke A, Schulten K. VMD - visual molecular dynamics. J Molec Graphics. 1996; 14:33–38.

75. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem. 1993; 97:10269–10280.

76. Bachrach, SM. Reviews in Computational Chemistry. John Wiley & Sons, Inc; 2007. p. 171-228.

77. Reed AE, Curtiss LA, Weinhold F. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. Chem Rev. 1988; 88:899–926.

78. Woods RJ, Khalil M, Pell W, Moffat SH, Smith VH. Derivation of net atomic charges from molecular electrostatic potentials. J Comput Chem. 1990; 11:297–310.

79. Stouch T, Williams DE. Conformational dependence of electrostatic potential derived charges of a lipid headgroup: Glycerylphosphorylcholine. J Comput Chem. 1992; 13:622–632.

80. Hu H, Lu Z, Yang W. Fitting molecular electrostatic potentials from quantum mechanical calculations. J Chem Theory Comput. 2007; 3:1004–1013. [PubMed: 26627419]

81. Dupradeau F-Y, Pigache A, Zaffran T, Savineau C, Lelong R, Grivel N, Le-long D, Rosanski W, Cieplak P. The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building. Phys Chem Chem Phys. 2010; 12:7821–7839. [PubMed: 20574571]

82. Case, DA., Babin, V., Berryman, JT., Betz, R., Cai, Q., Cerutti, D., Cheatham, T., Darden, T., Duke, R., Gohlke, H., Goetz, A., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossvry, I., Kovalenko, A., Lee, T., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, K., Paesani, F., Roe, D., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Seabra, G., Simmerling, C., Smith, W., Swails, J., Walker, R., Wang, J., Wolf, R., Wu, X., Kollman, P. AMBER 14. University of California; San Francisco: 2014.

83. Guvench O, MacKerell AD Jr. Automated conformational energy fitting for force-field development. J Mol Model. 2008; 14:667–679. [PubMed: 18458967]

84. Phillips JC, Braun R, Wang W, Gumbart JC, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]

85. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone $\phi$, $\psi$ and Side-Chain $\chi_1$ and $\chi_2$ Dihedral Angles. J Chem Theory Comput. 2012; 8:3257–3273. [PubMed: 23341755]

86. Klauda JB, Venable RM, Freites JA, OConnor JW, Tobias DJ, Mondragon-Ramirez C, Vorobyov I, MacKerell AD, Pastor RW. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. J Phys Chem B. 2010; 114:7830–7843. [PubMed: 20496934]

87. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; 79:926–935.

88. Darden T, York D, Pedersen L. Particle mesh Ewald: An N × log(N) method for Ewald sums in large systems. J Chem Phys. 1993; 98:10089–10092.

89. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comp Chem. 2008; 29:1859–1865. [PubMed: 18351591]

90. Reed AE, Weinstock RB, Weinhold F. Natural population analysis. J Chem Phys. 1985; 83:735–746.

91. Xu Z, Luo HH, Tieleman DP. Modifying the OPLS-AA force field to improve hydration free energies for several amino acid side chains using new atomic charges and an off-plane charge model for aromatic residues. J Comput Chem. 2007; 28:689–697. [PubMed: 17195160]

92. Masuda J, Shibata N, Morimoto Y, Toraya T, Yasuoka N. How a protein generates a catalytic radical from coenzyme B12: X-ray structure of a dioldehydrataseadeninylpentylcobalamin complex. Structure. 2000; 8:775–788. [PubMed: 10903944]

93. Chimento DP, Mohanty AK, Kadner RJ, Wiener MC. Substrate-induced transmembrane signaling in the cobalamin transporter BtuB. Nat Struct Biol. 2003; 10:394–401. [PubMed: 12652322]

94. Jensen KP, Ryde U. Theoretical Prediction of the Co-C Bond Strength in Cobalamins. J Phys Chem A. 2003; 107:7539–7545.

