JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

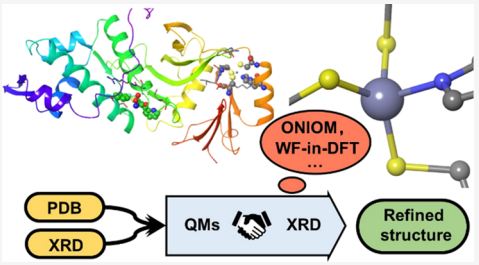
Multiscale Quantum Refinement Approaches for Metalloproteins

Zeyin Yan, Xin Li, and Lung Wa Chung*

Cite This: J. Chem. Theory Comput. 2021, 17, 3783−3796 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Biomolecules with metal ion(s) (e.g., metalloproteins) play many important biological roles. However, accurate structural determination of metalloproteins, particularly those containing transition metal ion(s), is challenging due to their complicated electronic structure, complex bonding of metal ions, and high number of conformations in biomolecules. Quantum refinement, which was proposed to combine crystallographic data with computational chemistry methods by several groups, can improve the local structures of some proteins. In this study, a quantum refinement method combining several multiscale computational schemes with experimental (X-ray diffraction) information was developed for metalloproteins. Various quantum refinement approaches using different ONIOM (our own *N*-layered integrated molecular orbital and molecular mechanics) combinations of quantum mechanics (QM), semiempirical (SE), and molecular mechanics (MM) methods were conducted to assess the performance and reliability on the refined local structure in two metalloproteins. The structures for two (Cu- or Zn-containing) metalloproteins were refined by combining two-layer ONIOM2(QM1/QM2) and ONIOM2(QM/MM) and three-layer ONIOM3(QM1/QM2/MM) schemes with experimental data. The accuracy of the quantum-refined metal binding sites was also examined and compared in these multiscale quantum refinement calculations. ONIOM3(QM/SE/MM) schemes were found to give good results with lower computational costs and were proposed to be a good choice for the multiscale computational scheme for quantum refinement calculations of metal binding site(s) in metalloproteins with high efficiency. Additionally, a two-center ONIOM approach was employed to speed up the quantum refinement calculations for the Zn metalloprotein with two remote active sites/ligands. Moreover, a recent quantum-embedding wavefunction-in-density functional theory (WF-in-DFT) method was also adopted as the high-level method in unprecedented ONIOM2(CCSD-in-B3LYP/MM) and ONIOM3(CCSD-in-B3LYP/SE/MM) calculations, which can be regarded as novel pseudo-three- and pseudo-four-layer ONIOM methods, respectively, to refine the key Zn binding site at the coupled-cluster singles and doubles (CCSD) level. These refined results indicate that multiscale quantum refinement schemes can be used to improve the structural accuracy obtained for local metal binding site(s) in metalloproteins with high efficiency.



1. INTRODUCTION

X-ray diffraction (XRD) has been widely used to determine the three-dimensional structures of various biomolecules (e.g., proteins and nuclei acids). The detailed atomic structures for biomolecules give important atomic information for us to understand their biological functions, properties, enzymatic reactions, and drug binding modes. In standard crystallographic refinement, classical force fields have been often used to aid in giving more reasonable chemical structures from the experimental (XRD) data.[1] Unfortunately, the determination of transition metal ion(s) in biomolecules (particularly with limited experimental resolution) by using force fields is challenging, as force fields cannot be used to well describe chemical bonding with transition metals.[2,3] Quantum mechanics (QM) methods can be helpful for reasonably describing chemical bonding with transition metals. However, the complete determination of protein structures by QM methods requires huge computational costs. In this regard, a combination of QM and molecular mechanics (MM) is one of the most popular and efficient computational chemistry methods for the study of biomolecules.[4−13] A small but important part of the biomolecule (active site, such as a metal binding site) is treated by a more accurate QM method in this hybrid method, while the rest of the biomolecule is described by a fast MM method.

Moreover, to realistically optimize the whole structure of biomolecules in an efficient manner, the combination of a computational chemistry method with experimental information was first proposed by Levitt and co-workers.[14] Subsequently, the Brunger, Ryde, and Merz groups pioneered a broadened joint refinement approach (energetically restrained refinement formalism) by using molecular dynamics, hybrid QM/MM methods, and/or a semiempirical (SE) QM method as the choice of the computational chemistry method.[15−25] A reliable local active-site structure (especially

Received: February 9, 2021
Published: May 25, 2021

https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796

containing metal ions) can be determined by QM/MM methods with low computational resources in such a multiscale quantum refinement approach. The balance between accuracy and cost can be adjusted by using different-sized QM regions and/or different levels for the QM method as defined by the user. Recently, as one important and successful application, the Ryde group applied their QM/MM quantum refinement method to override the widely accepted dicopper active site and suggested a monocopper active site in particulate methane monooxygenase (pMMO).[26] The monocopper active site in pMMO was later proven experimentally.[27] Moreover, his group further extended their joint quantum refinement method to X-ray-neutron refinement.[28] In addition, several other groups have also developed different quantum refinement methods and protocols,[29−31] including using fragmentation QM methods as the computational chemistry method.

On the other hand, ONIOM (our own *N*-layered integrated molecular orbital and molecular mechanics),[12,32,33] developed by Morokuma and coworkers, affords flexible and popular hybrid multiscale methods to investigate various chemical and biological systems. Apart from the typical two-layer QM/MM scheme ONIOM2(QM/MM), unique ONIOM methods also provide more flexible hybrid schemes with two, three, or multiple layers for different levels of computational chemistry methods such as QM, SE, and MM (e.g., ONIOM2(QM1/QM2) and ONIOM3(QM1/QM2/MM)).[12,33−35] ONIOM methods have been widely applied to evaluate energies, conduct geometry optimizations, and predict vibrational frequencies in many biomolecules.[12,32−43]

In this paper, we develop a joint X-ray quantum refinement method combined with flexible multiscale ONIOM schemes (including ONIOM2(QM/MM), ONIOM2(QM/SE), ONIOM3(QM/SE/MM), two-centered ONIOM (2C-ONIOM) variants, as well as unprecedented ONIOM2(WF-in-DFT/MM) and ONIOM3(WF-in-DFT/SE/MM) schemes) for metalloproteins. Apart from quantum refinements combined with the common two-layer QM/MM schemes, quantum refinements using various abovementioned ONIOM combinations of QM, SE, and MM were conducted to investigate the performance and reliability on the refined active-site structure(s) in two metalloproteins for the first time in this study. Geometries for three biomolecules (two metalloproteins and one peptide) were refined in this work by using various ONIOM-based or DFT methods with XRD experimental data. The key multiscale quantum refinement schemes in this study are briefly summarized in Section 2, and the computational details for our quantum refinement applications are then presented in Section 3. The key results and discussions for our quantum refinement applications on the two metalloproteins are discussed in Section 4.

2. THEORETICAL METHODS

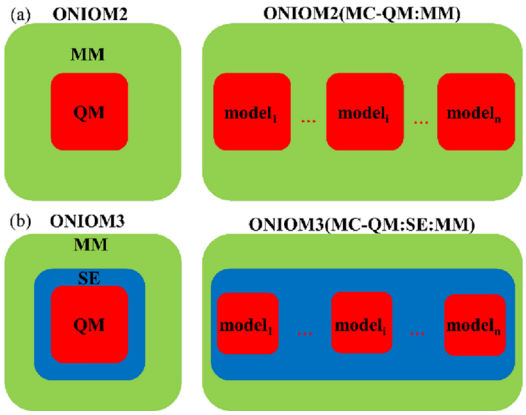
The multiscale quantum refinement method relies on geometry optimization based on combined gradient contributions from the computational chemistry and experimental (X-ray) parts, in which the total energy function of the whole system is expressed by eq 1.

$$E_{\text{total}} = E_{\text{comput}} + \omega_\alpha \times E_{\text{xray}} \tag{1}$$

where E_{comput} is the energy contribution from the computational chemistry method, E_{xray} is the energy contribution derived from the crystallographic penalty function, and ω_α is the weighting factor. The value of the weighting factor could be adjusted so as to increase or decrease the impact of the computational chemistry method in quantum refinements to achieve optimal refined structures. For the computational chemistry part, E_{comput} can be obtained from calculations using MM, SE, QM, fragmentation QM (or SE), hybrid QM/MM, or any multiscale ONIOM schemes depending on the accuracy and computational costs. A few multiscale ONIOM schemes, full SE, full QM, and fragmentation at the QM level, were combined and compared in this study. As ONIOM schemes are mainly adopted in this work, ONIOM schemes are briefly overviewed.[12]

As shown in Scheme 1a and eq 2, a high-level computational chemistry method is applied for the small model system (active site; red region), while a low-level computational chemistry method treats the large real system in the two-layer ONIOM scheme. The total energy of the two-layer ONIOM (such as ONIOM2(QM/MM) or ONIOM2(QM1/QM2)) schemes can be generalized as follows:

Scheme 1. (a) Conventional and (b) Multicentered (MC) ONIOM Methods for Two-Layer (ONIOM2) and Three-Layer (ONIOM3) Schemes



$$E_{\text{ONIOM2}} = E_{\text{high,model}} + E_{\text{low,real}} - E_{\text{low,model}} \tag{2}$$

Moreover, ONIOM schemes can be easily extended to three layers, e.g., ONIOM3(QM1/QM2/MM), see Scheme 1b and eq 3.[12]

$$E_{\text{ONIOM3}} = E_{\text{high,model}} + E_{\text{medium,intermediate}} - E_{\text{medium,model}} + E_{\text{low,real}} - E_{\text{low,intermediate}} \tag{3}$$

When some proteins contain two or multiple active sites with metal ion(s) and/or unnatural ligand(s), a larger model system including all active sites is desirable for quantum refinement calculations. However, this large model described by the high-level QM method should significantly increase computational costs. To alleviate the computational costs for those proteins without very close active sites, multicentered (MC) ONIOM schemes (ONIOM2(MC-QM:MM) and ONIOM3(MC-QM1:QM2:MM), Scheme 1) allow multiple model parts/fragments to be approximately treated together in a larger model system by the high-level approach, where the total energy of the larger model system with several fragments[44−46] can be described by a two-body expansion approximation (eq 4). When the active-site fragments are separated by a rather long distance, the total energy of the larger model system can be further simplified by energy

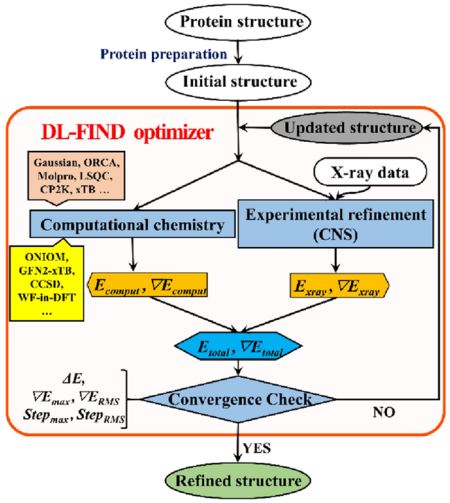
https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796

summation for each fragment with the high-level computational method and by approximately treating the interactions between different model fragments with the lower-level computational method in this MC ONIOM scheme.

