Impacts of Polarizable Continuum Models on the SCF Convergence and DFT Delocalization Error of Large Molecules

Fangning Ren, Fang Liu*

Department of Chemistry, Emory University, Atlanta, Georgia, 30322

Abstract: Advances in algorithm developments have enabled density functional theory (DFT) description of large molecules, including whole proteins, but the self-consistent field (SCF) convergence issues often hamper practical applications. The conductor-like polarizable continuum model (CPCM), although initially introduced as an implicit solvent model, was reported to improve SCF convergence in some large molecules. However, the underlying mechanisms and applicable use cases were unclear. We investigated the impacts of CPCM on the SCF convergence of 25 peptides and found that the CPCM only effectively reduced the SCF iterations for molecules with charge separations (e.g., the zwitterionic form of peptides) but had little effect on non-charge-separated molecules. We observed that CPCM increased the HOMO-LUMO gap of both the zwitterionic and non-charge-separated molecules, but only the charge-separated molecules suffered from the vanishing HOMO-LUMO gap problem in the gas phase which is the origin of the convergence issue. We revealed CPCM's gap-opening mechanism as the selective stabilization/destabilization of molecular orbitals (MO) based on their local electrostatic environment. Compared to level-shifting, a traditional SCF improvement technique, CPCM has superior performance because the stabilization/destabilization of MOs is consistent through SCF

* Electronic mail: fang.liu@emory.edu

iterations. Finally, we examined CPCM's impacts on DFT density delocalization error (DDE) when used as an SCF accelerator. CPCM can mitigate the DDE and reproduce the density-derived properties (e.g., dipole moments) matching high-level methods when a very low dielectric constant is used but tends to over-localize the electron density at higher dielectric constants.

I. INTRODUCTION

Recent advances in large-scale quantum chemistry development have enabled numerous *ab initio* calculations of large molecular and biomolecular systems, from the prediction of properties (protein structures,[1] protein-ligand binding affinities[2]) to the investigation of dynamics.[3] Density functional theory (DFT) is still the most widely employed method in large-scale quantum chemistry calculations for its balance of cost and accuracy in main group chemistry. Practical DFT calculations use density functional approximations (DFAs) for the exchange-correlation (xc) functional. However, most DFAs suffer from self-interaction errors, also called the delocalization error (DE). The DE affects the accuracy of DFT prediction of energetic properties (electron affinities,[4] band gaps,[5] barrier heights,[6] and dissociation energies[7]), density-derived properties (partial charges, molecular multipole moments), and geometry (bond lengths, bond angles).[8] For large molecules like biological systems, the DE is also known to cause DFT self-consistent field (SCF) convergence issues due to the vanishing HOMO-LUMO gap.[9] With a lot of computational resources wasted for each failed SCF, the convergence issue severely impacts the efficiency and usability of DFT in investigating biological systems.

One simple yet effective remedy to DFT convergence issues in large molecules is to apply dielectric screening via a polarizable continuum model (PCM), first introduced by Atony and Grimme[1] and later investigated by Payne *et al*.[2, 9] It was found that the PCM field could open the HOMO-LUMO gap and avoid the frequent mixture of virtual and occupied orbitals between SCF

iterations. However, Kohn-Sham (KS) orbital energies are little affected by the PCM reaction field in other contexts, such as the optimal tuning of long-range corrected functionals in PCM.[10, 11] The origin of PCM's effects on SCF convergence improvement and DE reduction has not been thoroughly investigated for large molecules. It is unclear whether PCM can be used as a universal SCF accelerator or only effective for specific large molecules. How the electronic structure of the large molecules can be impacted by the PCM field has not been scrutinized either.

Dale and Johnson recently showed that PCM reduces DEs and reproduces linear fractional charge behavior in DFAs because the polarization energy can be expressed as a quadratic concave down function[12]. Due to the overstabilization of fractional charge caused by DE, DFAs show a convex departure from the expected linear fractional charge behavior between N and N+1 electron systems. Therefore, common corrections counteract the DE by introducing a concave down fractional charge term that localizes electron density. Typical examples include the Hubbard U-correction in DFT+U,[13] an admixture of Hartree-Fock exchange globally[14] or with range separation in hybrid functionals,[15] and a general polynomial in jm-DFT.[16] The quadratic concave-down function form of PCM also has electron localization effects. Dale and Johnson demonstrated the impacts of PCM in reproducing the linear fractional charge behavior for 147 small molecules. The work provides an insightful explanation of the counterintuitive density localization observed in PCM calculations and has discussed the possibility of using PCM to reduce DEs to improve the prediction of molecular properties.

In this work, we investigate the origin of PCM's SCF convergence effects and identify applicable molecular systems. We have found that the PCM-induced SCF convergence improvement is most prominent in molecules with charge separations, such as zwitterions. We will analyze the formalism of the conductor-like PCM (CPCM)[17, 18] and show its preference for converging to a

polarized solute and the selective stabilization of molecular orbitals (MOs) based on the local electrostatic environment. Then we will use a set of small zwitterionic molecules and zwitterionic peptides to show the effects of CPCM in reducing density-driven DE and improving SCF convergence, which is superior to the traditional SCF convergence technique, level shifting.[19] We will show that two effects of PCM lead to improved SCF convergence: (1) an increased HOMO-LUMO gap that reduces the mixing of occupied and virtual orbitals, and (2) the stabilization/destabilization of selective MOs is consistent through all SCF iterations.

II. THEORY

Here, we analyze the electron localization effects of CPCM for a neutral molecule in the molecular cavity (fused Van der Waals spheres) and the impacts on KS orbital energies. It is worth noting that ε is used to denote dielectric constant, whereas ϵ is used to denote MOs.

1. CPCM formalism and its preference for a polarized solute

In CPCM, the reaction potential generated by the solute charge distribution is described in terms of an apparent surface charge (ASC) distribution spread over the solute cavity surface. Outside the surface is a dielectric continuum with a dielectric constant ε, whereas the cavity containing the solute has a unit dielectric constant. The cavity boundary is usually discretized into M surface segments (tesserae) with an ASCs, $\{q_k^{\infty}\}$, that describe the electric field of the polarized continuum. The values**, q,** are determined with a set of linear equations:[20, 21]

$$\mathbf{Aq} = -\frac{\varepsilon - 1}{\varepsilon}\mathbf{V}. \tag{1}$$

Here, $\mathbf{A} \in \mathbb{R}^{M\times M}$ is the Coulomb interaction between unit polarization charges on two cavity tesserae, and $\mathbf{V} \in \mathbb{R}^{M}$is a vector that represents the solute electrostatic potential at each tessera. The CPCM solvation energy at equilibrium is given by

$$\Delta G^{\mathrm{CPCM}} = -\frac{1}{2}\frac{\varepsilon - 1}{\varepsilon}\mathbf{V}^T\mathbf{A^{-1}}\mathbf{V} \tag{2}$$

In practical CPCM calculations, **V** is evaluated from the contribution of solute nuclei charge and solute electron density. To simplify the discussion, we represent the overall charge distribution of a neutral molecule as a dipole moment $\vec{\mu} = \mu\vec{l}$ at the center of mass of the molecule ($\vec{R}_{\mathrm{COM}}$), where $\vec{l}$ is the unit vector in the direction of the dipole moment and $\mu = |\vec{\mu}|$. Then the solute electrostatic potential can be expressed as

$$V_k = \frac{\vec{\mu}\cdot(\vec{r}_k - \vec{R}_{\mathrm{COM}})}{\left|\vec{r}_k - \vec{R}_{\mathrm{COM}}\right|^3} = \mu V_k^0, \tag{3}$$

where $V_k^0 = \frac{\vec{l}\cdot(\vec{r}_k - \vec{R}_{\mathrm{COM}})}{|\vec{r}_k - \vec{R}_{\mathrm{COM}}|^3}$ is the electrostatic potential of a unit dipole, $\vec{l}$, in the same direction as $\vec{\mu}$. The CPCM solvation energy of this molecule is then given by

$$\Delta G^{\mathrm{CPCM}} = -\mu^2\frac{\varepsilon - 1}{\varepsilon}\left(\frac{1}{2}(\mathbf{V^0})^T\mathbf{A^{-1}}\mathbf{V^0}\right) = -\mu^2\frac{\varepsilon - 1}{\varepsilon}C. \tag{4}$$