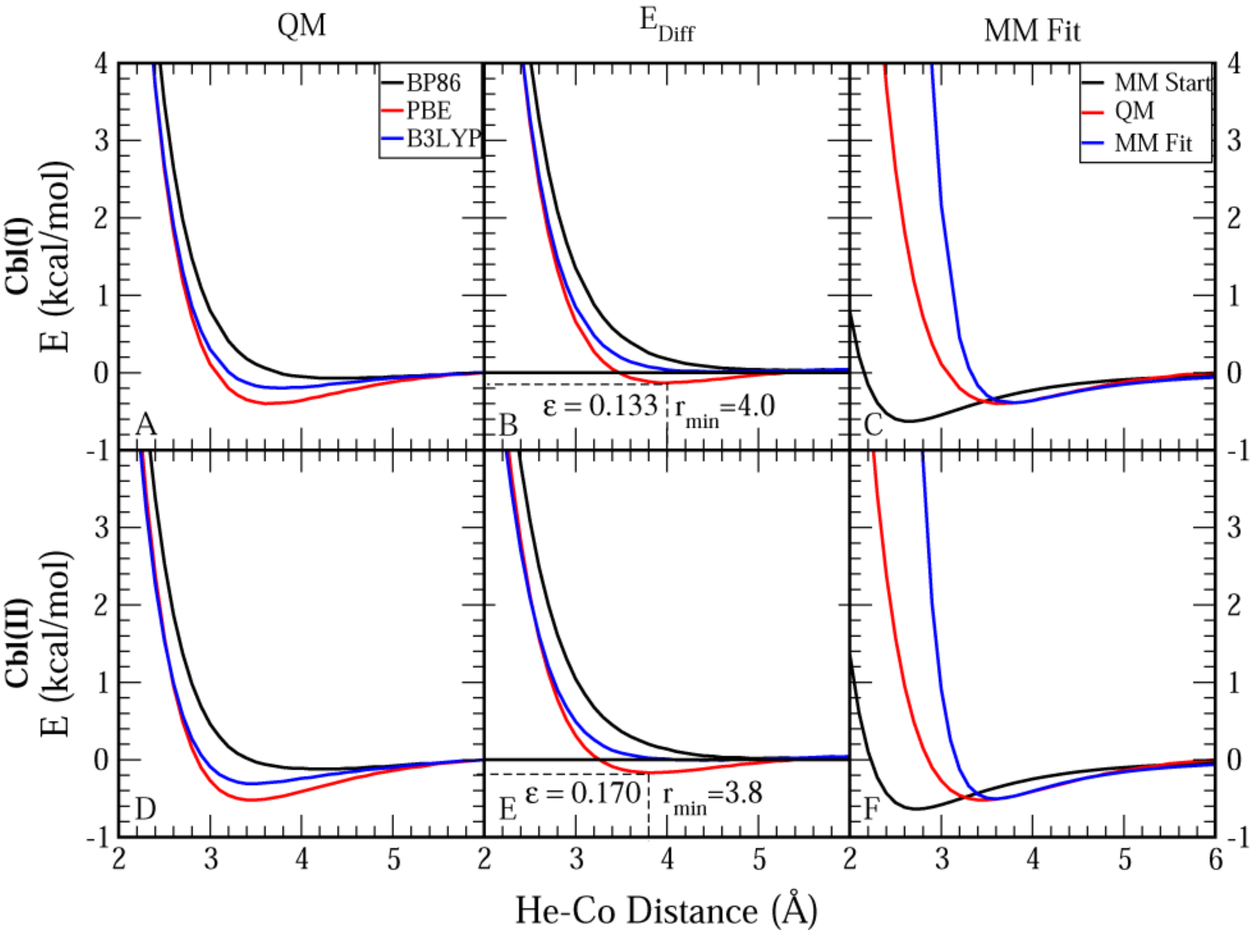


**Figure 1.**

Fitting of LJ parameters for Cbl(II) and Cbl(I) from QM calculations. D3 correction and Def2-TZVPP basis set were used for all QM calculations. See Figures S1 in Supporting Information for QM data at different basis sets. (A) Non-bonded interaction energies between He and Co(I)-cobalamin as a function of the He ⋯ Co distance for three different density functionals. (B) Difference between QM energies from (A) and the corresponding MM energies when the Co LJ potential is set to zero ($En_{Diff}$) The colors correspond to the same QM level of theory as in (A). The E = 0 line is shown to illustrate that only the red line (PBE) has a minimum below zero for both Cbl(II) and Cbl(I). (C) Comparison of non-bonded interaction energies between Cbl(I)-cobalamin and helium at BPE level of theory and using MM with (MM Fit) and without (MM Start) LJ potential for the Co atom. (D) Non-bonded interaction energies between He and HO-cob(II)alamin as a function of the He ⋯ Co distance at different levels of theory. (E) Difference between the QM energies from (D) and the corresponding MM energies with the LJ potential for Co set equal to zero. The colors correspond to the same level of theory as in (D). (F) Comparison of non-bonded interaction energies between He and HO-cob(II)alamin calculated at PBE level of theory (QM) and using MM with (MM Fit) and without (MM Start) the LJ potential for Co.