$$E_{\text{high,model}} = \sum_{i} E_{\text{high,model(i)}} + \sum_{\text{i,j>i}} E_{\text{high,model(i,j)}} \quad (4)$$

The experimental part (E_{xray}), which is related to the difference between the observed and calculated X-ray structure factors, can be obtained from the Crystallography and NMR System (CNS) program.[47] The default maximum likelihood target using amplitudes[48,49] was used. Our multiscale quantum refinement schemes and interface were implemented in an open-source DL-FIND optimizer (Scheme 2),[50] which

Scheme 2. Workflow for our Multiscale Quantum Refinement Calculations



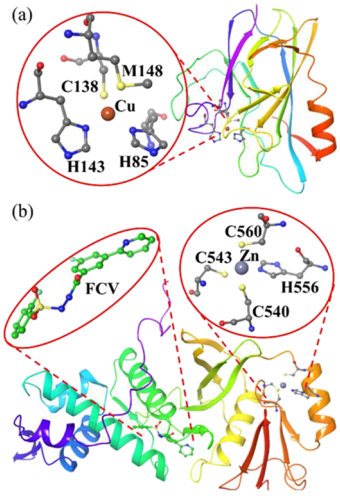
connects the CNS program and a few QM programs (such as Gaussian,[51] CP2K,[52] Molpro,[53] ORCA,[54] xTB,[55] LSQC[56]). In addition, the other QM programs can be easily combined and called through an external shell script provided by the user. The CNS program calculates only the crystallographic penalty function and its corresponding derivatives (with or without B factor correction) based on the given (initial or updated) structure and X-ray structure factors. The QM program evaluates the energy and gradients for the structure by the QM, QM/MM, or other computational chemistry methods. The DL-FIND optimizer then combines the energy and gradient components from the CNS and QM programs and minimizes E_{total} (eq 1) during the quantum refinement process, where the limited-memory L-BFGS algorithm[57,58] and delocalized internal coordinates[59−61] are preferentially employed. Iterations for the energy minimization are repeated until convergence of the root mean square (RMS) of the displacement (Step) and gradient (∇E), maximum (max) displacement and gradient, as well as variation of the energy (ΔE).

3. COMPUTATIONAL DETAILS

As shown in Scheme 2, a protein crystal structure was prepared and used as an initial structure for the quantum refinement calculations. The protein crystal structures and experimental data can be downloaded from the protein data bank (PDB).[62] In addition, the CNS ligand topology and parameter files for the protein structure can be obtained from the HIC-Up database,[2] and, for unknown ligands, these files can be generated through the AceDRG module[63] implemented in CCP4i2 7.0.[64]

Crystal structures of two metalloproteins (Scheme 3) were chosen to mainly evaluate the performance and reliability of

Scheme 3. (a) Rusticyanin (RC with a Cu(II) Ion) and (b) MYST Histone Acetyltransferase KAT6A/B (with a Zn(II) Ion and an Organic FCV Ligand/Inhibitor) in Complex with WM-1119



multiscale quantum refinements using different multiscale ONIOM combinations: (1) monocopper(II)-containing rusticyanin (RC) from *Thiobacillus ferrooxidans* (PDB ID: 1RCY; 1.90 Å resolution; ω_α value derived from CNS is 0.606)[65] and (2) zinc(II)-containing MYST histone acetyltransferase KAT6A/B in complex with WM-1119 (dual actives sites; PDB ID: 6CT2; 2.13 Å resolution; ω_α value derived from CNS is 1.667).[66] Another small protein amyloid-β(20−34) with L-isoAsp23 (PDB ID: 6NB9; 1.05 Å resolution; ω_α value derived from CNS is 0.269)[67] was also briefly studied (see details in the Supporting Information). These three protein structures were first prepared by determining the protonated state and the rotamer of some amino acid side chains (see details in the Supporting Information), as well as the addition of hydrogen atoms, for subsequent quantum refinement calculations combined with ONIOM schemes. The protonation of the protein is not required in standard crystallographic refinement. The protonated state of the titratable residues was assigned on the basis of the pKa results computed from the PROPKA 3.0 program at the pH of crystallization.[68] The rotamer of the amino acid side chain was determined by its nearby potential H-bonding and/or steric congestion. The addition of hydrogen atoms and optimization of the hydrogen-bond network were then carried out by using the PDB2PQR 2.0.0 program.[69]

Various computational methods were used to assess our multiscale quantum refinement schemes (Table 1). Our quantum refinement calculations combined with ONIOM-based QM/MM, QM/SE, and QM/SE/MM schemes (**M1**-**M6**) were mainly conducted, where efficient DFT methods (such as B3LYP,[70−72] B3LYP-D3[73] and M06-L[74] functionals with 6-31G(d)[75,76] and SDD[77] basis sets, denoted as BS1 and BS2, respectively) as the QM method, Grimme's tight-binding GFN2-xTB[55] as the SE method, and Amber94 force fields[78] as

https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796

Table 1. Computational Chemistry Methods in our Quantum Refinements[a,b,c]

	Methods
M1	ONIOM2(B3LYP/BS1:MM)[b]
M2	ONIOM2(B3LYP-D3/BS1:MM)
M3	ONIOM2(M06-L/BS1:MM)
M4	ONIOM2(B3LYP/BS2:MM)
M5	ONIOM3(B3LYP/BS1:SE:MM)[b]
M6	ONIOM2(B3LYP/BS1:SE)
M7	ONIOM2(CCSD-in-B3LYP/BS3:MM)[b]
M8	ONIOM3(CCSD-in-B3LYP/BS3:SE:MM)[b]

[a]Amber94 as the MM method and GFN2-xTB as the SE method. [b]Two-centered (2C) fragmentation applied in the model part (**2C-MX**). [c]BS1: 6-31G(d); BS2: SDD for Cu(II) and 6-31G(d) for the rest; BS3: def2-SVP for the WF-in-DFT (Zn(II) site) & B3LYP/BS1 (FCV).

the MM method were used. No cutoff for nonbonded (electrostatic and van der Waals) interactions was applied in the MM calculations. Apart from the popular DFT method as the QM level in the ONIOM scheme using Gaussian 16,[51] a recent projection-based quantum-embedding WF-in-DFT method[79−87] (CCSD-in-B3LYP[88] with def2-SVP[89,90] basis set (BS3) implemented in the Molpro 2019 program[91]) was also employed as the QM level in our ONIOM-based QM/MM and QM/SE/MM applications (**M7**-**M8**) for the first time, which can be regarded as pseudo-three-layer CCSD/B3LYP/MM and pseudo-four-layer CCSD/B3LYP/SE/MM methods, in which the key chemical bonds with the transition metals were described by a high-level coupled-cluster singles and doubles (CCSD) method.[85,86]

The convergence conditions (atomic unit) for our quantum refinement calculations were set to follow the default settings in many common QM codes (e.g., Gaussian 16): $\Delta E < 10^{-5}$, $\mathrm{Step}_{\max} < 1.8 \times 10^{-2}$, $\mathrm{Step}_{\mathrm{RMS}} < 1.2 \times 10^{-2}$, $\nabla E_{\max} < 4.5 \times 10^{-3}$, $\nabla E_{\mathrm{RMS}} < 3 \times 10^{-3}$. Such settings are quite similar to those used in the previous QR studies by the Ryde and Merz groups.[19,21] Two values of the weighting factor ω_α (those derived from CNS and 1.0) were applied and tested for our quantum refinement calculations in this proof-of-concept study (detailed results given in the Supporting Information). Our refinement result analysis (such as R factors, electron density maps, and real-space difference density Z score (RSZD)[92]) was finally conducted by using the Refmac5,[93] CCP4mg[94] and Edstats[92] modules implemented in CCP4i2 7.0.[64]

Two- and three-layer ONIOM schemes for our first application (RC,[65] Scheme 3a) were defined: the high-layer model part contains 117 atoms including 17 link atoms (seven neighboring residues within a radius of ∼4.0 Å to the Cu(II) center: PRO52, HIE85, SER86, CYM138, ILE140, HIE143, and MET148), and the medium-layer intermediate part for the three-layer ONIOM cases contains 449 atoms, including 15 link atoms and the high-layer atoms (all residues within a radius of ∼8.0 Å to the Cu(II) center). Protonation states of a few residues were set as HID39, HID57, HIE85, HID128, CYM138, and HIE143, whereas the other titratable residues remain in the standard protonation state. Three different optimized regions (R1a includes all atoms in the high layer, R2a includes all residues within a radius of ∼8.0 Å to the Cu(II) center, and R3a includes all atoms in the medium layer for the ONIOM3 cases; see the Supporting Information for details) around the copper active site were also set. ONIOM methods **M1**−**M4** with the R1a-optimized regions were first conducted to investigate the effects of different DFT functionals and basis sets. The effect of mechanical embedding (ME) and electrostatic embedding (EE) schemes was also compared for **M1** with the R1a- and R2a-optimized regions.[95] Moreover, quantum refinements using **M5** with the R1a- and R3a-optimized regions and **M6** with the R1a- and R2a-optimized regions were also performed.

Our second system (MYST histone acetyltransferase KAT6A/B with WM-1119)[66] contains two active sites: a zinc ion and an organic 3-fluoro-*N*′-[(2-fluorophenyl)-sulfonyl]-5-(pyridin-2-yl)benzohydrazide (FCV) ligand/inhibitor (Scheme 3b). This X-ray crystal structure contains 21 missing residues (485−505), which are terminal residues of the protein and very far away from the two active sites (roughly 26−37 Å). Therefore, we do not add these missing residues. The magnesium binding site was frozen and not refined in this study because main-group Mg ion has a simpler electronic structure (compared to transition metals) and thus can be treated reasonably by the force field.[96] Two- and three-layer ONIOM schemes for this system are defined as follows: the high-layer model part includes 112 atoms (zinc with its four neighboring residues: CYM540, CYM543, HID556, and CYM560 and an FCV ligand), and the medium-layer intermediate part for the three-layer cases contains 592 atoms, including 24 link atoms and the high-layer atoms (all residues within a radius of ∼3.0 Å to the FCV center and all residues within a radius of ∼5.0 Å to the Zn(II) center). Protonation states of a few residues were set as HID513, CYM540, CYM543, HID556, CYM563, HID579, HID586, HIE603, HIP627, HID743, HID755, and ASP520 (anionic amide form), while the other titratable residues remain in the standard protonation state. Three different optimized regions (R1b includes all atoms in the high layer, R2b includes all residues within a radius of ∼3.0 Å to the FCV center and all residues within a radius to ∼5.0 Å around the Zn(II) center, and R3b includes all atoms in the medium layer for ONIOM3 cases; see the Supporting Information for details) around the zinc and FCV active sites were also set. Quantum refinements using **M1** with the R1b- and R2b-optimized regions as well as **M5** with the R1b- and R3b-optimized regions were conducted. In addition, quantum refinements using **M2** with the R1b-optimized region were also performed to evaluate the effect of the dispersion correction. As the two FCV and zinc active sites are separated by more than ∼22 Å, quantum refinement calculations combined with a **2C**-ONIOM scheme (eq 5) were further applied and compared to the results obtained using the conventional (**1C**) ONIOM method to evaluate computational costs and the MC ONIOM approximation. For instance, the high layer, which contains 112 atoms (1148 basis functions at the B3LYP/6-31G(d) level) in the 1C-ONIOM approach, is divided into two smaller high-layer fragments in the 2C-ONIOM scheme: (1) a Zn(II) center (72 atoms, 710 basis functions) and (2) an FCV center (40 atoms, 435 basis functions). Due to the steep polynomial scaling for the QM method, the related computational cost to compute the whole high layer with 112 atoms is much larger than that to compute the two smaller high-layer fragments.[44] Energies/gradients of the two centers as well as interactions between the two centers (fragments) are described by the QM method in the 1C-ONIOM scheme. In comparison, only energies/gradients of the two centers are described by the high-level QM method in the 2C-ONIOM scheme, while long-range interactions