Since the Coulomb matrix, **A**, is a symmetric positive-definite matrix,[22] its inverse $\mathbf{A^{-1}}$ is also symmetric positive-definite, and the product $\mathbf{x^T A^{-1} x}$ for an arbitrary vector $\mathbf{x} \neq \mathbf{0}$ is always positive. Hence, the term inside the parathesis of Eq (4), $C = \frac{1}{2}(\mathbf{V^0})^T\mathbf{A^{-1}}\mathbf{V^0}$, is positive. For a fixed direction of the dipole moment, C is a positive constant, and $\Delta G^{\mathrm{C-PCM}}$ is a negative quadratic function of the magnitude of the molecule dipole moment, μ. Therefore, a neutral molecule with a larger dipole moment benefits more from CPCM stabilization, meaning that the formation of localized solute electron density with charge separation is energetically favored in CPCM calculation. The same conclusion can be reached when the derivation is done for the continuous form of the CPCM cavity surface charge distribution $\sigma(\vec{r})$ (see supplementary material Text S1).

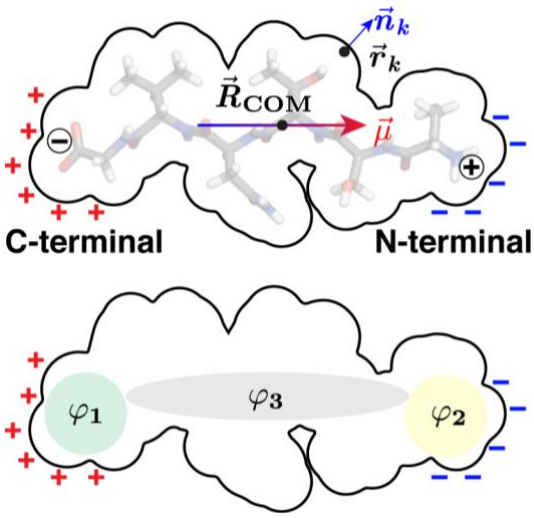

Figure 1. Illustration of CPCM cavity for a zwitterionic peptide (PDB ID: 3FTR). (upper) Presentation of the molecular dipole moment ($\vec{\mu}$), center of mass ($\vec{R}_{\mathrm{COM}}$), position ($\vec{r}_k$), and norm vector ($\vec{n}_k$) of the *k*th cavity tessera. Red plus signs around the C-terminal indicate positive surface polarization charges, and blue minus signs around the N-terminal indicate negative surface polarization charges. (lower) Examples of MOs that get stabilized (φ_1, green shade), destabilized (φ_2, yellow shade), or between the two extremes (φ_3, gray shade) in CPCM.

2. CPCM's selective stabilization of molecular orbitals

Here, we analyze CPCM's stabilization for individual MOs in DFT calculations. The 1-electron Kohn-Sham (KS) equation of a DFT calculation in CPCM can be expressed as

$$\hat{\mathrm{f}}^{\mathrm{CPCM}}\varphi_i = \epsilon_i\varphi_i, \tag{5}$$

where $\hat{\mathrm{f}}^{\mathrm{CPCM}} = \hat{\mathrm{f}}^{\mathrm{KS}} + \Delta\hat{\mathrm{f}}^{\mathrm{CPCM}}$ is the modified Kohn-Sham (KS) operator in CPCM, $\hat{\mathrm{f}}^{\mathrm{KS}}$ is the gas phase KS operator, and $\Delta\hat{\mathrm{f}}^{\mathrm{CPCM}}$ is the CPCM reaction field operator given by

$$\Delta\hat{\mathrm{f}}^{\mathrm{CPCM}} = -\sum_k^M \frac{q_k}{|\vec{r}-\vec{r}_k|} = -\frac{\varepsilon-1}{\varepsilon}\sum_k^M \frac{q_k^{\infty}}{|\vec{r}-\vec{r}_k|}. \tag{6}$$

Here, $\{q_k^{\infty}\}$ is the cavity surface polarization charge when the solvent is a conductor (dielectric constant $\varepsilon = \infty$). The orbital energy ϵ_i can be calculated as

$$\epsilon_i = \langle\varphi_i|\hat{\mathrm{f}}^{\mathrm{CPCM}}|\varphi_i\rangle = \langle i|\hat{\mathrm{f}}^{\mathrm{KS}}|i\rangle + \langle i|\Delta\hat{\mathrm{f}}^{\mathrm{CPCM}}|i\rangle = \epsilon_i^0 + \Delta\epsilon_i^{\mathrm{CPCM}}, \quad (7)$$

where the CPCM contribution to the orbital energy is

$$\Delta\epsilon_i^{\mathrm{CPCM}} = -\frac{\varepsilon - 1}{\varepsilon}\sum_k^M \int \frac{q_k^{\infty}}{|\vec{r} - \vec{r}_k|}|\varphi_i(\vec{r})|^2 d\vec{r}. \quad (8)$$

The sign of $\Delta\epsilon_i^{\mathrm{CPCM}}$ depends on the shape of the orbital $\varphi_i(\vec{r})$ and the distribution of the polarization charges $\{q_k^{\infty}\}$**.** Here, we qualitatively analyze the case for a large molecule with charge separation (Figure 1). Based on the CPCM formalism, the polarization charges $\{q_k^{\infty}\}$ have the opposite sign to the adjacent solute charge distribution. In a zwitterionic peptide, the C-terminal carries a negative partial charge, so the surrounding $\{q_k^{\infty}\}$ are positive. Hence, a MO localized around the C-terminal has $\Delta\epsilon_i^{\mathrm{CPCM}}$<0 and gets stabilized by CPCM (φ_1 in Figure 1). In contrast, a MO localized around the N-terminal gets destabilized ($\Delta\epsilon_i^{\mathrm{CPCM}}$>0) because of the surrounding negative $\{q_k^{\infty}\}$ (φ_2 in Figure 1). The CPCM stabilization of a delocalized MO is between these two extreme cases.

Furthermore, the change of orbital energy during dielectric tuning can be analyzed based on the derivative of orbital energy with respect to ε:

$$\left.\frac{\partial\Delta\epsilon_i^{\mathrm{CPCM}}}{\partial\varepsilon}\right|_{\{\varphi_j\}} = -\frac{1}{\varepsilon^2}\sum_k^M \int \frac{q_k^{\infty}}{|\vec{r} - \vec{r}_k|}|\varphi_i(\vec{r})|^2 d\vec{r} = \frac{1}{\varepsilon(\varepsilon - 1)}\Delta\epsilon_i^{\mathrm{CPCM}}. \quad (9)$$

Here the subscript $\{\varphi_j\}$ indicates that all MOs are assumed to be fixed when tuning ε. Then a MO like φ_1 in Figure 1 has $\Delta\epsilon_1^{\mathrm{CPCM}}$<0 [Eq. (7)] and $\left.\frac{\partial\Delta\epsilon_1^{\mathrm{CPCM}}}{\partial\varepsilon}\right|_{\{\varphi_j\}} < 0$ and will be more stabilized by CPCM as ε increases. In contrast, a MO like φ_2 in Figure 1 has $\Delta\epsilon_2^{\mathrm{CPCM}} > 0$ [Eq. (7)] and

$\left.\frac{\partial \Delta\epsilon_2^{\mathrm{CPCM}}}{\partial \varepsilon}\right|_{\{\varphi_i\}}$ >0 and will be destabilized even more as ε increases. Hence, the energy gap between these two types of orbitals, φ_1 and φ_2, will increase as ε increases.