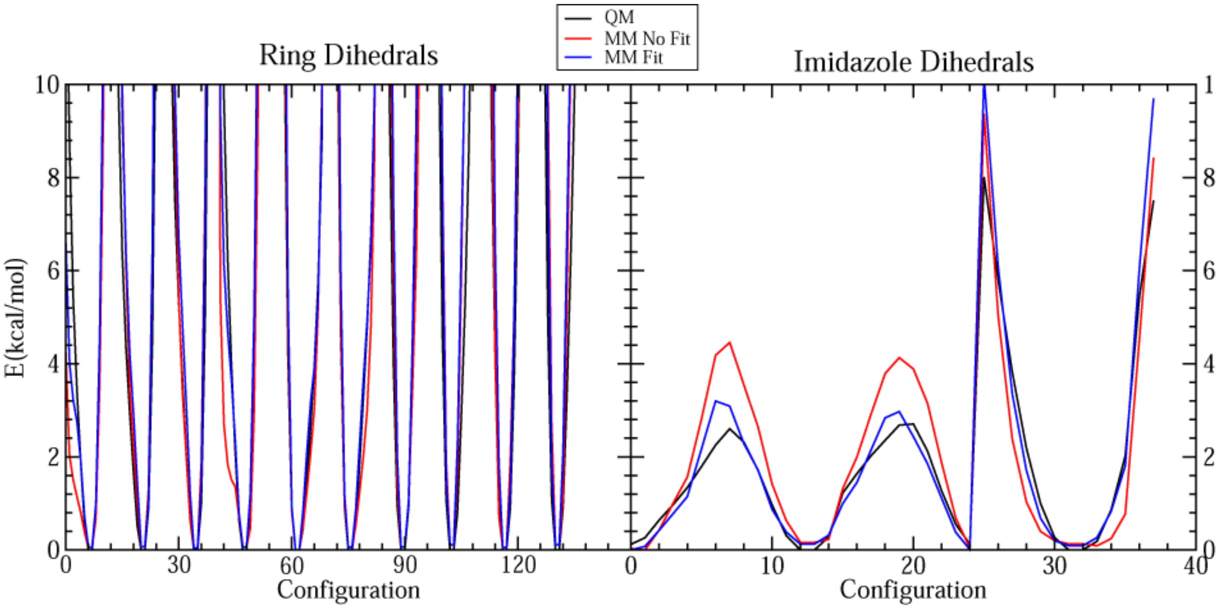


**Figure 2.**

Comparison of the QM PES to the MM PES before (No Fit) and after (Fit) fitting the dihedral constants for ring and imidazole dihedrals.

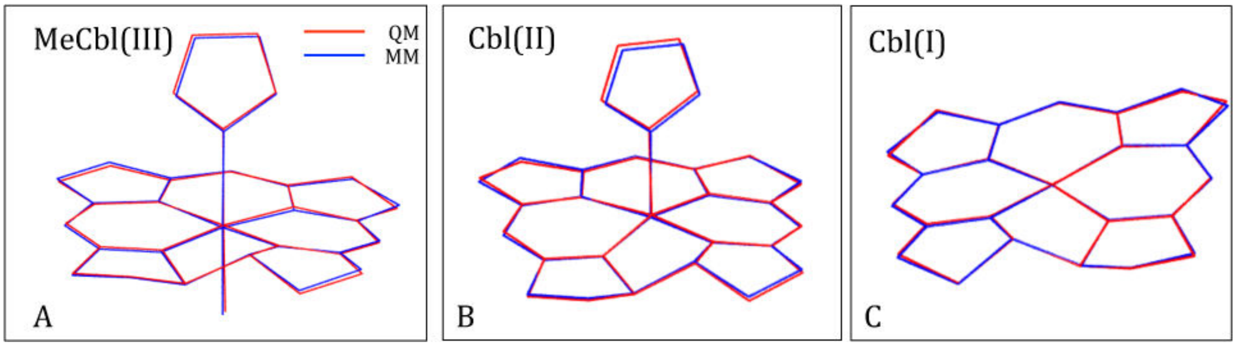


**Figure 3.**

Comparison of QM (red) and MM (blue) geometries for the small (A) MeCbl(III), (B) Cbl(II) and (C) Cbl(I) models. Hydrogens are omitted for clarity. See Figure S5 for a comparison of the CN-Cbl(III) and AdoCbl(III) geometries, and Table 6 for the RMSD between the MM and QM geometries.