https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796

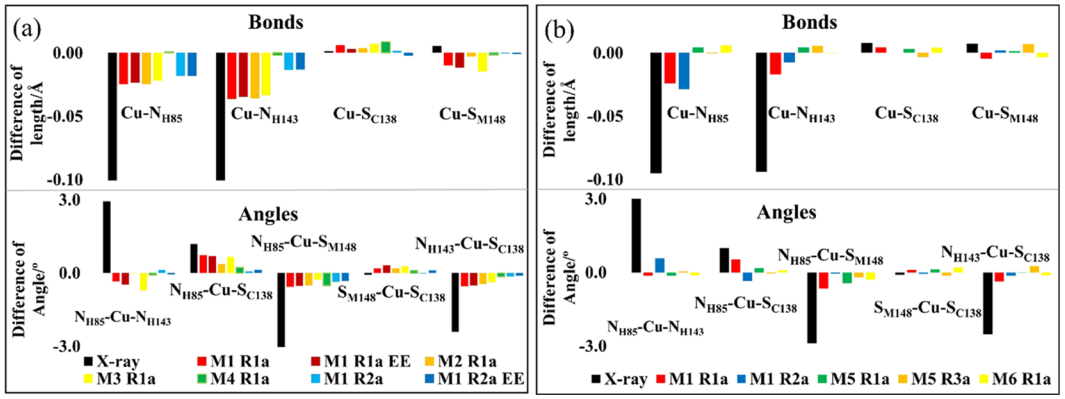


Figure 1. Difference in the key refined bond lengths and bond angles involving the Cu(II)-ligand bonds in rusticyanin protein by different ONIOM methods with ω_α values of (a) 1.0 and (b) 0.606 (detailed data are given in Tables S2, S5, and S6). The refined results for our theoretically most reliable **M6** scheme with the R2a-optimized region and their corresponding ω_α value were used as the reference to compare those refined by different ONIOM methods. Results for X-ray were taken from the experimental structure without our further refinement.

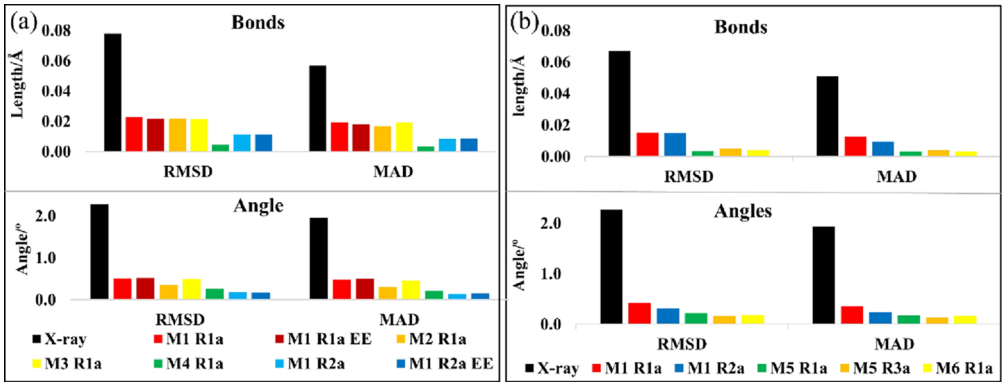


Figure 2. RMSD and MAD for the overall refined bond lengths and bond angles involving the Cu(II)-ligand bonds in rusticyanin protein by different ONIOM methods with ω_α values of (a) 1.0 and (b) 0.606 (detailed data are given in Tables S2, S5, and S6). The refined results for our theoretically most reliable **M6** scheme with the R2a-optimized region and their corresponding ω_α value were used as the reference to compare those refined by different ONIOM methods. Results for X-ray were taken from the experimental structure without our further refinement.

between the two remote centers and its gradients are approximated by the lower computational method (see SI for details). Therefore, such 2C-ONIOM approximation saves some computational time.

$$E_{\mathrm{QM,Zn+FCV}} = E_{\mathrm{QM,Zn}} + E_{\mathrm{QM,FCV}} + E_{\mathrm{QM,interaction}} \approx E_{\mathrm{QM,Zn}} + E_{\mathrm{QM,FCV}} + E_{\mathrm{SE\ or\ MM,interaction}} \quad (5)$$

Moreover, to describe chemical bonding with transition metals using the CCSD method, the quantum refinement calculations using ONIOM2(2C-CCSD-in-B3LYP:MM) (**2C-M7**) and ONIOM3(2C-CCSD-in-B3LYP:SE:MM) (**2C-M8**) with the R1b-optimized region were also carried out in our second application. Furthermore, for such refinement calculations by the CCSD-in-DFT method, Zn, S_{C540}, S_{C543}, N_{H556}, and S_{C560} atoms were chosen as the embedded atoms at the CCSD level. In addition, to save a substantial amount of computational time for the CCSD-in-DFT method, the refined results obtained by the **2C-M1** and **2C-M5** methods were taken as our initial quantum refinement structure. Currently, our quantum refinement calculations can apply the lower-cost CCSD with the small basis set only, although the CCSD method with a def2-SVP basis set (rather than the golden CCSD(T) method with a larger basis set) could be less reliable than some DFT methods. Such a combination is a proof of concept (the first example) to apply WF-in-DFT for quantum refinement calculations. We believe that the CCSD(T)-in-DFT implementation with larger basis sets for quantum refinement calculations will become feasible and practicable in the future.

Furthermore, the highest-level computational combination used in this study has been regarded as the best and most reliable computational method: ONIOM2(QM:SE) (**M6**) scheme with the R2a-optimized region for the first application and the ONIOM3(QM:SE:MM) (**M5**) scheme with the R1b-optimized region for the second application (due to too expensive ONIOM2(QM:SE) scheme and larger derivation found in the flexible zinc center for the larger R3b-optimized region, see the discussion below for details). These refined results were used to compute RMS deviation (RMSD) and mean absolute deviation (MAD) for the other refined structures as well as the crystal stuctures.

4. RESULTS AND DISCUSSION

4.1. Copper-Dependent RC Protein. Prior to our detailed comparison of various quantum refinement schemes, the effects of the different DFT functionals and basis sets on the Cu(II) active site in the RC protein were first evaluated by using ONIOM2(QM/MM) schemes (**M1-M4**) with an ω_α value of 1.0. When the B3LYP, B3LYP-D3, or M06-L functional was used as the QM method in our ONIOM2-

https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796

Table 2. RSZD Scores of Residues around the Cu(II) Binding Site in the Rusticyanin Protein from the Six Quantum Refinement Approaches with ω_α Values of 0.606 and 1.0 (in Parenthesis)[a]

	M1 R1a	**M1** R2a	**M5** R1a	**M5** R3a	**M6** R1a	**M6** R2a	X-ray
PRO52	1.4(1.3)	1.4(1.4)	1.4(1.3)	1.5(1.3)	1.4(1.3)	1.6(1.4)	1.3
HIE85	0.4(0.2)	0.2(0.2)	0.7(0.3)	0.2(0.4)	0.6(0.4)	0.3(0.4)	1.4
SER86	0.6(0.5)	1.0(0.8)	0.6(0.6)	0.7(0.6)	0.7(0.6)	0.9(0.8)	2.3
CYM138	0.8(0.8)	0.7(0.8)	0.9(0.9)	0.9(1.0)	0.8(0.9)	0.7(0.8)	1.0
ILE140	0.3(0.3)	0.4(0.5)	0.3(0.2)	0.4(0.4)	0.2(0.2)	0.5(0.5)	0.3
HIE143	0.4(0.4)	0.3(0.2)	0.3(0.3)	0.1(0.1)	0.4(0.3)	0.2(0.1)	1.5
MET148	0.9(0.9)	0.6(0.7)	0.8(0.8)	0.7(0.7)	0.8(0.8)	0.7(0.6)	0.8
Cu(II)	3.0(3.1)	3.2(3.2)	3.0(3.0)	3.2(3.3)	3.0(3.0)	3.2(3.2)	2.7
sum	7.8(7.5)	7.8(7.8)	8.0(7.4)	7.7(7.8)	7.9(7.5)	8.1(7.8)	11.3

[a]Results for X-ray were taken from the experimental structure without our further refinement.

(QM/MM) scheme (Figure 1a), the refined bond lengths and bond angles involving the Cu(II) and its ligated residues were found to be quite similar for these different DFT functionals. The Cu-S_{M148} bond length was found to be the only ligated bond length that changed most; it changed by 0.02 Å at most, and the largest discrepancy in the bond angles was less than 0.7° for these methods. In addition, compared to the highest-level **M6** R2a scheme with an ω_α value of 1.0, the RMSD and MAD analysis for these key geometrical features in the R1a-optimized regions using different DFT functionals as the high-level QM method of the ONIOM2(QM/MM) scheme also gives small RMSD (bonds: 0.02 Å; angles: 0.4−0.5°) and MAD (bonds: 0.02 Å; angles: 0.3−0.5°) values. In contrast, the X-ray structure gives the highest RSMD and MAD values (RMSD: 0.07 Å and 2.3°; MAD: 0.05 Å and 1.9°; see Figure 2a). These computational results show an insignificant effect of the chosen DFT functionals on the copper active-site bonding. In comparison, a slightly larger effect of the basis set for the Cu(II) metal on the bond lengths and bond angles was observed (**M4**), which gives a more similar local structure compared to the **M6** R2a results (RMSD: 0.00 Å and 0.3°; MAD: 0.00 Å and 0.2°) with an ω_α value of 1.0.[97] Moreover, to investigate the influence of dispersion correction on the Cu(II) active site, the optimized structures of the R1a-optimized region in the protein and in the gas phase using the B3LYP and B3LYP-D3 methods were compared. As shown in Table S3, the maximum difference for the ligated bond lengths and angles in the protein refined by **M1** and **M2** schemes first without B factor correction is less than 0.01 Å and 0.4°, respectively. Such key geometric differences in the gas phase by the B3LYP and B3LYP-D3 methods are increased up to 0.05 Å and 4.8°. Therefore, the dispersion correction was found to have a smaller effect on the Cu(II) active site in the protein compared to that in the gas phase. Owing to the similar Cu-ligand bonding obtained by using different DFT functionals or basis sets; the B3LYP/6-31G(d) was then used as the QM method in subsequent various multiscale quantum refinement calculations.