III. COMPUTATIONAL DETAILS

A. Peptides and Small Zwitterions. To test CPCM's SCF improvement effects for large molecules, we selected a set of 25 canonical zwitterionic peptides from the RCSB PDB[23] database obtained with aqueous solution nuclear magnetic resonance (NMR) or X-ray diffraction. The peptides range from 6 to 26 in the number of residues and 64 to 440 in the number of atoms, and most were found to converge within fewer SCF iterations in CPCM and XP-PCM[24] than in the gas phase in previous studies.[1, 25] Detailed properties of the peptides (PDB ID, number of residues and atoms, charge, and secondary structure) are summarized in the supplementary material, Table S1. We also made the non-charge-separated form for each zwitterionic peptide to investigate how CPCM's SCF improvement effect depends on the molecule's charge distribution. The non-charge-separated structures were generated by transferring one proton from the N-terminal to the C-terminal with the remaining atom coordinates fixed using Avogadro 1.2.0.[26] To investigate the impacts of PCM on the electron density in detail, we selected four typical small zwitterions (glycine, β-Alanine, glycocyamine, and taurine) that can represent the charged residuals in peptides and are small enough to be calculated at the coupled-cluster level of theory. All the small zwitterion structures were downloaded from PubChem.[27]

B. DFT Calculations. All DFT calculations in the gas phase and CPCM were carried out in the GPU accelerated quantum chemistry package TeraChem.[28] The SCF convergence of peptides in the gas phase and CPCM were tested at the B3LYP[29]/6-31+G*[30-32] level, where the global hybrid functional, B3LYP, has its DDE partially mitigated by HF exchange but still benefits from CPCM's effects. The larger basis set, aug-cc-pVDZ,[33, 34] was used when comparing the DFT

electron density of small zwitterions with coupled-cluster reference results. The DIIS[35] algorithm was used for all SCF calculations and iterated until the largest component of the DIIS error vector was smaller than 1×10^{-6} Hartree. All electron integrals were evaluated on the GPUs using double precision arithmetic to isolate CPCM's impacts on SCF, avoiding potential influence by TeraChem's mixed precision or dynamic precision techniques.[36] This also guarantees that conclusions in this work also apply to calculations done with CPU-based quantum chemistry packages where only double-precision arithmetic is available.

For CPCM calculations, the solute cavity was built using defaults available in TeraChem (i.e., 1.2 times Bondi's van der Waals radii). The relative dielectric constant of the solvent, ε, was varied from 1 (gas-phase) to 80 (water), with values taken at 1, 2, 4, 10, and 80 for all the 25 peptides, because the CPCM field strength is proportional to $\frac{\varepsilon-1}{\varepsilon}$. To show the change of MO with ε, additional calculations at ε=1.2, 1.5, 3.0, and 6.0 were performed for several small peptides. For small zwitterions, calculations were performed at ε=1.0, 1.2, 1.5, 2.0, 3.0, 6.0, 10.0, 20.0, 40.0, and 80.0 to show a clear trend of the changes in molecule dipole moments and atomic charges.

The performance of PCM in improving SCF convergence was compared to the widely used level-shifting technique.[19] The level-shifting value for each peptide was set as the difference of HOMO-LUMO gaps calculated in aqueous solvation (ε=80) and the gas phase.

C. Wave Function Theory References. The gas-phase reference calculations for all small zwitterions were obtained with orbital-optimized coupled-cluster singles and doubles with perturbative triples (orb-opt CCSD(T))[37-39] theory using the aug-cc-pVTZ[33, 34] basis set with ORCA 5.0.2.[40] All small zwitterions studied in this work are singlets and were calculated in a restricted formalism.

D. Density Analysis. We employed Hirshfeld charges[41] to obtain a systematic estimate of the charge based on the real-space total density, simplifying the comparison of partial charges across distinct basis sets and electronic structure methods or software. To compare the electron density calculated with CPCM DFT at different dielectric constants relative to the gas-phase reference electron density calculated by orb-opt CCSD(T), we performed Hirshfeld population analysis[41] on all small zwitterions using Multiwfn,[42] which read molecular orbital information from molden files generated by TeraChem or converted from ORCA gbw files. Additional comparisons were done with other charge methods, including atomic dipole moment corrected Hirshfeld (ADCH),[43] Bader,[44] charge model 5 (CM5),[45] and Merz-Singh-Kollman (MK).[46] The Gaussian cube file's grid spacing was set to 0.04 Bohr to ensure convergence (supplementary material Table S2).

IV. RESULTS AND DISCUSSION

In the following subsections, we will investigate the applicable molecular systems and mechanisms for CPCM as an SCF convergence accelerator. We will first test the performance of CPCM in improving SCF convergence on zwitterionic and non-charge-separate peptides and amino acids. We will then compare CPCM with a conventional SCF convergence accelerator, level-shifting, and investigate the origin of their different performance despite their similar effects in opening the HOMO-LUMO gap. Finally, we will demonstrate the CPCM's reduction of DDE in the tested molecules while being used to accelerate SCF convergence.

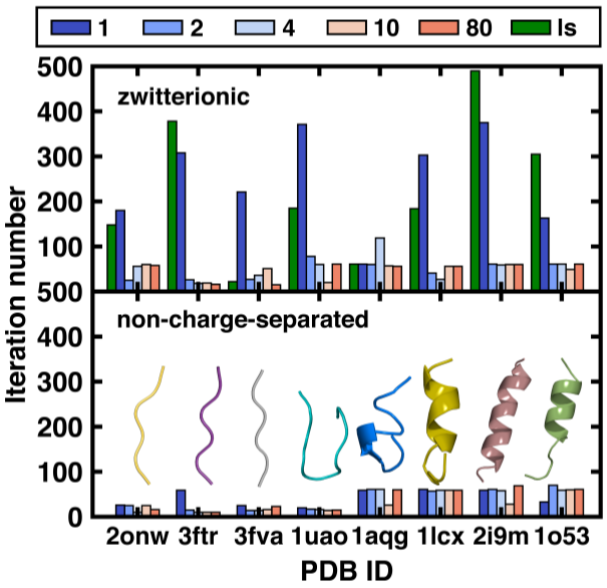

Figure 2. SCF iteration number for B3LYP/6-31+G* calculation in the gas phase (ε=1) and CPCM at different ε values (2, 4, 10, 80) for 8 out of 25 zwitterionic peptides (upper) and their non-charge-separated forms (lower). The level-shifting results for zwitterionic peptides are shown in green bars labeled with "ls". Peptide structures are shown in the inset.

A. The Impact of CPCM on SCF Convergence. We compared the SCF convergence of the B3LYP/6-31+G* calculations for 25 zwitterionic peptides and their non-charge-separated counterparts in the gas phase and CPCM (Figure 2 and supplementary material Table S3). It is known that CPCM's convergence improvement effect is most prominent for pure GGA functionals[9] like BLYP[47, 48] and PBE.[49] However, neither functional can generate converged gas-phase calculations for any of our benchmark molecules, making it hard to compare the SCF convergence quantitatively. Therefore, we chose the B3LYP functional, which had converged SCF but in an undesirably large number of iterations. In the gas phase (ε=1), the charge-separated peptides are always harder to converge than their non-charge-separated counterpart: all non-charge-separated peptides converge within 61 DIIS iterations while 17 out of 25 peptide zwitterions take over 100 iterations or do not converge within 1000 steps. Applying CPCM with

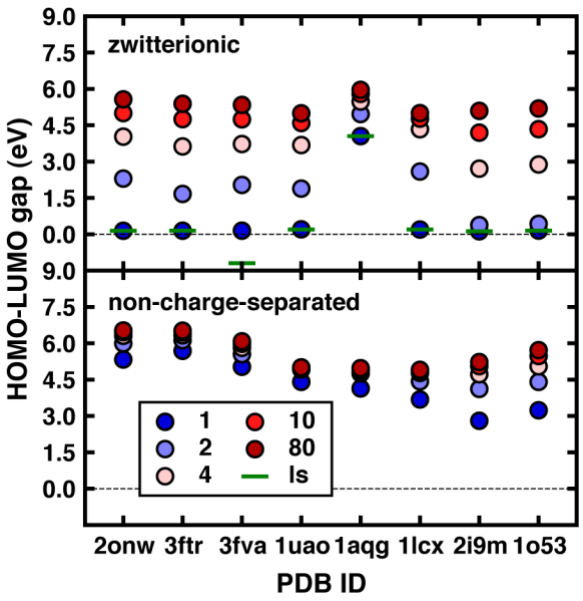

Figure 3. HOMO-LUMO gap (in eV) for 8 representative zwitterionic peptides (upper) and their non-charge-separated forms (lower) calculated with B3LYP/6-31+G*. Gas-phase (ε=1) and CPCM (ε=2, 4, 10, 80) calculations without level-shifting are denoted by circles. For zwitterionic peptides, gas phase calculations with level-shifting were denoted by green bars with the label "ls".