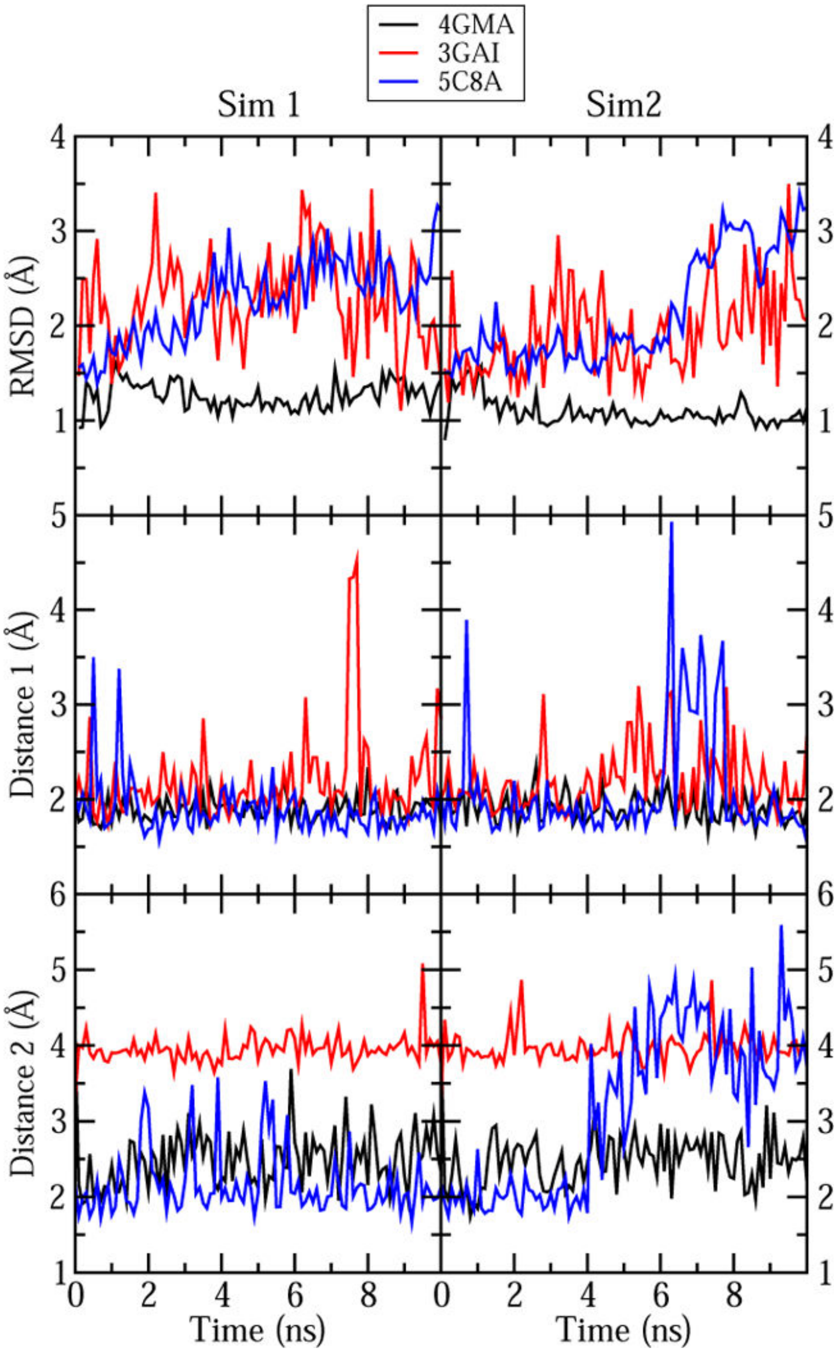


**Figure 4.**

Measurements of the cobalamin RMSD and the two selected distances in our simulations of cobalamin-containing proteins (see Table 9). The results of the first and the second simulation are shown in the left and right graphs, respectively. For 5C8A we only show the distances for chain A; see Figure S8 for the corresponding distances in chain B. Time 0 corresponds to the values in the X-ray crystal structures.