In addition, the effect of ME and EE schemes in the ONIOM(B3LYP/MM) method on the active-site bonding was also examined by using the R1a- and R2a-optimized regions with an ω_α value of 1.0 (Figure 1a and Figure 2a). Our computational refinement results reveal that the ONIOM-EE scheme mainly leads to slightly longer Cu−N_{H143} and Cu−S_{M148} bonds by 0.02 and 0.01 Å, respectively, compared to those computed by the corresponding ONIOM-ME scheme. Additionally, the two ONIOM-ME and ONIOM-EE schemes show very similar bond angles with a maximum difference of no more than 0.2° (Table S5).[98] Hence, the effect of EE on the active-site bonding was found to be insignificant in this Cu(II) protein, and the ONIOM-ME scheme was applied to the following quantum refinements with various multiscale ONIOM schemes due to a lower computational cost for the efficient ONIOM-ME scheme.

Notably, for our quantum refinement calculations with an ω_α value of either 0.606 (suggested by CNS) or 1.0 (Table S6), the bonding on the copper site is very similar with a maximum distance difference of 0.03 Å and a maximum angle difference of 0.4° for different quantum refinement approaches (**M1**, **M5**, and **M6**). In addition, for different ω_α cases, their RSZD scores were computed to be very close with a deviation of less than 0.6 at the same computational scheme (Table 2).[99] Hereafter, the refined results for the ω_α value of 0.606 will be mainly discussed.

The key results obtained for the two- and three-layer ONIOM-based quantum refinements with different optimized regions are also summarized in Figure 1b. Generally, various quantum refinement schemes (**M1**, **M5**, and **M6**) gave similar results for the active-site bonding. For instance, compared to the X-ray crystal structure, the main geometric difference for these multiscale quantum refinement schemes lies in the Cu−N bonds between the copper and two ligated histidines (85 and 143), which are 0.07−0.13 Å longer than those in the crystal structure. In comparison, the two refined Cu−S bonds calculated by various schemes and the crystal structure are very similar (with a deviation of less than 0.01 Å). Additionally, the refined N_{H85}−Cu−N_{H143} and N_{H85}−Cu−S_{M148} angles by various schemes were found to slightly deviate from the crystal structure by 2.5−3.9°. Compared to the crystal structure, the difference in the refined N_{H143}−Cu−S_{C138} angle obtained by different schemes was 1.8−2.7°. The differences in the refined N_{H85}−Cu−S_{C138} and S_{M148}−Cu−S_{C138} angles were found to be even smaller (0.1−1.2°). Moreover, RMSD and MAD (**M6** R2a as reference) for these key ligated bond lengths and bond angles determined by various schemes are quite similar (RMSD: 0.00−0.02 Å and 0.0−0.5°; MAD: 0.00−0.02 and 0.0−0.5°, see Figure 2b), which further supports a very similar refined copper active-site structure obtained by various quantum refinement schemes.[100]

Compared to the most computationally expensive ONIOM2(B3LYP/SE) method (**M6**) with a larger optimized (R2a) region, the ONIOM3(B3LYP/SE/MM) method (**M5**) was found to give the closest refinement results (identical RMSD and MAD values for ligated bond lengths, only 0.1−0.3° difference for the RMSD and MAD of bond angles) with a much lower computational cost.

https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796

As shown in Figures 3 and S4 and S5, after these quantum refinement calculations, the disappearance of the negative (red

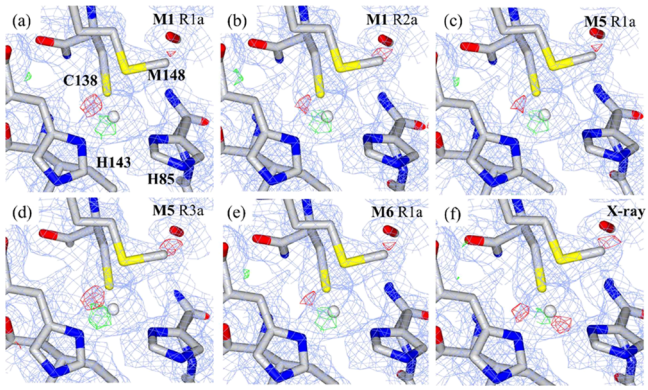


Figure 3. Structures for the Cu(II) binding site in the rusticyanin protein from various quantum refinements with an ω_α value of 0.606 using (a) **M1** (with the R1a-optimized region), (b) **M1** (with the R2a-optimized region), (c) **M5** (with the R1a-optimized region), (d) **M5** (with the R3a-optimized region), and (e) X-ray crystal structure, including the electron density map ($2mF_o-DF_c$ maps, contoured at 1.0 σ (blue), mF_o-DF_c maps, contoured at +3.0 σ (green), and mF_o-DF_c maps, contoured at −3.0 σ (red)). Results for the X-ray were taken from the experimental structure without our further refinement.

contours) difference density between copper and HIE85 in the resultant electron density maps possibly results from the elongated $Cu-N_{HIE85}$ (by 0.07−0.13 Å, see Tables S6 and S8). It indicates that the discrepancy with experimental observations was reduced through quantum refinements (**M1**, **M5**, and **M6**). As shown in Table 2, the maximum absolute RSZD score for the copper ion and its seven neighboring residues in the R1a-optimized region are decreased after our quantum refinement calculations, compared to the RSZD score for the X-ray crystal structure. The RSZD score for the sum of the ligand and seven residues reduced from 11.3 (the crystal structure) to 7.4−8.1 after different quantum refinement calculations. Interestingly, different quantum refinement schemes gave very similar RSZD scores, indicating similar refined geometries. Moreover, the RSZD scores of our refined HIE85, SER86, and HIE143 residues after different quantum refinements were decreased to 0.7−1.2, 1.3−1.7, and 1.1−1.4 in comparison with the crystal structure, respectively. Furthermore, the RSZD score for the ONIOM3(DFT/SE/MM) case (**M5** with R3a) is the lowest (7.7). Likewise, the scores for the ONIOM2(DFT/SE) and ONIOM2(DFT/MM) cases (**M6** and **M1**: 7.8−8.1) are very similar to that for the ONIOM3(DFT/SE/MM) case. These overall results suggest that various ONIOM-based quantum refinement schemes can improve the local structures.

Furthermore, from an energy point of view, compared to the X-ray crystal structure, the strain energy[19] of the R1a-optimized active-site regions (which was calculated as the QM energy difference between the fully optimized R1a-optimized geometry in vacuum and the quantum-refined R1a-optimized geometry extracted from the protein at the B3LYP/6-31G(d) level) is considerably decreased by more than 175.1 kcal/mol with the ONIOM-based schemes through these quantum refinement processes (Figure 4). The strain energies (Table S7) for the R1a-optimized active-site region with an ω_α value of 0.606 are slightly smaller than the cases with an ω_α value of 1.0 because of the similar weighting for the

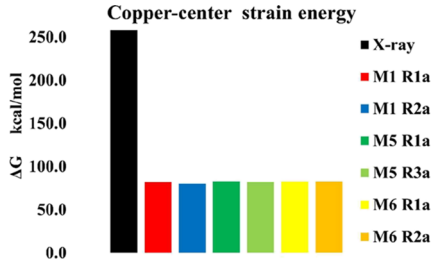


Figure 4. Computed strain energy (ΔE, kcal/mol) at the B3LYP/6-31G(d) level for the R1a-optimized active-site region in the copper(II) protein for various quantum refinement schemes with an ω_α value of 0.606 (detailed data are given in Table S7). Results for the X-ray were taken from the experimental structure without our further refinement.

experimental part. Overall, these multiscale quantum refinements improve the Cu-ligand bonding in the active site. Moreover, our quantum refinement protocol based on combining various multiscale ONIOM schemes was shown to be feasible and comparable for this Cu(II) protein application. On the basis of a balance between the computational costs and accuracy, the ONIOM3(DFT/SE/MM) method is also suggested to be a good computational approach for quantum refinement.

Compared to the experimental structure (X-ray crystal structure without further refinement), the RSZD scores of the Cu(II) center are reduced from 11.3 (X-ray crystal structure) to 8.1 after quantum refinement calculations using our theoretically most reliable scheme (**M6** with the R2a-optimized region, see Table 2). According to the decreases of the positive and negative contours in the electron density maps (Figure 3), the refined local structures by various quantum refinements (**M1**, **M5**, and **M6**) fit the XRD data better than the X-ray structure. Moreover, the strain energy of the R1a-optimized active-site region is much higher for the X-ray structure and can be considerably decreased by about 175.5 kcal/mol through the same quantum refinement scheme **M6** (Figure 4). Therefore, these computational results reveal that our ONIOM-based quantum refinements can obviously improve the local Cu(II) binding site from the X-ray crystal structure.

4.2. Histone Acetyltransferase. The influence of the dispersion correction on the geometry in the Zn(II) active site (the R1b-optimized regions) in the protein and gas phase is first compared. The largest difference for the ligated bond lengths and angles in the protein refined by **M1** and **M2** schemes first without the B factor correction is less than 0.02 Å and 0.7° (Table S9), respectively. In contrast, these key geometric differences in the gas phase by the B3LYP and B3LYP-D3 methods become 0.03 Å and 5.6°. Moreover, our quantum refinement calculations using **M1** and **M2** schemes show that the difference of their RSZD scores is less than 0.1 (Table S10) and their electron density maps are similar (Figure S6). Because the dispersion correction was found to have no significant effect on the Zn(II) active site in the protein compared to the gas phase, the B3LYP method was then used as the QM method in the subsequent various multiscale quantum refinement calculations. However, we recommend the inclusion of dispersion correction in practical quantum refinement calculations.

The key results for the conventional two- and three-layer ONIOM-based quantum refinements (**M1** and **M5**) are

https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796

compared and summarized in Figures S7 and S8 and Tables S12−S17. In comparison with our quantum refinement calculations with an ω_α value of 1.667 (suggested by CNS) and 1.0 (Tables S12 and S13), their maximum difference of the metal−ligand bond lengths on the Zn(II) site between various quantum refinement schemes is less than 0.03 Å and that of the related angle is less than 0.7°. The computed RSZD scores for these ω_α values are also similar with a deviation of less than 0.2 (Table S11).[101] These results reveal that our quantum refinement calculations give similar local geometries for the two different ω_α cases. Hereafter, the refined results for the ω_α value of 1.667 are mainly discussed.