even a low dielectric constant of ε=2 (corresponding to non-polar solvents like cyclohexane) significantly reduces the SCF iterations for all zwitterionic peptides by at least 63% (Figure 2), with the only exception, 1aqg. When a high dielectric constant of ε=80 (corresponding to water) is applied, the DIIS iterations for 19 out of 25 peptides are reduced to <50% of the gas-phase results. In contrast, the SCF convergence of non-charge-separated peptides seems to benefit little from CPCM in the whole range of ε values tested (2 to 80), with only 3 out of 25 non-charge-separated peptides getting marginally reduced iterations. Our result demonstrates that CPCM can only effectively reduce the SCF iterations for zwitterionic peptides.

To understand the mechanism behind this, we investigated the HOMO-LUMO of these peptides as a function of ε values (Figure 3 and supplementary material Table S4). The gas phase

calculations of 21 out of 25 zwitterionic peptides have small HOMO-LUMO gaps less than 0.5 eV, agreeing with the vanishing HOMO-LUMO gap problem reported in previous studies of DFT convergence issues in proteins.[9] The HOMO-LUMO gap increases as ε increases, with an approximately linear relationship to (ε−1)/ε, consistent with our analysis in Section 2. The gap opening usually results in faster convergence than the gas-phase SCF. However, the DIIS iteration does not decrease monotonically as the gap (or ε) increases, remaining almost unchanged for ε values from 2 to 80. The increased gap is also observed for the non-charge separated peptides, but the increment is much smaller than that observed in zwitterions.

Previous research attributed the vanishing HOMO-LUMO gap to two main reasons: the intrinsic density delocalization error for DFAs[50] and the electric field caused by the surface dipole of

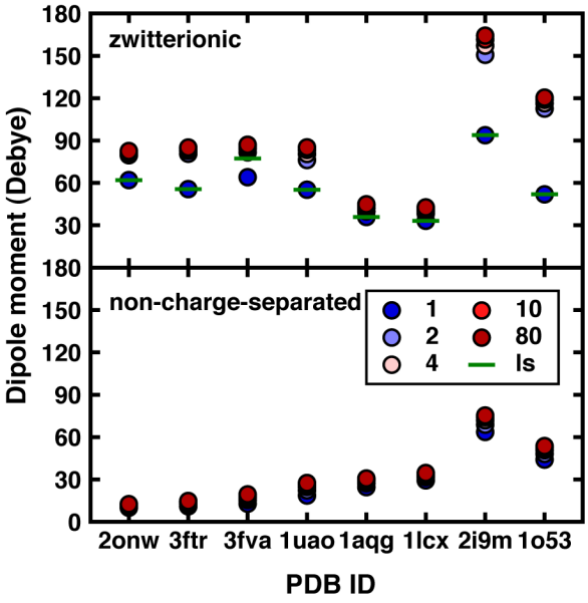

Figure 4. Dipole moment (in Debye) for 8 representative zwitterionic peptides (upper) and their non-charge-separated forms (lower) calculated by B3LYP/6-31+G*. Gas-phase (ε=1) and CPCM (ε=2, 4, 10, 80) calculations without level-shifting are denoted by circles. For zwitterionic peptides, gas phase calculations with level-shifting were denoted by green bars with the label "ls".

proteins.[51] CPCM could eliminate the second factor by inducing solvent polarization charges to screen this electric field. Therefore, the HOMO-LUMO gap can be gradually recovered when ε increases. In contrast, the non-charge-separated peptides' surface charges have much smaller magnitudes, as indicated by the relatively low dipole moments (Figure 4 and supplementary material Table S5). Hence, the non-charge-separated peptides do not suffer from the vanishing gap problem and benefit little from CPCM's SCF acceleration effect.

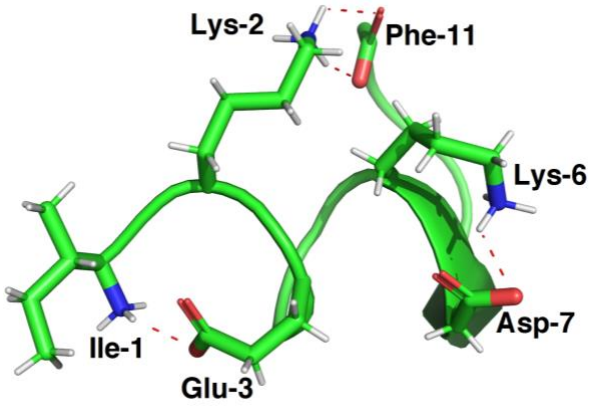

Figure 5. The structure of 1aqg peptide zwitterion. Only charged residues and terminals are specifically shown in sticks. The hydrogen bonds are shown in red dashed lines.

We also investigated why the exception, zwitterionic 1aqg, does not get SCF acceleration from CPCM. Compared to other peptide zwitterions, 1aqg has a significantly larger HOMO-LUMO gap of 4.049 eV and can converge easily in the gas phase within similar iterations as the non-charge-separated counterpart. As shown in Figure 5, further investigation of this peptide's structure shows strong salt bridges and ionic hydrogen bonds between oppositely charged residues. All three carboxylate anions are within 2.2 Å of an amino cation, electrostatically stabilizing each other and making this peptide resemble a non-charge-separated molecule. In contrast, other peptide zwitterions have at least a pair of oppositely charged residues separated by more than 6.0 Å.

B. Comparison to Level-Shifting. Although we have already observed the increased HOMO-LUMO gap and reduced SCF iterations in CPCM calculations, it is unclear whether CPCM's acceleration effects can be attributed solely to the increased gap. To deepen our understanding of the acceleration mechanism, we compared the performance of CPCM to level-shifting, a more conventional technique with a similar mechanism focusing on increasing the HOMO-LUMO gap. Level-shifting was initially introduced by Saunders and Hillier to address oscillatory SCF iterations caused by rapid mixing between the occupied (occ.) and virtual (virt.) space.[19] The algorithm shifts up the virtual-space diagonal elements of the Hamiltonian represented in the previous iteration's MO basis by a constant value λ (typically around 0.2 Hartree), and therefore limits vigorous occupied-virtual mixture upon diagonalization, making the SCF iterations more stable. Mathematically, level-shifting for a closed-shell molecule can be expressed as:

$$\begin{aligned} \mathbf{H}_{(k)}^{\mathrm{MO}} &= \mathbf{C}_{(k-1)}^{\dagger}\mathbf{H}_{(k)}^{\mathrm{AO}}\mathbf{C}_{(k-1)}, \\ H_{pp}^{\mathrm{MO,shifted}} &= \begin{cases} H_{pp}^{\mathrm{MO}}, & p \in \text{occ.} \\ H_{pp}^{\mathrm{MO}} + \lambda, & p \in \text{virt.} \end{cases} \end{aligned} \tag{10}$$

Here, $\mathbf{H}_{(k)}^{\mathrm{AO}}$ denotes the *k*-th iteration's Hamiltonian in AO basis, and $\mathbf{C}_{(k-1)}$ denotes the MO coefficients of the (*k*-1)-th iteration. The shift value, λ, is equivalent to an artificial increase of the HOMO-LUMO gap and is removed from the virtual orbital energies after each diagonalization. To make a fair comparison between CPCM and level-shifting, we set λ for each molecule to match the increased HOMO-LUMO gap caused by CPCM at ε=80:

$$\lambda = [\epsilon_{\mathrm{HOMO}}(\varepsilon = 80) - \epsilon_{\mathrm{LUMO}}(\varepsilon = 80)] - [\epsilon_{\mathrm{HOMO}}(\mathrm{gas}) - \epsilon_{\mathrm{LUMO}}(\mathrm{gas})] \tag{11}$$

In the situation where gas phase calculation is not converged, we assume that $\epsilon_{\mathrm{HOMO}}(\mathrm{gas}) - \epsilon_{\mathrm{LUMO}}(\mathrm{gas}) = 0$. If their SCF acceleration mechanisms are the same, they should give similar performance. The applied λ values range from 0.07 to 0.25 Hartree (1.9 to 6.8 eV) for the 25 tested

peptides. For gas-phase calculations that cannot converge within 1000 steps, we assume that the gas phase HOMO-LUMO gap is zero.