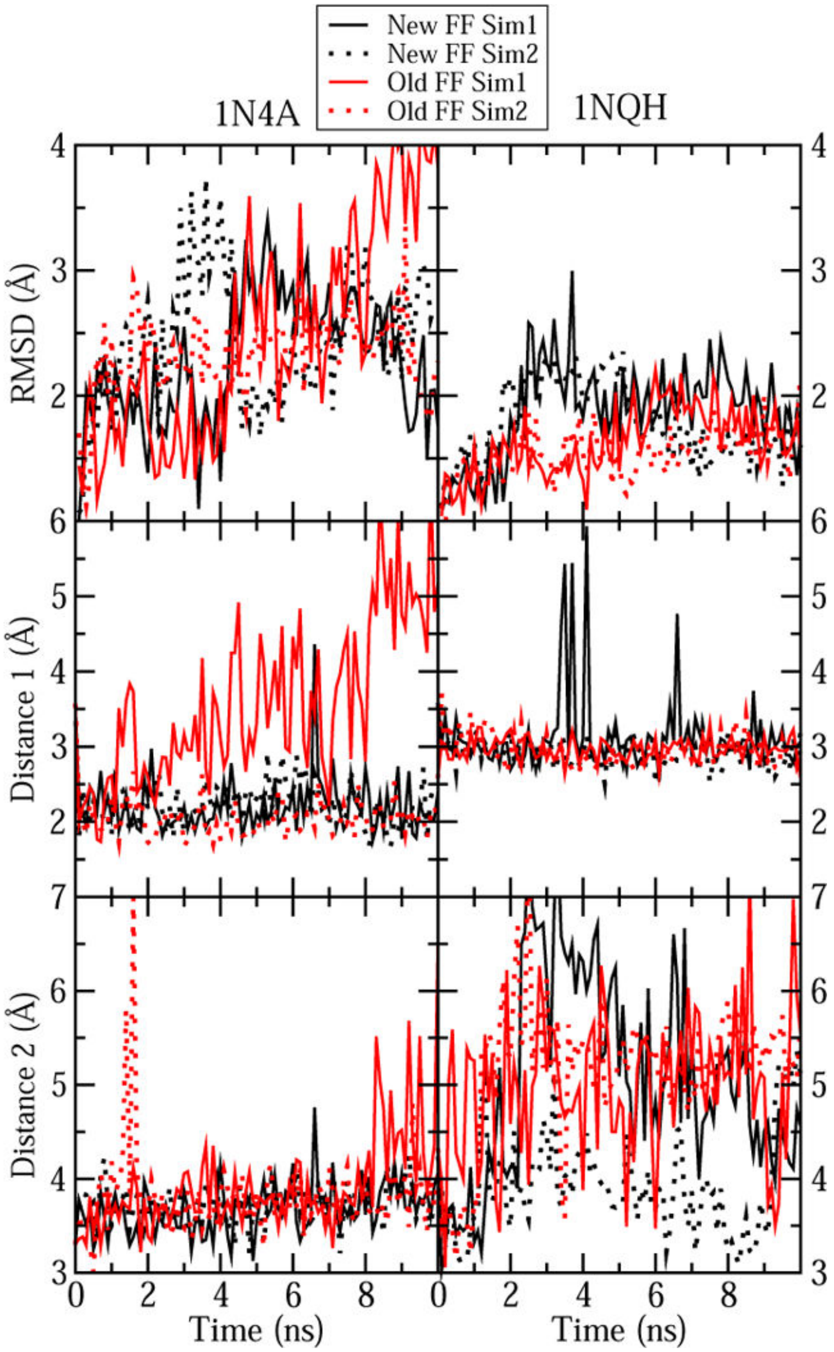
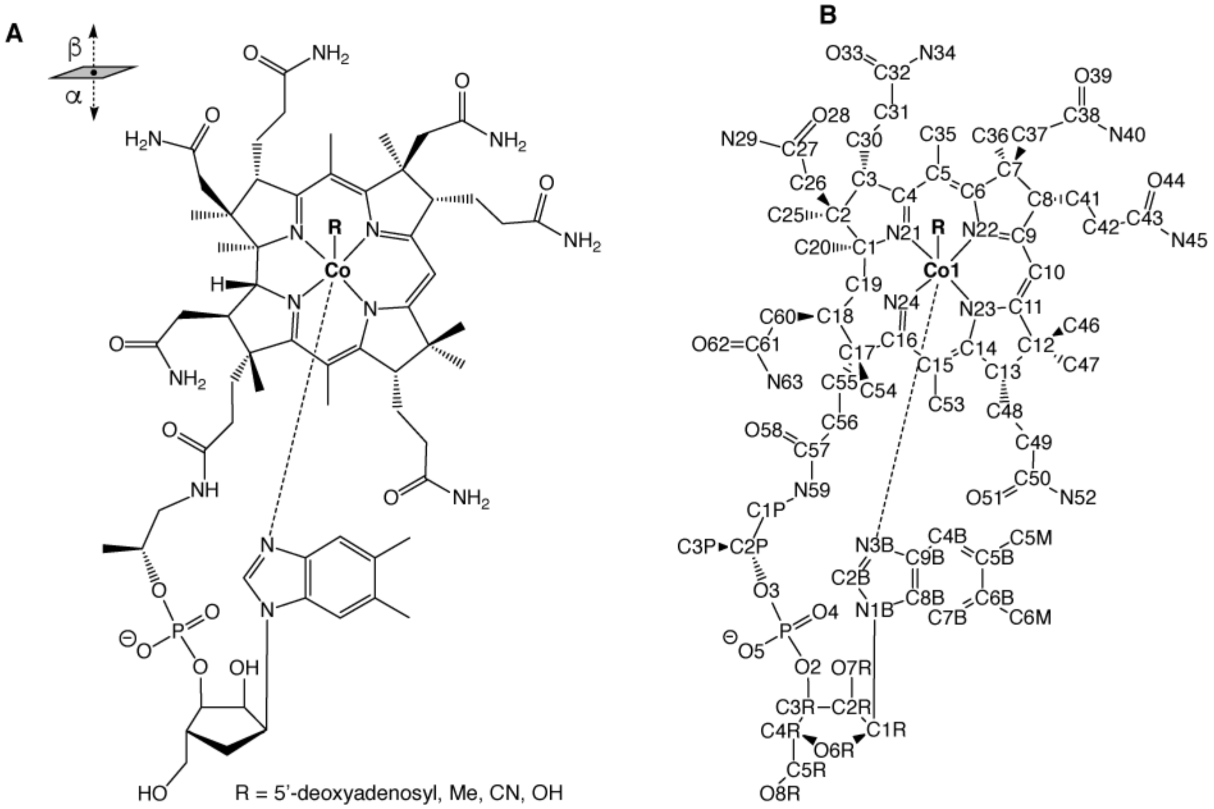