However, the computed RZSD scores for the Zn(II) center by different quantum refinement schemes (**M1** and **M5**) become higher than the X-ray crystal structure by 0.4−1.7 (see details in the Supporting Information). As shown in Figures S10 and S11, a less positive difference density around the zinc ion was also observed after these quantum refinement calculations. Moreover, the key atomic displacement for these refined results (Figure S12a and Tables S21 and S22) with respect to the X-ray crystal structure shows that the zinc metal moves only 0.07−0.08 Å, whereas the Zn-ligated $N\varepsilon_{HID556}$ atom was found to have the largest displacement (0.18−0.22 Å) compared to the other Zn-ligated S_{CYM540}, S_{CYM543}, and S_{CYM560} atoms (0.12−0.17 Å) and the Zn(II) metal. Notably, as reported previously and suggested by one reviewer,[19,20] geometry optimization including a restrained refinement of individual B factors for all atoms can improve the RSZD score for the Zn(II) binding site. Therefore, all subsequent results and discussion are based on quantum refinement calculations with an ω_α value of 1.667 including individual B factors for all atoms.[102]

Compared to the X-ray crystal structure, the refined Zn-S_{C540} bond obtained by different multiscale schemes (**M1** and **M5**) becomes longer by 0.09−0.11 Å, while the Zn−S_{C543}, Zn−N_{H556}, and Zn−S_{C560} bonds become shorter by 0.13−0.14, 0.04−0.07, and 0.09−0.11 Å, respectively (Figure 5). In addition, our different quantum refinement calculations generally provide similar active-site bonding based on their RMSD, MAD, and RSZD scores. Compared to our theoretically most reliable computational combination (**1C-M5** R1b), the RMSD and MAD values for the bonds involving the zinc ion refined by different approaches deviate by 0.01 and 0.01 Å only (Figure 6), respectively. Similarly, the RMSD and MAD values for the bond angles refined by different approaches deviate by approximately 0.5−1.1 and 0.3−1.0°, respectively. Likewise, the RSZD scores of the zinc and FCV centers refined by different approaches deviate by less than 0.2 and 0.4 (Table 3), respectively. The electron density maps are almost identical relative to those by **1C-M5** R1b (Figures 7 and 8). These computational results show again that different quantum refinement schemes generally lead to the similar active-site geometry.

As shown in Figure 7, after these quantum refinement calculations by including the individual B factors, the negative (red contours) difference density around CYM543, HID556, and zinc ion mostly disappears in the resultant electron density maps. The positive (green contours) difference electron density close to HID556 becomes smaller. Similar improvement was also observed in the FCV center (as shown in Figure 8); the negative and positive difference electron densities were reduced around the S and O atoms. It indicates that the discrepancy with experimental observations was reduced

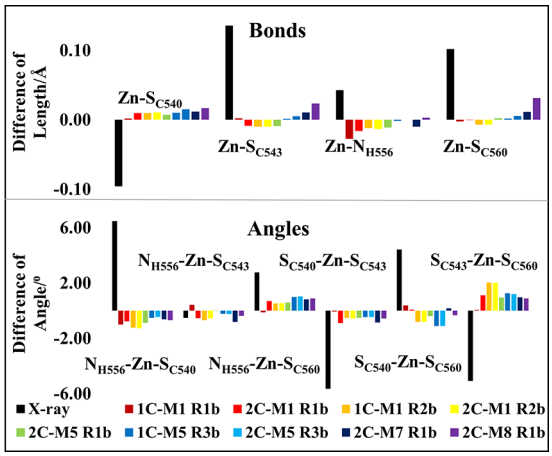


Figure 5. Difference in the refined key bond lengths and bond angles involving the zinc(II) metal−ligand in histone acetyltransferase protein using different ONIOM methods with an ω_α value of 1.667 (detailed data are given in Table S28). The refined results for our theoretically most reliable **1C-M5** scheme with the R1b-optimized region were used as the reference to compare those refined by different ONIOM methods. Individual B factors for all atoms were included in quantum refinement calculations. Results for the X-ray were taken from the experimental structure without our further refinement.

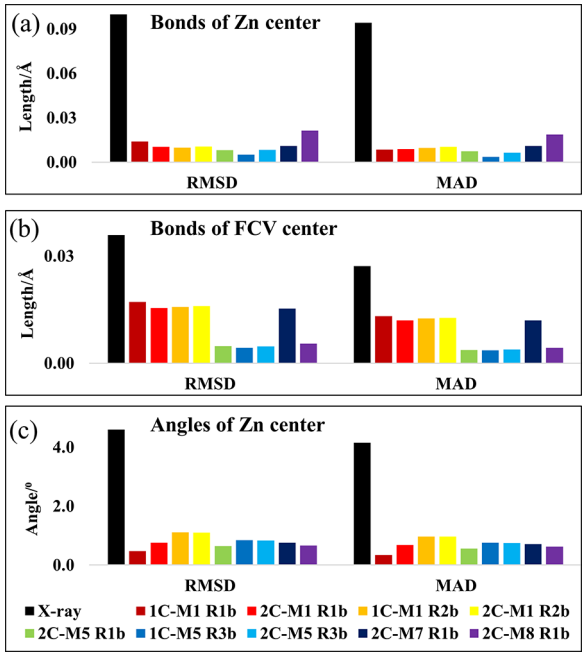


Figure 6. RMSD and MAD analysis of the refined (a) bond lengths for the Zn(II) center; (b) bond lengths for the FCV center; and (c) bond angles around the Zn(II) center of the histone acetyltransferase obtained using different ONIOM methods with an ω_α value of 1.667, which are compared to that obtained from our theoretical most reliable **1C-M5** scheme with the R1b-optimized region (detailed data are given in Tables S28 and S29). Individual B factors for all atoms were included in quantum refinement calculations. Results for the X-ray were taken from the experimental structure without our further refinement.

through quantum refinements (**M1** and **M5**). As shown in Table 3, the maximum absolute RSZD score for the zinc ion and its four neighboring residues in the R1b-optimized region are decreased after our quantum refinement calculations,

https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796

Table 3. RSZD Scores of Residues around the Zn(II) Binding Site in Histone Acetyltransferase from Various Quantum Refinement Approaches with an ω_α Value of 1.667[a]

	M1 R1b	M1 R2b	M5 R1b	M5 R3b	M7 R1b[b]	M8 R1b[b]	X-ray
CYM540	1.5(1.5)	1.5(1.5)	1.6(1.5)	1.6(1.7)	(1.6)	(1.6)	0.6
CYM543	1.2(1.3)	1.3(1.3)	1.2(1.1)	1.2(1.2)	(1.3)	(1.2)	1.5
HID556	0.3(0.3)	0.4(0.4)	0.3(0.3)	0.4(0.4)	(0.3)	(0.3)	3.0
CYM560	1.3(1.3)	1.2(1.2)	1.4(1.3)	1.2(1.3)	(1.3)	(1.4)	1.7
Zn(II)	2.1(2.2)	2.4(2.3)	2.1(2.1)	2.2(2.2)	(2.1)	(1.5)	2.7
sum[c](Zn center)	6.4(6.6)	6.8(6.7)	6.6(6.3)	6.6(6.8)	(6.6)	(6.0)	9.5
FCV	1.2(1.2)	1.0(1.0)	1.3(1.2)	0.9(0.9)	(1.2)	(1.2)	1.7

[a]Individual B factors for all atoms were included in quantum refinement calculations. The two-center (2C) ONIOM approximation was also used (values given in parenthesis). RSZD scores of the FCV binding site were also given. Results for X-ray were taken from the experimental structure without our further refinement [b]Values computed by the **2C-M7** and **2C-M8** schemes with the CCSD-in-DFT method as the QM method to describe the key Zn(II) metal center. [c]The sum for the Zn center is the sum of CYM540, CYM543, HID556, CYM560, and Zn.

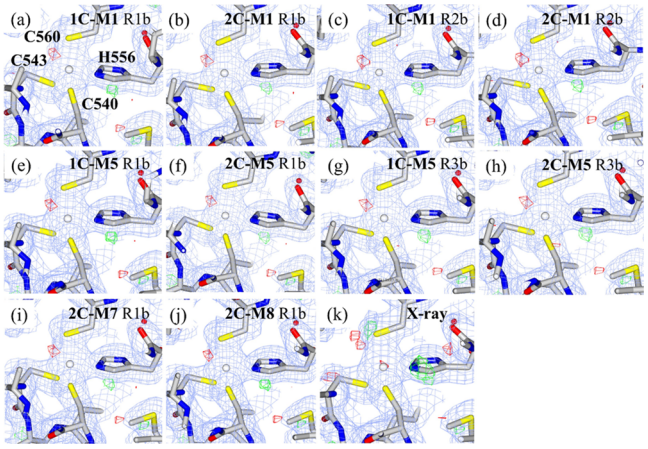


Figure 7. Structures for the zinc(II) active site in the MYST histone acetyltransferase KAT6A/B protein from various quantum refinement schemes with an ω_α value of 1.667 using (a) **1C-M1** (with the R1b-optimized region), (b) **2C-M1** (with the R1b-optimized region), (c) **1C-M1** (with the R2b-optimized region), (d) **2C-M1** (with the R2b-optimized region), (e) **1C-M5** (with the R1b-optimized region), (f) **2C-M5** (with the R1b-optimized region), (g) **1C-M5** (with the R3b-optimized region), (h) **2C-M5** (with the R3b-optimized region), (i) **2C-M7** (with the R1b-optimized region), (j) **2C-M8** (with the R1b-optimized region), and (k) the X-ray crystal structure, including the electron density map ($2mF_o-DF_c$ maps, contoured at 1.0 σ (blue), mF_o-DF_c maps, contoured at +3.0 σ (green), and mF_o-DF_c maps, contoured at −3.0 σ (red)). Individual B factors for all atoms were included in quantum refinement calculations. Results for the X-ray were taken from the experimental structure without our further refinement.

compared to the RSZD score for the X-ray crystal structure. The RSZD score for the sum of zinc and four residues reduced from 9.5 (the crystal structure) to 6.4−6.8, and that of the FCV ligand reduced from 1.7 to 0.9−1.3 after different quantum refinement calculations (**M1** and **M5**). Hence, compared to the crystal structure, these results suggest that our quantum refinement calculations improve the local structure of the zinc and FCV binding sites, particularly the flexible zinc(II) binding site by including the individual B factors.