Contrary to our intuition, level-shifting has significantly worse performance than CPCM, with only 4 out of 25 peptides (1uao, 2ceh, 2jof, 3fva) showing non-trivial reduction (more than 50%) of DIIS iterations (Figure 2). Moreover, 3vfa and 2jof yield unphysical negative HOMO-LUMO gaps with level-shifting, and their energies are also 0.11 eV and 2.80 eV higher than their gas-phase result, indicating that level-shifting may cause the electronic structure converges to a different state (Figure 3). Therefore, CPCM performs better in some tricky molecules that cannot be handled by level-shifting. CPCM's acceleration mechanism cannot be explained in the same way as level-shifting by merely opening the HOMO-LUMO gap. This motivates us to investigate the different mechanisms in CPCM vs. level shifting in opening the gap in Section III C.

C. The Selective Stabilization of Kohn-Sham Orbitals by CPCM. Although both lead to the gap-opening effect, CPCM and level shifting modify orbital energies differently. Level-shifting applies the same shift to all virtual orbitals regardless of their local electrostatic environment (Eq (10)). In contrast, the change of orbital energy caused by CPCM, $\Delta\epsilon_i^{\mathrm{CPCM}}$, depends on the shape of the orbital $\varphi_i(\vec{r})$ and the distribution of the solvent polarization charges, as detailed in our theoretical analysis in Section II. To verify this, we investigated the change of HOMO and LUMO energies and orbital shapes as functions of CPCM dielectric constant, ε, for a representative zwitterionic peptide, 3ftr. As shown in Figure 6, The HOMO is mainly localized at the negatively charged C-terminal and therefore is expected to get electrostatically stabilized by surrounding positive CPCM polarization charges. In contrast, the LUMO is closer to the positively charged N-

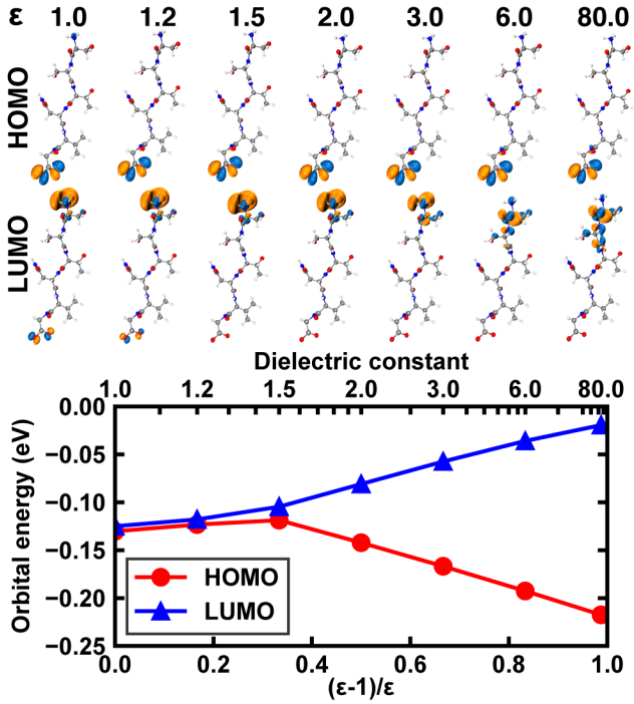

Figure 6. Frontier orbitals for 3ftr with different dielectric constants and their energies (in Hartree). (upper) Shape of the HOMO and LUMO at ε=1.0, 1.2, 1.5, 2.0, 3.0, 6.0, and 80.0. (lower) The HOMO (in red) and LUMO (in blue) energies plotted against $\frac{\varepsilon-1}{\varepsilon}$.

terminal and is expected to be destabilized by surrounding negative CPCM polarization charges. Indeed, $\epsilon_{\mathrm{HOMO}}^{\mathrm{CPCM}}$ decreases with ε, while $\epsilon_{\mathrm{LUMO}}^{\mathrm{CPCM}}$ increases with ε, leading to the enlarged HOMO-LUMO gap (Figure 6). The relation between the orbital energy, $\epsilon_i^{\mathrm{CPCM}}$, and (ε-1)/ε deviates slightly from the linear relationship shown by Eq (8) since orbital relaxation was not considered when deriving Eq. (8). Specifically, there is a significant change to the shape of 3ftr LUMO when ε increases (Figure 6). Due to the increasingly unfavorable electrostatic environment at the N-terminal at higher ε values, the LUMO gradually moves away from that region.

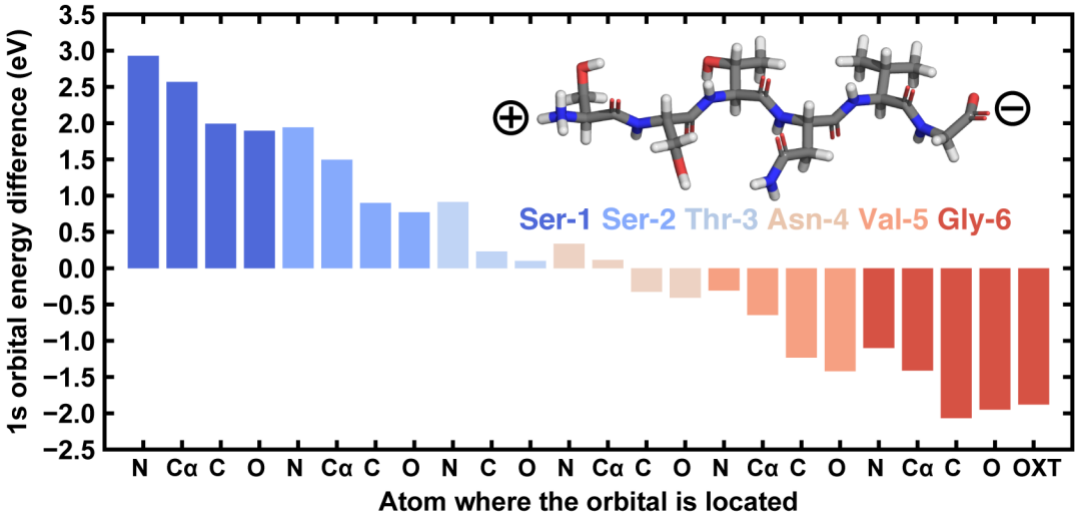

Figure 7. The energy change (in eV) of the 1s orbitals on 3FTR backbone atoms after adding CPCM with ε=80. The structure and sequence of 3FTR and positions of charged functional groups are shown in the inset. Bars are colored according to the amino acid label to which they belong. The 1s orbitals are marked with their corresponding atom names on the x-axis.

To further validate CPCM's selective stabilization of MOs based on their local electrostatic environments, we extended our analysis to more MOs in the zwitterionic 3ftr. We focused on core orbitals that were minimally impacted by orbital relaxation at different ε values and, hence, are expected to follow the trend of Eqs. (8) and (9). To further simplify the comparison, we limited our analysis to MOs with very similar shapes but distinct local electrostatic environments. Based

on these criteria, we selected the MOs corresponding to the 1s orbitals of the backbone atoms (N, Cα, C, O, N) of each residue, whose similarities and insensitivity to ε change were verified (supplementary material Text S2). Clear trends showed up as we plotted the orbital energy contribution by CPCM, $\Delta\epsilon_i^{\mathrm{CPCM}}$, ordered by the spatial location of these orbitals (Figure 7). The orbital energy change, $\Delta\epsilon_i^{\mathrm{CPCM}}$, gradually transits from negative (stabilization) to positive (destabilization) as the location of the 1s orbital changes from the C-terminal (GLY-6) to the N-terminal (SER-1), meeting the expectation of Eq. (8) and Eq. (9). The same trend was observed for another three small zwitterionic peptides, 2ol9, 2onw, and 3fva (supplementary material Figure S1).