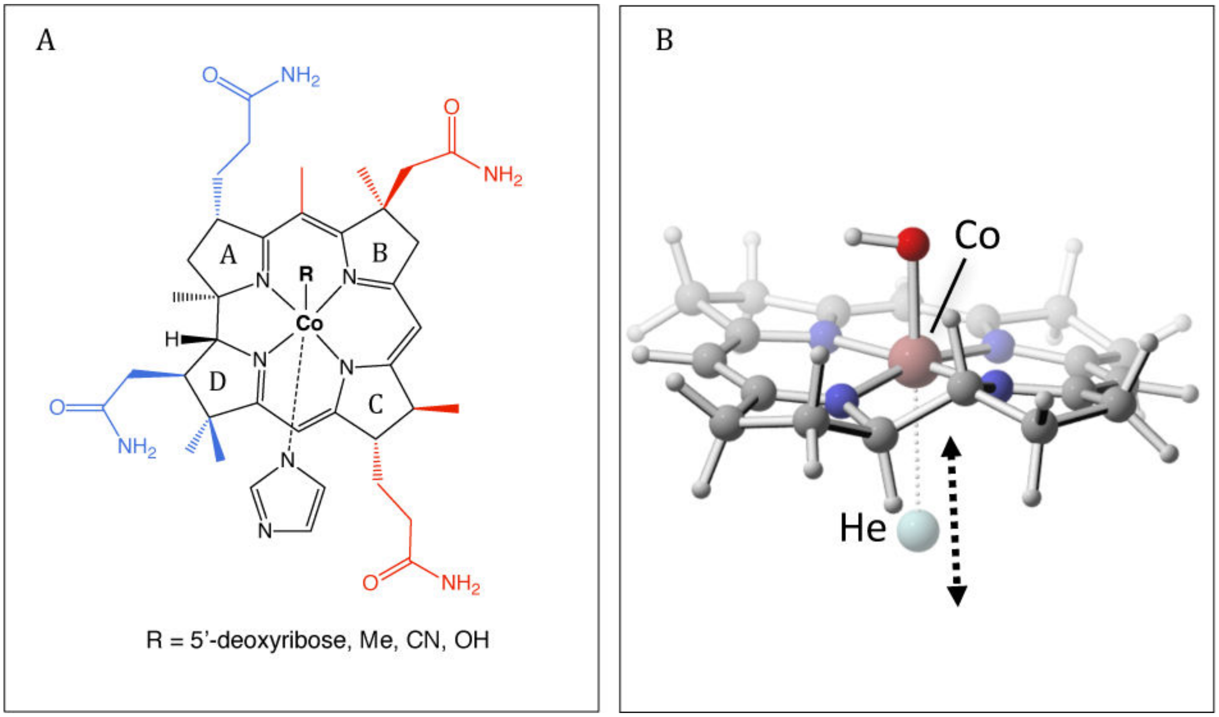
**Figure 5.**

Comparison between our new force field parameters for cobalamin and the previous parameters for simulation of crystal structure 1N4A and 1NQH. Cobalamin RMSD and the two selected distances from Table 9 were monitored for both cases.