Furthermore, compared to the X-ray crystal structure, the strain energy[19] of the R1b-optimized region (which was calculated as the QM energy difference between the fully optimized R1b-optimized geometry in vacuum and the quantum-refined R1b-optimized geometry extracted from the protein at the B3LYP/6-31G(d) level) is considerably

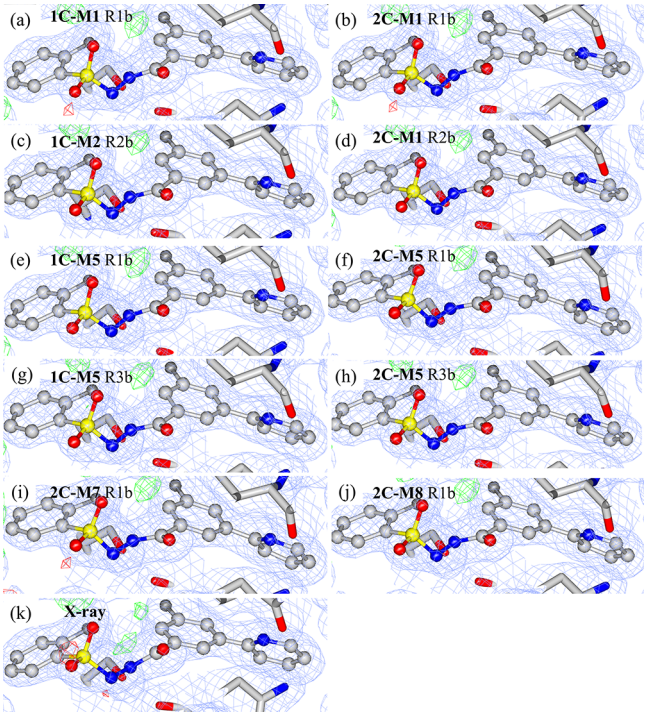


Figure 8. Structures for the FCV binding site in the MYST histone acetyltransferase KAT6A/B protein from various quantum refinement calculations with an ω_α value of 1.667 using (a) **1C-M1** (with the R1b-optimized region), (b) **2C-M1** (with the R1b-optimized region), (c) **1C-M1** (with the R2b-optimized region), (d) **2C-M1** (with the R2b-optimized region), (e) **1C-M5** (with the R1b-optimized region), (f) **2C-M5** (with the R1b-optimized region), (g) **1C-M5** (with the R3b-optimized region), (h) **2C-M5** (with the R3b-optimized region), (i) **2C-M7** (with the R1b-optimized region), (j) **2C-M8** (with the R1b-optimized region), and (k) the X-ray crystal structure, including the electron density map ($2mF_o-DF_c$ maps, contoured at 1.0 σ (blue), mF_o-DF_c maps, contoured at +3.0 σ (green), and mF_o-DF_c maps, contoured at −3.0 σ (red)). Individual B factors for all atoms were included in quantum refinement calculations. Results for the X-ray were taken from the experimental structure without our further refinement.

decreased by 17.4−23.0 and 64.1−66.6 kcal/mol for the FCV and zinc centers, respectively, after these quantum refinement calculations (**M1** and **M5**; Figure 9 and Table S30).

In comparison to the experimental structure (X-ray crystal structure in the absence of further refinement), the RSZD scores of the Zn(II) center were found to be decreased from 9.5 (X-ray crystal structure) to 6.6 after the quantum

https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796

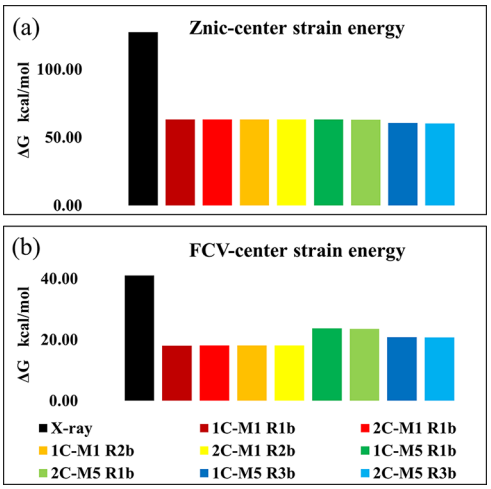


Figure 9. Strain energy (ΔE, kcal/mol) at the B3LYP/6-31G(d) level for the R1b-optimized region for (a) Zn(II) center and (b) FCV center in histone acetyltransferase determined by various quantum refinement schemes. Results for the X-ray were taken from the experimental structure without our further refinement.

refinement (**M5**) scheme (with the R1b-optimized region, Table 3). In addition, the RSZD score of the FCV ligand is modestly reduced from 1.7 (X-ray crystal structure) to 1.3 after the same quantum refinement (**M5**) approach. The refined local structures by various quantum refinements (**M1** and **M5**) fit the XRD data better than the X-ray structure according to the electron density maps (Figures 7 and 8). Furthermore, the strain energy of the R1b-optimized active-site regions for the Zn(II) and FCV centers refined by the **M5** scheme is considerably lower by 64.0 and 17.4 kcal/mol (Figure 9), respectively, than those in the X-ray crystal structure. Consequently, the local Zn(II) and FCV binding sites can be apparently improved by our ONIOM-based quantum refinements including the individual B factors.

Histone acetyltransferase contains a Zn(II)-binding site and one FCV inhibitor-binding site, which are separated by approximately 22 Å. In addition to the conventional ONIOM scheme, the 2C-ONIOM approximation was also employed to describe the two active sites for histone acetyltransferase to reduce computational costs. Our quantum refinement calculations using **M1** with the R1b-optimized region (using 24 CPUs, Intel(R) Xeon(R) Gold 5118 CPU @ 2.30 GHz) require about 34 h 52 m and 42 h 20 m for the 2C-ONIOM and conventional (1C) ONIOM schemes, respectively. These results show a lower computational cost for the 2C-ONIOM scheme. As shown in Figures 5 and 6, the deviation for the Zn-ligated bonds is less than 0.01 Å and that for the Zn-ligated bond angles is less than 1.0° between the conventional (1C-) and 2C-ONIOM schemes (**M1** and **M5**). In addition, these two types of the ONIOM schemes give a small difference for the strain energy[19] (0.0−0.3 kcal/mol, Figure 9). The RSZD scores are almost identical with a difference less than 0.3 (Table 3). Moreover, the ONIOM3-(2C-DFT/SE/MM) (**2C-M5**) quantum refinement scheme, which provides a similar active-site bonding (RMSD: 0.01 Å and 0.6°; MAD: 0.01 Å and 0.6°; identical R factors; deviation of RSZD scores: < 0.3; the strain energy difference: < 0.3 kcal/mol) to that obtained by the conventional ONIOM3 scheme (**1C-M5**) can be an efficient quantum refinement method for proteins with several far active sites. Long-range interactions between the active sites can be approximately treated by the GFN2-xTB method in such a multicenter ONIOM approximation. Therefore, these computational results show that the 2C-ONIOM approximation can achieve very similar quantum refined results with a lower computational cost when multiple active sites in the biomolecules are not very close to each other. On the basis of a balance between the computational costs and accuracy, the 2C-ONIOM3(DFT/SE/MM) is also suggested to be a good computational method for quantum refinement.

Moreover, compared to our refined structure by the **1C-M5** scheme, the CCSD-in-DFT method as the high-level QM method in our two- and three-layer ONIOM schemes with the 2C-ONIOM approximation (**2C-M7** and **2C-M8**) to describe the zinc center leads to the shorter Zn-N_{HID} bond and the longer Zn-S_{CYM} bonds by 0.01 and 0.01−0.03 Å only, respectively (Table S28). The RMSD and MAD values for the bonds involving the zinc ion refined by the computational approaches with the CCSD-in-DFT method deviate by 0.01 and 0.01−0.02 Å only (Table S28), respectively, compared to **1C-M5**. Likewise, the RMSD and MAD values for the bond angles refined by **2C-M7** and **2C-M8** deviate by approximately 0.7−0.8 and 0.6−0.7° (Table S28), respectively.[103] Moreover, the RSZD scores for the Zn(II) center by **2C-M7** and **2C-M8** were decreased to 0.0−0.6 (Table 3), and the corresponding electron density maps are almost identical (Figure 7), relative to those by **1C-M5**. These RSZD results suggest that our quantum refinement calculations combined with the CCSD-in-DFT method (**2C-M7** and **2C-M8**) slightly improves the local Zn-binding site compared to quantum refinements with **1C-M5**.

Apart from the two abovementioned applications using the ONIOM-based quantum refinement approaches, to extend and assess our quantum refinement with complete DFT,[54,104] fragmentation DFT,[56,105−107] or GFN2-xTB methods,[55] a small amyloid-β(20−34) protein with L-isoAsp23 was also refined (see the Supporting Information for details).

5. CONCLUSIONS

We have developed a multiscale quantum refinement method for metalloprotein structures by combining several multiscale ONIOM schemes and XRD refinement techniques. Compared to the standard crystallographic refinement, the classical force fields were replaced by multiscale ONIOM schemes to aid in giving more reasonable chemical structures, especially for active sites described by the QM method, along with the experimental (XRD) data. This approach takes advantage of various flexible ONIOM schemes and experimental (XRD) information, in which the challenging transition-metal binding site is described by a relatively high-accuracy and efficient QM method (e.g., DFT methods) and the remaining structures/interactions are approximated by faster low-level computational method(s) (such as SE and/or MM methods). Quantum refinement combined with two- and three-layer ONIOM schemes was first applied to refine the local metal binding site in one Cu-protein system. Our quantum refinement using the high-level QM/SE combination method yields our theoretically most reliable results based on the electron density, R factors, and RSZD scores. Additionally, our computational results show that different DFT functionals and the choice of the embedding scheme used in our ONIOM calculations yield similar Cu(II)-binding structures. Moreover, the QM/SE/MM method was found to provide a very good

https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796

refined protein structure with lower computational costs, compared to the expensive QM/SE method. Alternatively, to treat proteins with multiple local active sites, our second application for one Zn-containing protein with an inhibitor was examined by using the 2C-ONIOM approximation, which was found to speed up quantum refinement calculations with a similar accuracy, compared to the conventional ONIOM method. Although the long-range interactions between the two active sites are not described by the high-level QM method in this 2C-ONIOM scheme, such interactions can be still approximated by the lower-level computational method. Moreover, the recent quantum-embedding CCSD-in-DFT method was introduced as the QM method in our two- and three-layer ONIOM methods to describe the key Zn(II) metal center in our proof-of-concept multiscale quantum refinement calculations. Apart from the common quantum refinement combined with two-layer QM/MM method, our quantum refinement combined with several ONIOM schemes (two-layer 2C-ONIOM2(QM:MM), ONIOM2(QM:SE); three-layer ONIOM3(QM:SE:MM), 2C-ONIOM3(QM:SE:MM); pseudo-three-layer ONIOM3(WF-in-DFT:MM); pseudo-four-layer ONIOM4(WF-in-DFT:SE:MM)) were applied to evaluate the performance and reliability of various quantum refinement schemes on two metalloproteins for the first time. Considering the computational costs and accuracy, the ONIOM3(QM/SE/MM)-based refinement scheme was suggested to be a good choice. In principle, different QM methods (including fragmentation QM methods,[39,107−117] local correlation QM methods[118−122] as well as quantum embedding QM-in-QM methods[79−87]) can also replace ONIOM or become one of the computational methods used in the ONIOM schemes for quantum refinement calculations. We believe that multiscale quantum refinements can enable ones to refine more reasonable structures of metalloprotein (including those that contain inhibitors, substrates, drug candidates, etc.).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00148.

Details for the protein structure preparation, CNS refinement preparation and molecular mechanics force field parameters for our quantum refinement calculations, and Cartesian coordinates (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Lung Wa Chung − *Shenzhen Grubbs Institute, Department of Chemistry and Guangdong Provincial Key Laboratory of Catalysis, Southern University of Science and Technology, Shenzhen 518055, China;* orcid.org/0000-0001-9460-7812; Email: oscarchung@sustech.edu.cn

Authors

Zeyin Yan − *Shenzhen Grubbs Institute, Department of Chemistry and Guangdong Provincial Key Laboratory of Catalysis, Southern University of Science and Technology, Shenzhen 518055, China*

Xin Li − *Shenzhen Grubbs Institute, Department of Chemistry and Guangdong Provincial Key Laboratory of Catalysis, Southern University of Science and Technology, Shenzhen 518055, China;* orcid.org/0000-0003-4026-9365

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c00148

Notes

The authors declare no competing financial interest.