We further compared CPCM and level-shifting by tracking the frontier orbital energy changes over SCF iterations. For the gas phase SCF of zwitterionic 3ftr without level-shifting, the HOMO and LUMO energies are constantly oscillating and very close to each other, with less than 0.03 eV difference in most iterations (Figure 8). Because of the small HOMO-LUMO gap, the LUMO switch between the MO localized on the N-terminal and the MO localized on the C-terminal. This near degeneracy of HOMO and LUMO also occasionally causes a vigorous mixture of the virtual and occupied spaces, leading to drastic character change for multiple frontier orbitals (e.g., iteration 51 of Figure 8), requiring many more iterations to reconverge. Level-shifting leads to more stable orbital energies in the short term but still suffers from occasional vigorous oscillations that cause the SCF to continue for many iterations. The reason is that the energetic order of frontier MOs in the first few iterations may significantly deviate from the converged result, but level-shifting reinforces this wrong order by separating the virtual and occupied space, which delays the convergence. For example, the converged LUMO (localized on the N-terminal) used to be HOMO-2 in iterations 27-38, and the level-shifting keeps it in the occupied space for a long time (Figure

8). In contrast, CPCM (ε=80) has the HOMO and LUMO well separated throughout all iterations and converges within 14 steps. CPCM's selective stabilization of MOs is consistent over different iterations because it recognizes each MO by its character rather than the current energy ordering (Figure 8).

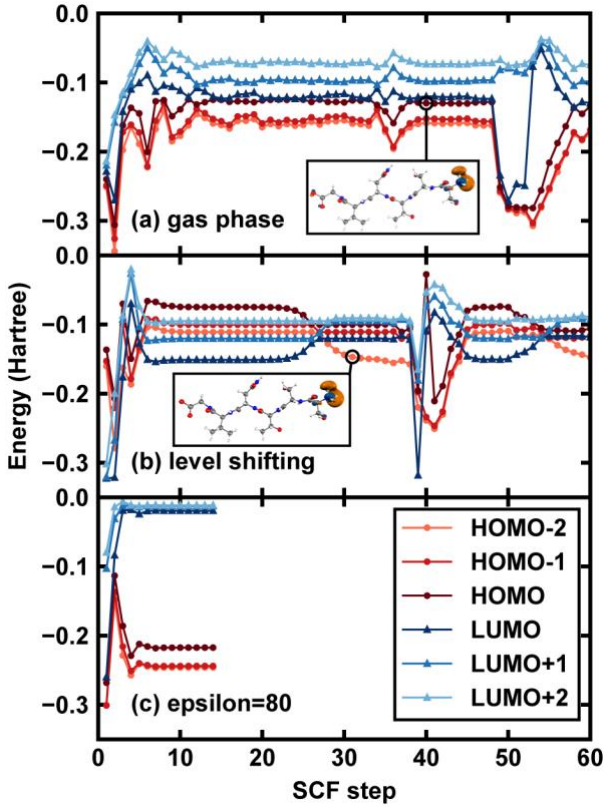

Figure 8. Frontier orbital energy for 3ftr during the SCF for single point energy calculation in (a) gas phase, (b) level shifting, and (c) ε=80. Orbital energies for the level-shifting result are already subtracted by the used level-shifting value. The gas-phase HOMO of iteration 40 and the level-shifting HOMO-2 of iteration 31 are shown in the insets of (a) and (b), respectively.

In summary, level-shifting artificially increases the energy for all virtual orbitals at the current iteration without considering orbital character. Therefore, it often keeps the wrong order of orbitals obtained in the first few SCF iterations, delaying the convergence. In contrast, CPCM has

consistent HOMO/LUMO definitions during SCF iterations, increasing the gap and preventing the SCF oscillation, therefore addressing the convergence problem.

D. Impacts of CPCM on electron density.

While CPCM could be useful for improving SCF convergence of large molecules, it also impacts the electronic structure. We selected four representative zwitterions (glycine, β-alanine, taurine, and glycocyamine) and computed their Hirshfeld partial charge on oppositely charged groups. For these zwitterions, the negatively charged terminal is the carboxyl anion (or sulfonic anion for taurine), while the positively charged terminal is the amino cation (or guanidino cation for glycocyamine). It is known that GGA functionals suffer from density delocalization errors (DDE),[8] which is reflected as the underestimated partial charges of gas-phase BLYP calculations compared

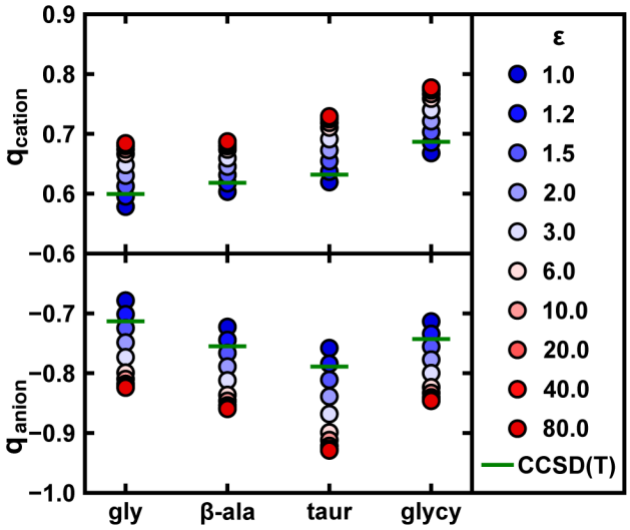

Figure 9. Partial charge (in e) on the cation N-terminal (upper) and the anion N-terminal (lower) for the four zwitterions for the CPCM B3LYP/aug-cc-pVDZ calculations with ε ranges from 1.0 to 80.0 (denoted by circles) and the orb-opt CCSD(T)/aug-cc-pVTZ reference (denoted by green bars). The abbreviations gly, β-ala, taur, and glycy on the x-axis correspond to glycine, β-alanine, taurine, and glycocyamine, respectively.

to the CCSD(T) reference for the four zwitterions (supplementary material Figure S2). Although adding the Hartree-Fock exchange can mitigate DDE, the needed percentage varies for different molecules and is often larger than the optimal percentage for accurate energy, e.g., 20% for B3LYP.[8] Hence, the partial charges by gas-phase B3LYP calculations are also underestimated (Figure 9). As CPCM is applied with increasing dielectric constant, ε, the magnitude of partial charges on both terminals monotonically increases, agreeing with CPCM's density localization effects as reported in the literature,[12, 52] as well as our analysis in Section II-A. For all four zwitterions, the cationic N-terminals and anionic C-terminals can get their B3LYP calculated partial charges recovered to the CCSD(T) reference values by an ε of about 1.2. The corresponding BLYP calculations need a slightly higher ε of 1.5 due to the lack of Hartree-Fock exchange mitigation (supplementary material Figure S2). It is worth noting that these optimal ε values to eliminate DDE are very small compared to even the most nonpolar solvents, e.g., ε=2.0 for cyclohexane at 298.15K.[53]

To further verify CPCM's density localization effects, we repeated the partial charge analysis for B3LYP and BLYP with another four charge models (see Section III D). All charge models agree on the trend, and the optimal ε for correcting DDE is always 1.2 and 1.5 for B3LYP and BLYP, respectively. The only exception is the Bader charges for cations, which are significantly lower than other charge models. The Bader charges for cationic C-terminals obtained with CCSD(T) are lower than that for gas-phase BLYP, meaning that further applying CPCM will worsen the results compared to CCSD(T). We believe that this exception is caused by the nature of the Bader charge itself, as some studies have shown that the Bader charge has very insufficient reproducibility for molecular properties such as dipole moment and electrostatic potential.[54]