**Scheme 1.**

(**A**) Chemical structure of cobalamin. (**B**) Standard PDB atom numbering for cobalamin.



**Scheme 2.**

(**A**) Models used to develop parameters for cobalamin. The smaller models of cobalamin used for fitting most of the force-field parameters are shown in black. Side chains that were added to this model to create the structures ExtAD and ExtBC are shown in blue and red, respectively. (**B**) Displacement vector of the He probe used to fit the LJ parameters for Co(I) and Co(II).

## Table 1

Comparison of the minimum interaction energies and distances for our helium-cobalamin models at the QM and MM levels before and after the fit. The resulting $\varepsilon$ and $r_{min}$ parameters for Co are also shown. The LJ parameters are provided in the standard CHARMM parameter file format as $-\varepsilon_{Co}$ and $r_{min,\ Co}/2$.

| State | Minimum (kcal mol$^{-1}$) | | | Distance (Å) | | | $-\varepsilon_{Co}$ (kcal mol$^{-1}$) | $r_{min;Co}/2$ (Å) |
|---|---|---|---|---|---|---|---|---|
| | QM | $MM_{init}$ | $MM_{Fit}$ | QM | $MM_{init}$ | $MM_{Fit}$ | | |
| Cbl(I) | −0.399 | −0.628 | −0.390 | 3.7 | 2.7 | 3.8 | −0.832 | 2.52 |
| Cbl(II) | −0.523 | −0.637 | −0.506 | 3.5 | 2.7 | 3.6 | −1.325 | 2.32 |

## Table 2

Partial atomic charges for Co, the Co-bound nitrogens N21-N34, and N3B for five cobalamin complexes. Partial charges for all atoms are provided in the Supporting Information.

| | **$B_{12}$ State** | | | |
|---|---|---|---|---|
| **Atom** | **Me- and AdoCbl(III)** | **CN-Cbl(III)** | **Cbl(II)** | **Cbl(I)** |
| System Charge | +1 | +1 | +1 | 0 |
| Co | 0.382 | 0.234 | 0.613 | 0.424 |
| N21,N24 | −0.403 | −0.387 | −0.463 | −0.479 |
| N23,N22 | −0.444 | −0.435 | −0.496 | −0.504 |
| N3B | −0.492 | −0.386 | −0.509 | NA |

## Table 3

Comparison of selected bond lengths for minimized geometries of our MeCbl(III), Cbl(II), and Cbl(I) models. All bond lengths are in Å. NP means that the bond is not present in the model. See the parameter file in the Supporting Information for all optimized bond parameters and 1B for the atom numbering.

| | **B12 State** | | | | | |
|---|---|---|---|---|---|---|
| | **MeCbl(III)** | | **Cbl(II)** | | **Cbl(I)** | |
| **Bond** | **QM** | **MM** | **QM** | **MM** | **QM** | **MM** |
| Co–N21 | 1.88 | 1.87 | 1.88 | 1.88 | 1.84 | 1.84 |
| Co–N22 | 1.94 | 1.98 | 1.94 | 1.93 | 1.91 | 1.92 |
| Co–N23 | 1.94 | 1.97 | 1.94 | 1.92 | 1.91 | 1.92 |
| Co–N24 | 1.88 | 1.86 | 1.88 | 1.88 | 1.84 | 1.84 |
| Co–N3B | 2.17 | 2.16 | 2.13 | 2.12 | NP | NP |

## Table 4

Comparison of selected angles for minimized geometries of our MeCbl(III), Cbl(II), and Cbl(I) models. All angle values are in deg. NP means that the angle is not present in the model. See the parameter file in the Supporting Information for all optimized angle parameters and 1B for the atom numbering.