Our interface is freely available and may be accessed at https://github.com/oscarchung-lab/ONIOM_QR.

■ ACKNOWLEDGMENTS

We gratefully acknowledge the financial support from the National Natural Science Foundation of China (21873043 and 21933003), Southern University of Science and Technology, the Shenzhen Nobel Prize Scientists Laboratory Project (C17783101), Guangdong Provincial Key Laboratory of Catalysis (2020B121201002), and Natural Science Foundation of Shenzhen Innovation Committee (JCYJ20170817104736009). We thank the Center for Computational Science and Engineering of Southern University of Science and Technology for partly supporting this work as well as the anonymous reviewers for many suggestions. This work is dedicated to memory of Professor Keiji Morokuma sensei.

■ REFERENCES

(1) Kleywegt, G. J.; Jones, T. A. Databases in protein crystallography. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **1998**, *54*, 1119−1131.

(2) Kleywegt, G. J. Crystallographic refinement of ligand complexes. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **2007**, *63*, 94−100.

(3) Hu, L. H.; Ryde, U. Comparison of Methods to Obtain Force-Field Parameters for Metal Sites. *J. Chem. Theory Comput.* **2011**, *7*, 2452−2463.

(4) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(5) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.

(6) Monard, G.; Merz, K. M. Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Acc. Chem. Res.* **1999**, *32*, 904−911.

(7) Gao, J.; Truhlar, D. G. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.

(8) Friesner, R. A.; Guallar, V. Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389−427.

(9) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185−199.

(10) Hu, H.; Yang, W. T. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573−601.

(11) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198−1229.

(12) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z. F.; Liu, F. Y.; Li, H. B.; Ding, L. N.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678−5796.

(13) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.* **2015**, *115*, 6217−6263.

(14) Jack, A.; Levitt, M. Refinement of Large Structures by Simultaneous Minimization of Energy and R-Factor. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1978**, *34*, 931−935.
(15) Brunger, A. T.; Kuriyan, J.; Karplus, M. Crystallographic R factor refinement by molecular dynamics. *Science* **1987**, *235*, 458−460.
(16) Brunger, A. T.; Karplus, M.; Petsko, G. A. Crystallographic Refinement by Simulated Annealing - Application to Crambin. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1989**, *45*, 50−61.
(17) Brunger, A. T.; Krukowski, A.; Erickson, J. W. Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1990**, *46*, 585−593.
(18) Weis, W. I.; Brunger, A. T.; Skehel, J. J.; Wiley, D. C. Refinement of the influenza virus hemagglutinin by simulated annealing. *J. Mol. Biol.* **1990**, *212*, 737−761.
(19) Ryde, U.; Olsen, L.; Nilsson, K. Quantum chemical geometry optimizations in proteins using crystallographic raw data. *J. Comput. Chem.* **2002**, *23*, 1058−1070.
(20) Ryde, U.; Nilsson, K. Quantum chemistry can locally improve protein crystal structures. *J. Am. Chem. Soc.* **2003**, *125*, 14232−14233.
(21) Yu, N.; Yennawar, H. P.; Merz, K. M. Refinement of protein crystal structures using energy restraints derived from linear-scaling quantum mechanics. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **2005**, *61*, 322−332.
(22) Yu, N.; Li, X.; Cui, G. L.; Hayik, S. A.; Merz, K. M. Critical assessment of quantum mechanics based energy restraints in protein crystal structure refinement. *Protein Sci.* **2006**, *15*, 2773−2784.
(23) Li, X.; Hayik, S. A.; Merz, K. M. QM/MM X-ray refinement of zinc metalloenzymes. *J. Inorg. Biochem.* **2010**, *104*, 512−522.
(24) Fu, Z.; Li, X.; Miao, Y. P.; Merz, K. M. Conformational Analysis and Parallel QM/MM X-ray Refinement of Protein Bound Anti-Alzheimer Drug Donepezil. *J. Chem. Theory Comput.* **2013**, *9*, 1686−1693.
(25) Fu, Z.; Li, X.; Merz, K. M. Accurate Assessment of the Strain Energy in a Protein-Bound Drug Using QM/MM X-ray Refinement and Converged Quantum Chemistry. *J. Comput. Chem.* **2011**, *32*, 2587−2597.
(26) Cao, L. L.; Caldararu, O.; Rosenzweig, A. C.; Ryde, U. Quantum Refinement Does Not Support Dinuclear Copper Sites in Crystal Structures of Particulate Methane Monooxygenase. *Angew. Chem. Int. Ed.* **2018**, *57*, 162−166.
(27) Ross, M. O.; MacMillan, F.; Wang, J. Z.; Nisthal, A.; Lawton, T. J.; Olafson, B. D.; Mayo, S. L.; Rosenzweig, A. C.; Hoffman, B. M. Particulate methane monooxygenase contains only mononuclear copper centers. *Science* **2019**, *364*, 566−570.
(28) Caldararu, O.; Oksanen, E.; Ryde, U.; Hedegard, E. D. Mechanism of hydrogen peroxide formation by lytic polysaccharide monooxygenase. *Chem. Sci.* **2019**, *10*, 576−586.
(29) Hsiao, Y. W.; Sanchez-Garcia, E.; Doerr, M.; Thiel, W. Quantum refinement of protein structures: implementation and application to the red fluorescent protein DsRed.M1. *J. Phys. Chem. B* **2010**, *114*, 15413−15423.
(30) Zheng, M.; Moriarty, N. W.; Xu, Y. T.; Reimers, J. R.; Afonine, P. V.; Waller, M. P. Solving the scalability issue in quantum-based refinement: Q|R#1. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **2017**, *73*, 1020−1028.
(31) Zheng, M.; Reimers, J. R.; Waller, M. P.; Afonine, P. V. Q|R: quantum-based refinement. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **2017**, *73*, 45−52.
(32) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A multilayered integrated MO +MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ oxidative addition. *J. Phys. Chem.* **1996**, *100*, 19357−19363.
(33) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct.: THEOCHEM* **1999**, *461*, 1−21.
(34) Vreven, T.; Frisch, M. J.; Kudin, K. N.; Schlegel, H. B.; Morokuma, K. Geometry optimization with QM/MM methods II: Explicit quadratic coupling. *Mol. Phys.* **2006**, *104*, 701−714.
(35) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815−826.
(36) Maseras, F.; Morokuma, K. IMOMM: A New Integrated Ab-Initio + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comput. Chem.* **1995**, *16*, 1170−1179.
(37) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. Geometry optimization with QM/MM, ONIOM, and other combined methods I. Microiterations and constraints. *J. Comput. Chem.* **2003**, *24*, 760−769.
(38) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. The ONIOM method: its foundation and applications to metalloenzymes and photobiology. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 327−350.
(39) Canfield, P.; Dahlbom, M. G.; Hush, N. S.; Reimers, J. R. Density-functional geometry optimization of the 150,000-atom photosystem-I trimer. *J. Chem. Phys.* **2006**, *124*, No. 024301.
(40) Morokuma, K.; Wang, Q.; Vreven, T. Performance evaluation of the three-layer ONIOM method: Case study for a zwitterionic peptide. *J. Chem. Theory Comput.* **2006**, *2*, 1317−1324.
(41) Lundberg, M.; Sasakura, Y.; Zheng, G.; Morokuma, K. Case Studies of ONIOM(DFT:DFTB) and ONIOM(DFT:DFTB:MM) for Enzymes and Enzyme Mimics. *J. Chem. Theory Comput.* **2010**, *6*, 1413−1427.
(42) Sousa, S. F.; Ribeiro, A. J. M.; Neves, R. P. P.; Brás, N. F.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, *7*, No. e1281.
(43) Nogueira, J. J.; Rassbach, S.; Ochsenfeld, C.; Gonzalez, L. Effect of DNA Environment on Electronically Excited States of Methylene Blue Evaluated by a Three-Layered QM/QM/MM ONIOM Scheme. *J. Chem. Theory Comput.* **2018**, *14*, 4298−4308.
(44) Hopkins, B. W.; Tschumper, G. S. A multicentered approach to integrated QM/QM calculations Applications to multiply hydrogen bonded systems. *J. Comput. Chem.* **2003**, *24*, 1563−1568.
(45) Hopkins, B. W.; Tschumper, G. S. Multicentred QM/QM methods for overlapping model systems. *Mol. Phys.* **2005**, *103*, 309−315.
(46) Bates, D. M.; Smith, J. R.; Tschumper, G. S. Efficient and Accurate Methods for the Geometry Optimization of Water Clusters: Application of Analytic Gradients for the Two-Body:Many-Body QM:QM Fragmentation Method to $(H_2O)_n$, n=3-10. *J. Chem. Theory Comput.* **2011**, *7*, 2753−2760.
(47) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **1998**, *54*, 905−921.
(48) Pannu, N. S.; Read, R. J. Improved structure refinement through maximum likelihood. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1996**, *52*, 659−668.
(49) Adams, P. D.; Pannu, N. S.; Read, R. J.; Brunger, A. T. Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 5018−5023.
(50) Kastner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *J. Phys. Chem. A* **2009**, *113*, 11856−11865.
(51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J.

https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796

V.; Izmaylov, A. F.; Sonnenberg, J. L.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; 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, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. A.03*; Gaussian Inc.: Wallingford CT, 2016.

(52) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. CP2K: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 15−25.

(53) Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schutz, M. Molpro: a general-purpose quantum chemistry program package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 242−253.

(54) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1327.

(55) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB-An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.

(56) Li, W.; Chen, C. H.; Zhao, D. B.; Li, S. H. LSQC: Low Scaling Quantum Chemistry Program. *Int. J. Quantum Chem.* **2015**, *115*, 641−646.

(57) Nocedal, J. Updating Quasi-Newton Matrices with Limited Storage. *Math. Comput.* **1980**, *35*, 773−782.

(58) Liu, D. C.; Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Program.* **1989**, *45*, 503−528.

(59) Pulay, P.; Fogarasi, G. Geometry optimization in redundant internal coordinates. *J. Chem. Phys.* **1992**, *96*, 2856−2860.

(60) Baker, J.; Kessi, A.; Delley, B. The generation and use of delocalized internal coordinates in geometry optimization. *J. Chem. Phys.* **1996**, *105*, 192−212.

(61) Peng, C. Y.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *J. Comput. Chem.* **1996**, *17*, 49−56.

(62) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. The Protein Data Bank. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **2002**, *58*, 899−907.

(63) Long, F.; Nicholls, R. A.; Emsley, P.; Grazulis, S.; Merkys, A.; Vaitkus, A.; Murshudov, G. N. AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **2017**, *73*, 112−122.