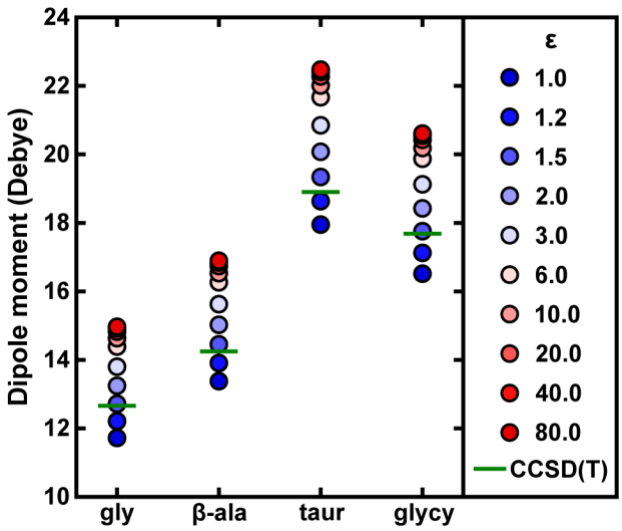

Figure 10. Dipole moments (in Debye) of four zwitterions calculated with CPCM B3LYP/aug-cc-pVDZ calculations at different ε (denoted by circles) and the gas phase orb-opt CCSD(T)/aug-cc-pVTZ reference (denoted by green bars). The abbreviations gly, β-ala, taur, and glycy on the x-axis correspond to glycine, β-alanine, taurine, and glycocyamine, respectively.

We also investigated these zwitterions' dipole moments, which indicate the polarization of charge distribution (Figure 10). Consistent with the partial charge results, the dipole moment increases with ε. The optimal ε for recovering the dipole moment values to the CCSD(T) reference for different zwitterions is always between 1.2 and 1.5 for B3LYP calculations, whereas the optimal ε for BLYP is larger, ranging from 1.2 to 6.0 for different molecules (supplementary material Figure S5). In summary, the SCF tends to converge to a more localized electron density distribution when CPCM is applied, increasing the partial charges and dipole moments. This localization effect could recover some gas-phase density properties affected by DDE when a very small dielectric constant is applied.

V. CONCLUSIONS

This work seeks to understand the mechanisms for CPCM's SCF improvement effects previously observed in the DFT calculations of proteins and identify the applicable type of molecules to be accelerated by CPCM.

We investigated the DFT convergence of 25 peptides in their zwitterionic and non-charge-separated form and concluded that CPCM could only effectively reduce the SCF iteration steps for charge-separate systems. The reason is that CPCM can open the near vanishing HOMO-LUMO gap in the zwitterionic peptide and avoid the vigorous virtual-occupied space mixing over iterations, the origin of convergence issues. Non-charge-separated molecules do not suffer from the vanishing gap problem and, therefore, their convergence does not benefit from CPCM even though their HOMO-LUMO gaps are also slightly increased by CPCM.

Further comparison with the level-shifting technique revealed that CPCM and level-shifting have different mechanisms for opening the HOMO-LUMO gap. Level-shifting artificially lifts the energies of all virtual orbitals by a constant value regardless of the shape of the MOs, which may mistakenly keep some MOs in the virtual space for many iterations because of the wrong energetic ordering of MOs at the early stage of SCF. This issue is not likely for CPCM, whose selective stabilization or destabilization of specific MOs is based on the MO's local electrostatic environment and is consistent over iterations. Because of this, CPCM performs better than level-shifting in reducing SCF iterations for charge-separated molecules.

Finally, we emphasized that CPCM can significantly impact the converged electronic structure while being used as an SCF accelerator. CPCM tends to localize the electron density, increasing the extent of charge separation and the dipole moment magnitude. A very low dielectric constant of 1.2 (or 1.5) can mitigate DDE and recover density-driven molecular properties to the orb-opt CCSD(T) reference results. However, these values are lower than the typical epsilon of real

solvents. Using CPCM at larger epsilon values to improve SCF convergence is likely to over-localize the electron density if the purpose is to study the molecule in the gas phase.

These conclusions provide guidelines about when CPCM is useful for fixing convergence issues in large-scale DFT calculations, especially when traditional SCF convergence techniques fail. The epsilon value can now be selected more transparently: it should fall in the range efficient for improving convergence and appropriate for DDE mitigation. Further investigations are needed to determine the effect of CPCM on the DFT convergence of other systems, such as transition metal-containing molecules. In addition, it is also possible to develop new SCF convergence techniques based on the mechanism of CPCM's convergence acceleration effects.

SUPPLEMENTARY MATERIAL

See the supplementary material for the proof of Eq. (4) in the continuous form of CPCM, the characteristics of the 25 tested proteins, Bader charge convergence tests, the SCF iteration number, HOMO-LUMO gap, and dipole moment for all the 25 peptides, the criteria for selecting comparable 1s orbitals, energy variation of 1s orbitals for the other three peptides, the partial charge and dipole moment for the 4 small zwitterions calculated by BLYP, and the comparison of charge calculation methods using the BLYP and B3LYP result for the 4 zwitterions (PDF).

The structures of all tested 25 zwitterionic proteins and their non-charge-separated counterparts, and the structure of the 4 zwitterions (ZIP).

ACKNOWLEDGMENTS

Acknowledgment is made to the Donors of the American Chemical Society Petroleum Research Fund for support of this research through grant number PRF #65858-DNI4. This work used the Extreme Science and Engineering Discovery Environment[55] (XSEDE) Bridges-2 at Pittsburgh

Supercomputing Center through allocation CHE210036, which is supported by National Science Foundation grant number ACI-1548562.

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material.

AUTHOR DECLARATIONS

The authors have no conflicts to disclose.

REFERENCES

[1] H. J. Kulik, N. Luehr, I. S. Ufimtsev, and T. J. Martinez, "Ab Initio Quantum Chemistry for Protein Structures", The Journal of Physical Chemistry B **116**, 12501 (2012).
[2] J. Antony, and S. Grimme, "Fully ab initio protein-ligand interaction energies with dispersion corrected density functional theory", Journal of Computational Chemistry **33**, 1730 (2012).
[3] I. S. Ufimtsev, N. Luehr, and T. J. Martinez, "Charge Transfer and Polarization in Solvated Proteins from Ab Initio Molecular Dynamics", The Journal of Physical Chemistry Letters **2**, 1789 (2011).
[4] D. J. Tozer, and F. De Proft, "Computation of the hardness and the problem of negative electron affinities in density functional theory", The Journal of Physical Chemistry A **109**, 8923 (2005).
[5] P. Mori-Sánchez, A. J. Cohen, and W. Yang, "Localization and delocalization errors in density functional theory and implications for band-gap prediction", Physical review letters **100**, 146401 (2008).
[6] B. G. Johnson, C. A. Gonzales, P. M. Gill, and J. A. Pople, "A density functional study of the simplest hydrogen abstraction reaction. Effect of self-interaction correction", Chemical physics letters **221**, 100 (1994).
[7] A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov, and G. E. Scuseria, "Density functionals that are one-and two-are not always many-electron self-interaction-free, as shown for H 2+, He 2+, Li H+, and Ne 2+", The Journal of chemical physics **126**, 104102 (2007).
[8] F. Liu, and H. J. Kulik, "Impact of Approximate DFT Density Delocalization Error on Potential Energy Surfaces in Transition Metal Chemistry", Journal of Chemical Theory and Computation **16**, 264 (2020).
[9] E. Rudberg, "Difficulties in applying pure Kohn–Sham density functional theory electronic structure methods to protein molecules", Journal of Physics: Condensed Matter **24**, 072202 (2012).