| | **B12 State** | | | | | |
|---|---|---|---|---|---|---|
| | **MeCbl(III)** | | **Cbl(II)** | | **Cbl(I)** | |
| **Angle** | **QM** | **MM** | **QM** | **MM** | **QM** | **MM** |
| N22–Co–N21 | 91.1 | 90.7 | 91.1 | 90.7 | 91.3 | 91.3 |
| N24–Co–N21 | 82.5 | 84.0 | 82.5 | 82.3 | 83.5 | 83.9 |
| N23–Co–N22 | 95.2 | 93.9 | 94.7 | 95.7 | 94.5 | 94.0 |
| N24–Co–N23 | 91.4 | 91.8 | 90.7 | 91.1 | 91.3 | 91.4 |
| N21–Co–N3B | 88.4 | 87.2 | 92.6 | 89.1 | NP | NP |
| N22–Co–N3B | 90.6 | 91.3 | 95.9 | 95.1 | NP | NP |
| N23–Co–N3B | 92.6 | 87.7 | 92.2 | 92.0 | NP | NP |
| N24–Co–N3B | 88.7 | 93.3 | 98.0 | 96.6 | NP | NP |

## Table 5

RMSE of all dihedral constant optimizations performed in this work.

| **Dihedrals** | **RMSE (kcal mol$^{-1}$)** |
|---|---|
| Imidazole | 0.56 |
| Ring | 2.2 |
| Connection (Con1) | 0.22 |
| Connection (Con2) | 0.82 |
| Ado | 0.40 |
| Cbl(II) refit | 1.8 |
| Cbl(I) refit | 1.2 |

## Table 6

RMSD between QM and MM geometries of the five small cobalamin models.

| **Model** | **RMSD (Å)** |
|---|---|
| MeCbl(III) | 0.070 |
| CN-Cbl(III) | 0.117 |
| AdoCbl(III) | 0.117 |
| Cbl(II) | 0.076 |
| CblI) | 0.047 |

## Table 7

Comparison of fold angles (deg) for different cobalamins in aqueous MD simulations before and after fitting the improper force constants ($MD_{In}$ and $MD_{Fit}$) to the experimental fold angles (Exp).[92]

| System | $MD_{In}$ | $MD_{Fit}$ | Exp |
|---|---|---|---|
| MeCbl(III) | 18.4 | 13.5 | 14.8 |
| CN-Cbl(III) | 18.1 | 12.1 | 14.1 |
| AdoCbl(III) | 15.9 | 10.6 | 10.0 |
| Cbl(II) | 26.5 | 14.0 | 11.5 |
| Cbl(I) | 21.1 | 11.3 | NA |

## Table 8

Validation of the improper force constants by comparing the corrin fold angles (deg) from MD simulations before and after fitting of improper constants ($MD_{In}$ and $MD_{Fit}$) to the corrin fold angles in the crystal structures (CS). No parameters were altered based on the bound-state simulations. NA indicates that the simulation was not performed.

| **System** | **$MD_{In}$** | **$MD_{Fit}$** | **CS** |
|---|---|---|---|
| 1N4A | 15.5 | 10.7 | 3.2 |
| 4GMA | 15.3 | 12.7 | 12.5 |
| 3GAI | 13.0 | 8.6 | 11.2 |
| 5C8A(Chain A) | 17.1 | 7.1 | 2.5 |
| 5C8A(Chain B) | 15.8 | 9.3 | 15.9 |
| 1NQH | NA | 12.3 | 4.7 |

## Table 9

Definitions of the distances monitored during each MD simulation. The residue and atom name for the two atoms involved in each distance are specified. Heavy atoms are specified for the hydrogen bond interactions, but the distances were measured between the acceptor atom and the hydrogen bound to the donor atom with the exception of Distance 2 in 1NQH, for which we used the serine hydroxy oxygen due to a switch between a hydrogen bond donor and an acceptor to atoms proximal to $B_{12}$ C43.

| **Crystal structure** | **Distance 1** | | **Distance 2** | |
|---|---|---|---|---|
| | **Atom 1** | **Atom 2** | **Atom 1** | **Atom2** |
| 1N4A | $B_{12}$, O44 | A10, N | $B_{12}$, C7B | W63, CZ3 |
| 4GMA | $B_{12}$, O39 | G49, N2 | $B_{12}$, N34 | C75, O2’ |
| 3GAI | $B_{12}$, O44 | I113, N | $B_{12}$, Co | ATP, C5’ |
| 5C8A | $B_{12}$, O51 | Q178, N | $B_{12}$, O65 | $B_{12}$, N59 |
| 1NQH | $B_{12}$, P | S61, OG | $B_{12}$, C43 | S91, OG |