(64) Potterton, L.; Agirre, J.; Ballard, C.; Cowtan, K.; Dodson, E.; Evans, P. R.; Jenkins, H. T.; Keegan, R.; Krissinel, E.; Stevenson, K.; Lebedev, A.; McNicholas, S. J.; Nicholls, R. A.; Noble, M.; Pannu, N. S.; Roth, C.; Sheldrick, G.; Skubak, P.; Turkenburg, J.; Uski, V.; von Delft, F.; Waterman, D.; Wilson, K.; Winn, M.; Wojdyr, M. CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **2018**, *74*, 68−84.

(65) Walter, R. L.; Ealick, S. E.; Friedman, A. M.; Blake, R. C.; Proctor, P.; Shoham, M. Multiple wavelength anomalous diffraction (MAD) crystal structure of rusticyanin: A highly oxidizing cupredoxin with extreme acid stability. *J. Mol. Biol.* **1996**, *263*, 730−751.

(66) Baell, J. B.; Leaver, D. J.; Hermans, S. J.; Kelly, G. L.; Brennan, M. S.; Downer, N. L.; Nguyen, N.; Wichmann, J.; McRae, H. M.; Yang, Y.; Cleary, B.; Lagiakos, H. R.; Mieruszynski, S.; Pacini, G.; Vanyai, H. K.; Bergamasco, M. I.; May, R. E.; Davey, B. K.; Morgan, K. J.; Sealey, A. J.; Wang, B.; Zamudio, N.; Wilcox, S.; Garnham, A. L.; Sheikh, B. N.; Aubrey, B. J.; Doggett, K.; Chung, M. C.; de Silva, M.; Bentley, J.; Pilling, P.; Hattarki, M.; Dolezal, O.; Dennis, M. L.; Falk, H.; Ren, B.; Charman, S. A.; White, K. L.; Rautela, J.; Newbold, A.; Hawkins, E. D.; Johnstone, R. W.; Huntington, N. D.; Peat, T. S.; Heath, J. K.; Strasser, A.; Parker, M. W.; Smyth, G. K.; Street, I. P.; Monahan, B. J.; Voss, A. K.; Thomas, T. Inhibitors of histone acetyltransferases KAT6A/B induce senescence and arrest tumour growth. *Nature* **2018**, *560*, 253−257.

(67) Warmack, R. A.; Boyer, D. R.; Zee, C. T.; Richards, L. S.; Sawaya, M. R.; Cascio, D.; Gonen, T.; Eisenberg, D. S.; Clarke, S. G. Structure of amyloid-β(20-34) with Alzheimer's-associated isomerization at Asp23 reveals a distinct protofilament interface. *Nat. Commun.* **2019**, *10*, 3357.

(68) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.

(69) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665−W667.

(70) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.

(71) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(72) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(73) 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*, No. 154104.

(74) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods .9. Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724−728.

(75) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213−222.

(76) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. A complete basis set model chemistry .5. Extensions to six or more heavy atoms. *J. Chem. Phys.* **1996**, *104*, 2598−2619.

(77) Stoll, H.; Fuentealba, P.; Schwerdtfeger, P.; Flad, J.; Szentpály, L. V.; Preuss, H. Cu and Ag as one-valence-electron atoms: CI results and quadrupole corrections for Cu_2, Ag_2, CuH, and AgH. *J. Chem. Phys.* **1984**, *81*, 2732−2736.

(78) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(79) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050−8053.

(80) Libisch, F.; Huang, C.; Carter, E. A. Embedded Correlated Wavefunction Schemes: Theory and Applications. *Acc. Chem. Res.* **2014**, *47*, 2768−2775.

(81) Jacob, C. R.; Neugebauer, J. Subsystem density-functional theory. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 325−362.

(82) Wesolowski, T. A.; Shedge, S.; Zhou, X. W. Frozen-Density Embedding Strategy for Multilevel Simulations of Electronic Structure. *Chem. Rev.* **2015**, *115*, 5891−5928.

(83) Sun, Q. N.; Chan, G. K. L. Quantum Embedding Theories. *Acc. Chem. Res.* **2016**, *49*, 2705−2712.

(84) Hegely, B.; Nagy, P. R.; Kallay, M. Dual Basis Set Approach for Density Functional and Wave Function Embedding Schemes. *J. Chem. Theory Comput.* **2018**, *14*, 4600−4615.

(85) Lee, S. J. R.; Welborn, M.; Manby, F. R.; Miller, T. F. Projection-Based Wavefunction-in-DFT Embedding. *Acc. Chem. Res.* **2019**, *52*, 1359−1368.

https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796

(86) Lee, S. J. R.; Ding, F.; Manby, F. R.; Miller, T. F. Analytical gradients for projection-based wavefunction-in-DFT embedding. *J. Chem. Phys.* **2019**, *151*, No. 064112.
(87) Jones, L. O.; Mosquera, M. A.; Schatz, G. C.; Ratner, M. A. Embedding Methods for Quantum Chemistry: Applications from Materials to Life Sciences. *J. Am. Chem. Soc.* **2020**, *142*, 3281−3295.
(88) Schutz, M.; Lindh, R.; Werner, H. J. Integral-direct electron correlation methods. *Mol. Phys.* **1999**, *96*, 719−733.
(89) 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.
(90) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057−1065.
(91) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Györffy, W.; Kats, D.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bennie, S. J.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Lee, S. J. R.; Liu, Y.; Lloyd, A. W.; Ma, Q.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Welborn, M. *MOLPRO, version 2019.2, a package of ab initio programs*, Wallingford, CT, 2019.
(92) Tickle, I. J. Statistical quality indicators for electron-density maps. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **2012**, *68*, 454−467.
(93) Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **2011**, *67*, 355−367.
(94) McNicholas, S.; Potterton, E.; Wilson, K. S.; Noble, M. E. M. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **2011**, *67*, 386−394.
(95) Rossbach, S.; Ochsenfeld, C. Influence of Coupling and Embedding Schemes on QM Size Convergence in QM/MM Approaches for the Example of a Proton Transfer in DNA. *J. Chem. Theory Comput.* **2017**, *13*, 1102−1107.
(96) Li, P. F.; Merz, K. M. Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* **2017**, *117*, 1564−1686.
(97) Interestingly, all these quantum refinement calculations (**M1**-**M4**) give very similar local active-site structures and R/R_{free} factors (0.162−0.163/0.220−0.221, Table S2).
(98) The two ONIOM-ME and ONIOM-EE schemes show essentially identical R and R_{free} factors.
(99) Moreover, the R/R_{free} factors remain unchanged for different ω_α values at the same computational scheme, because only a small region in the protein was optimized.
(100) The slightly decreased R/R_{free} factors by 0.001−0.002/0.002−0.008 (Table S6) further support the similar and slightly improved active-site structure refined by different ONIOM schemes.
(101) In addition, the R/R_{free} factors for these ω_α values are computed to be identical.
(102) Additional quantum refinement calculations using larger ω_α values (3.0, 5.0, 10.0, and 20.0) have been conducted (Figure S16 and Tables S26 and S27). However, these computational results do not have significant improvement around the Zn center.
(103) The R/R_{free} factors refined by **2C-M7** and **2C-M8** are the same as those for the **1C-M5** results (Table S28).
(104) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree-Fock exchange. *Chem. Phys.* **2009**, *356*, 98−109.
(105) Li, W.; Li, S. H.; Jiang, Y. S. Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules. *J. Phys. Chem. A* **2007**, *111*, 2193−2199.
(106) Yang, Z.; Hua, S. G.; Hua, W. J.; Li, S. H. Low-Lying Structures and Stabilities of Large Water Clusters: Investigation Based on the Combination of the AMOEBA Potential and Generalized Energy-Based Fragmentation Approach. *J. Phys. Chem. A* **2010**, *114*, 9253−9261.
(107) Li, S. H.; Li, W.; Ma, J. Generalized Energy-Based Fragmentation Approach and Its Applications to Macromolecules and Molecular Aggregates. *Acc. Chem. Res.* **2014**, *47*, 2712−2720.
(108) Yang, W.; Lee, T. S. A density-matrix divide-and-conquer approach for electronic structure calculations of large molecules. *J. Chem. Phys.* **1995**, *103*, 5674−5678.
(109) Collins, M. A.; Cvitkovic, M. W.; Bettens, R. P. A. The Combined Fragmentation and Systematic Molecular Fragmentation Methods. *Acc. Chem. Res.* **2014**, *47*, 2776−2785.
(110) Fedorov, D. G.; Asada, N.; Nakanishi, I.; Kitaura, K. The Use of Many-Body Expansions and Geometry Optimizations in Fragment-Based Methods. *Acc. Chem. Res.* **2014**, *47*, 2846−2856.
(111) Gao, J. L.; Truhlar, D. G.; Wang, Y. J.; Mazack, M. J. M.; Loffler, P.; Provorse, M. R.; Rehak, P. Explicit Polarization: A Quantum Mechanical Framework for Developing Next Generation Force Fields. *Acc. Chem. Res.* **2014**, *47*, 2837−2845.
(112) He, X.; Zhu, T.; Wang, X. W.; Liu, J. F.; Zhang, J. Z. H. Fragment Quantum Mechanical Calculation of Proteins and Its Applications. *Acc. Chem. Res.* **2014**, *47*, 2748−2757.
(113) Merz, K. M. Using Quantum Mechanical Approaches to Study Biological Systems. *Acc. Chem. Res.* **2014**, *47*, 2804−2811.
(114) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643−5677.
(115) Macetti, G.; Genoni, A. Quantum Mechanics/Extremely Localized Molecular Orbital Method: A Fully Quantum Mechanical Embedding Approach for Macromolecules. *J. Phys. Chem. A* **2019**, *123*, 9420−9428.
(116) Guo, W.; Wu, A.; Zhang, I.; Xu, X. XO: an extended ONIOM method for accurate and efficient modeling of large systems. *J. Comput. Chem.* **2012**, *33*, 2142−2160.
(117) Guo, W.; Wu, A.; Xu, X. XO: An extended ONIOM method for accurate and efficient geometry optimization of large molecules. *Chem. Phys. Lett.* **2010**, *498*, 203−208.
(118) Saebo, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213−236.
(119) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps-A systematic infrastructure for reduced-scaling electronic structure methods II. Linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* **2016**, *144*, No. 024109.
(120) Schwilk, M.; Ma, Q. L.; Koppl, C.; Werner, H. J. Scalable Electron Correlation Methods. 3. Efficient and Accurate Parallel Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD). *J. Chem. Theory Comput.* **2017**, *13*, 3650−3675.
(121) Ma, Q. L.; Werner, H. J. Explicitly correlated local coupled-cluster methods using pair natural orbitals. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1371.
(122) Nagy, P. R.; Samu, G.; Kallay, M. Optimization of the Linear-Scaling Local Natural Orbital CCSD(T) Method: Improved Algorithm and Benchmark Applications. *J. Chem. Theory Comput.* **2018**, *14*, 4193−4215.

https://doi.org/10.1021/acs.jctc.1c00148
J. Chem. Theory Comput. 2021, 17, 3783−3796