[10] Z. Zheng, J.-L. Brédas, and V. Coropceanu, “Description of the Charge Transfer States at the Pentacene/C60 Interface: Combining Range-Separated Hybrid Functionals with the Polarizable Continuum Model”, The Journal of Physical Chemistry Letters **7**, 2616 (2016).
[11] T. B. de Queiroz, and S. Kümmel, “Charge-transfer excitations in low-gap systems under the influence of solvation and conformational disorder: Exploring range-separation tuning”, The Journal of Chemical Physics **141**, 084303 (2014).
[12] L. O. Hemmingsen, O. A. J. Hervir, and S. G. Dale, “Linear fractional charge behavior in density functional theory through dielectric tuning of conductor-like polarizable continuum model”, The Journal of Chemical Physics **156**, 014106 (2022).
[13] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. Humphreys, and A. P. Sutton, “Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+ U study”, Physical Review B **57**, 1505 (1998).
[14] A. D. Becke, “A new mixing of Hartree–Fock and local density‐functional theories”, The Journal of chemical physics **98**, 1372 (1993).
[15] A. V. Krukau, G. E. Scuseria, J. P. Perdew, and A. Savin, “Hybrid functionals with local range separation”, The Journal of chemical physics **129**, 124103 (2008).
[16] A. Bajaj, J. P. Janet, and H. J. Kulik, “Communication: Recovering the flat-plane condition in electronic structure theory at semi-local DFT cost”, The Journal of Chemical Physics **147**, 191101 (2017).
[17] D. M. York, and M. Karplus, “A Smooth Solvation Potential Based on the Conductor-Like Screening Model”, The Journal of Physical Chemistry A **103**, 11060 (1999).
[18] A. W. Lange, and J. M. Herbert, “PCM-iswig”, The Journal of Chemical Physics **133**, 244111 (2010).
[19] V. R. Saunders, and I. H. Hillier, “Level Shifting”, International Journal of Quantum Chemistry **7**, 699 (1973).
[20] A. W. Lange, and J. M. Herbert, “A smooth, nonsingular, and faithful discretization scheme for polarizable continuum models: the switching/Gaussian approach”, The Journal of chemical physics **133**, 244111 (2010).
[21] B. Mennucci, “Polarizable continuum model”, Wiley Interdisciplinary Reviews: Computational Molecular Science **2**, 386 (2012).
[22] G. Scalmani, and M. J. Frisch, “Continuous surface charge polarizable continuum models of solvation. I. General formalism”, The Journal of Chemical Physics **132**, 114110 (2010).
[23] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, “The Protein Data Bank”, Nucleic Acids Research **28**, 235 (2000).
[24] R. Cammi, “A new extension of the polarizable continuum model: Toward a quantum chemical description of chemical reactions at extreme high pressure”, Journal of Computational Chemistry **36**, 2246 (2015).
[25] A. Gale, E. Hruska, and F. Liu, “Quantum chemistry for molecules at extreme pressure on graphical processing units: Implementation of extreme-pressure polarizable continuum model”, The Journal of Chemical Physics **154**, 244103 (2021).
[26] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, “Avogadro: an advanced semantic chemical editor, visualization, and analysis platform”, Journal of Cheminformatics **4**, 17 (2012).
[27] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, “PubChem in 2021: new data content and improved web interfaces”, Nucleic Acids Research **49**, D1388 (2021).

[28] S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, C. M. Isborn, S. I. L. Kokkila‐Schumacher, X. Li, F. Liu, N. Luehr, J. W. Snyder, C. Song, A. V. Titov, I. S. Ufimtsev, L. P. Wang, and T. J. Martínez, "TeraChem : A graphical processing unit ‐accelerated electronic structure package for large‐scale ab initio molecular dynamics", WIREs Computational Molecular Science **11** (2021).
[29] A. D. Becke, "Density‐functional thermochemistry. III. The role of exact exchange", The Journal of Chemical Physics **98**, 5648 (1993).
[30] W. J. Hehre, R. Ditchfield, and J. A. Pople, "Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules", J. Chem. Phys. **56**, 2257 (1972).
[31] P. C. Hariharan, and J. A. Pople, "The influence of polarization functions on molecular orbital hydrogenation energies", Theor. Chim. Acta **28**, 213 (1973).
[32] T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. V. R. Schleyer, "Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F", Journal of Computational Chemistry **4**, 294 (1983).
[33] T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen", J. Chem. Phys. **90**, 1007 (1989).
[34] R. A. Kendall, T. H. Dunning, and R. J. Harrison, "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions", J. Chem. Phys. **96**, 6796 (1992).
[35] P. Pulay, "Convergence acceleration of iterative sequences. the case of scf iteration", Chemical Physics Letters **73**, 393 (1980).
[36] N. Luehr, I. S. Ufimtsev, and T. J. Martínez, "Dynamic Precision for Electron Repulsion Integral Evaluation on Graphical Processing Units (GPUs)", Journal of Chemical Theory and Computation **7**, 949 (2011).
[37] G. D. Purvis, and R. J. Bartlett, "A full coupled‐cluster singles and doubles model: The inclusion of disconnected triples", The Journal of Chemical Physics **76**, 1910 (1982).
[38] G. E. Scuseria, and H. F. Schaefer, "The optimization of molecular orbitals for coupled cluster wavefunctions", Chemical Physics Letters **142**, 354 (1987).
[39] C. D. Sherrill, A. I. Krylov, E. F. C. Byrd, and M. Head-Gordon, "Energies and analytic gradients for a coupled-cluster doubles model using variational Brueckner orbitals: Application to symmetry breaking in O4+", The Journal of Chemical Physics **109**, 4171 (1998).
[40] F. Neese, "Software update: The ORCA program system—Version 5.0", WIREs Computational Molecular Science 10.1002/wcms.1606 (2022).
[41] F. L. Hirshfeld, "charge-Hirshfeld", Theoretica Chimica Acta **44**, 129 (1977).
[42] T. Lu, and F. Chen, "Multiwfn: A multifunctional wavefunction analyzer", Journal of Computational Chemistry **33**, 580 (2012).
[43] T. Lu, and F. Chen, "charge-ADCH", Journal of Theoretical and Computational Chemistry **11**, 163 (2012).
[44] R. F. W. Bader, "charge-Bader", Chemical Reviews **91**, 893 (1991).
[45] A. V. Marenich, S. V. Jerome, C. J. Cramer, and D. G. Truhlar, "Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases", Journal of Chemical Theory and Computation **8**, 527 (2012).
[46] B. H. Besler, K. M. Merz, and P. A. Kollman, "charge-MK", Journal of Computational Chemistry **11**, 431 (1990).

[47] A. D. Becke, “Density-functional exchange-energy approximation with correct asymptotic behavior”, Physical Review A **38**, 3098 (1988).
[48] C. Lee, W. Yang, and R. G. Parr, “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density”, Physical Review B **37**, 785 (1988).
[49] J. P. Perdew, K. Burke, and M. Ernzerhof, “Generalized Gradient Approximation Made Simple”, Physical Review Letters **77**, 3865 (1996).
[50] J. Antony, and S. Grimme, “Fully ab initio protein‐ligand interaction energies with dispersion corrected density functional theory”, Journal of computational chemistry **33**, 1730 (2012).
[51] G. Lever, D. J. Cole, N. D. M. Hine, P. D. Haynes, and M. C. Payne, “Electrostatic considerations affecting the calculated HOMO–LUMO gap in protein molecules”, Journal of Physics: Condensed Matter **25**, 152101 (2013).
[52] S. G. Dale, and E. R. Johnson, “Counterintuitive electron localisation from density-functional theory with polarisable solvent models”, The Journal of Chemical Physics **143**, 184112 (2015).
[53] H. T. French, M. Koshla, and K. N. Marsh, “Dielectric constants and apparent dipole moments of (butan-1-ol or butan-2-ol + cyclohexane) at 298.15 and 318.15 K and of (2-methylpropan-2-ol + cyclohexane) at 299.15 and 318.15 K”, The Journal of Chemical Thermodynamics **20**, 1175 (1988).
[54] T. Lu, and F. Chen, “Comparison of Computational Methods for Atomic Charges”, Acta Physico-Chimica Sinica **28**, 1 (2012).
[55] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, and N. Wilkins-Diehr, “XSEDE: Accelerating Scientific Discovery”, Comput. Sci. Eng. **16**, 62 (2014).