SIAM J. SCI. COMPUT.
Vol. 43, No. 1, pp. A415–A445

REGULARIZATION OF POISSON–BOLTZMANN TYPE EQUATIONS WITH SINGULAR SOURCE TERMS USING THE RANGE-SEPARATED TENSOR FORMAT*

PETER BENNER†, VENERA KHOROMSKAIA‡, BORIS KHOROMSKIJ§, CLEOPHAS KWEYU¶, AND MATTHIAS STEIN†

Abstract. In this paper, we present a new regularization scheme for the linearized Poisson–Boltzmann equation (PBE) which models the electrostatic potential of biomolecules in a solvent. This scheme is based on the splitting of the target potential into the short- and long-range components localized in the molecular region by using the range-separated (RS) tensor format [P. Benner, V. Khoromskaia, and B. N. Khoromskij, *SIAM J. Comput.*, 2 (2018), pp. A1034–A1062] for representation of the discretized multiparticle Dirac delta [B. N. Khoromskij, *J. Comput. Phys.*, 401 (2020), 108998] constituting the highly singular right-hand side in the PBE. From the computational point of view our regularization approach requires only the modification of the right-hand side in the PBE so that it can be implemented within any open-source grid-based software package for solving PBE that already includes some FEM/FDM disretization scheme for elliptic PDE and solver for the arising linear system of equations. The main computational benefits are twofold. First, one applies the chosen PBE solver only for the smooth long-range (regularized) part of the collective potential with the regular right-hand side represented by a low-rank RS tensor with a controllable precision. Thus, we eliminate the numerical treatment of the singularities in the right-hand side and do not change the interface and boundary conditions. And second, the elliptic PDE need not be solved for the singular part in the right-hand side at all, since the short-range part of the target potential of the biomolecule is precomputed independently on a computational grid by simple one-dimensional tensor operations. The total potential is then obtained by adding the numerical solution of the PBE for the smooth long-range part to the directly precomputed tensor representation for the short-range contribution. Numerical tests illustrate that the new regularization scheme, implemented by a simple modification of the right-hand side in the chosen PBE solver, improves the accuracy of the approximate solution on rather coarse grids. The scheme also demonstrates good convergence behavior on a sequence of refined grids.

Key words. Poisson–Boltzmann equation, Coulomb potential, summation of electrostatic potentials, long-range many-particle interactions, low-rank tensor decompositions, range-separated tensor formats

AMS subject classifications. 65F30, 65F50, 65N35, 65F10

DOI. 10.1137/19M1281435

1. Introduction.

Numerical treatment of long-range interaction potentials is a challenging task in computer modeling of multiparticle systems, for example, in

*Submitted to the journal's Methods and Algorithms for Scientific Computing section August 15, 2019; accepted for publication (in revised form) December 7, 2020; published electronically January 27, 2021.

https://doi.org/10.1137/19M1281435

Funding: The work of the fourth author was supported by the International Max Planck Research School (IMPRS) for Advanced Methods in Process and Systems Engineering and the Max Planck Society for the Advancement of Science (MPG).

†Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany (benner@mpi-magdeburg.mpg.de, matthias.stein@mpi-magdeburg.mpg.de).

‡Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany, and Max Planck Institute for Mathematics in the Sciences, Inselstr. 22-26, D-04103 Leipzig, Germany (vekh@mis.mpg.de).

§Max Planck Institute for Mathematics in the Sciences, Inselstr. 22-26, D-04103 Leipzig, Germany (bokh@mis.mpg.de).

¶Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany, and Moi University, Department of Mathematics and Physics, P.O. Box 3900-30100, Eldoret, Kenya (kweyu@mpi-magdeburg.mpg.de).

calculation of electrostatics in large solvated biological systems, or in many-particle dynamics simulations [15, 50, 17, 28, 53, 43]. The Poisson–Boltzmann equation (PBE) [25, 51, 21, 13] is one of the most popular implicit solvent models for computation of the electrostatic potential in proteins [37, 29, 55, 1, 45, 23, 44]. Other models include the generalized Born methods [4] and the polarizable continuum models [3]. The PBE computes the electrostatic potential both in the protein and in the surrounding solvent, and it is widely used in protein docking, in classification problems, and for computation of the free energy of biomolecules in a self-consistent way [18, 27].

The main difficulty in the traditional finite element method (FEM) approximation schemes for the three-dimensional (3D) PBE problem is related to the presence of a highly singular source term that includes a large sum of Dirac delta distributions which need to be resolved using rather coarse 3D grids. To overcome these limitations, a number of regularization schemes for the FEM applied to the PBE, based on the full grid representation of all functional data, have been considered in the literature; see [24, 57] and references therein. Consequently, we note that the PBE theory has recently received major improvement in terms of accuracy by the introduction of solution decomposition techniques which have been developed, for example, in [57, 47, 12], where the PBE is treated as an interface problem. This aims at avoiding the discontinuities in traces and fluxes at the interface between the biomolecule and the solvent and also to circumvent constructing the numerical approximations corresponding to the Dirac delta distributions because of the existence of analytical expansions in the solute subregion.

However, these techniques still face the following computational challenges. First, jumps in the interface conditions, arising due to regularizing splitting of the solution, need to be incorporated to eliminate the solution discontinuity (e.g., Cauchy data) at the interface. Second, the boundary conditions have to be specified using some analytical representation of the solution of the PBE. And third, in regularization-based techniques [57], one has to solve multiple algebraic systems for the linear (or nonlinear) boundary value problems before summing up the partial solutions, which increases the computational costs. We provide an overview of the regularization techniques in Appendix A.

Here, we present a new approach for the regularization of the PBE by using the range-separated (RS) canonical tensor format introduced and analyzed in [7]. The RS tensor format relies on the independent grid-based low-rank tensor representation of the long- and short-range parts in the total sum of single-particle electrostatic potentials discretized on a fine 3D $n \times n \times n$ Cartesian grid Ω_n in the computational box $\Omega \subset \mathbb{R}^3$. This representation is based on the splitting of a single reference potential, defined by a radial function like $p(\|\bar{x}\|) = 1/\|\bar{x}\|$, into a sum of localized and long-range low-rank canonical tensors both represented on the computational grid Ω_n. This RS splitting is justified by the previous approximation results for the low-rank decomposition of function related tensors [22, 19, 33, 32]. The long-range part in the collective potential of a many-particle system is represented as a low-rank canonical tensor with the rank only logarithmically depending on the number of particles in the system. The rank-reduction algorithm is performed by rank-structured tensor operations including the reduced higher order singular value decomposition (RHOSVD) [36]. The short-range contributions to a many-particle potential are parametrized by a single low-rank canonical tensor of local support. Notice that in [6], it was already sketched how the RS tensor formats can be utilized for calculation of the free interaction energy of protein-type systems, while the idea

for regularized formulation of the PBE by using the smooth long-range part of the free space electrostatic potential was outlined in [35].

In this paper, we introduce the new regularization scheme for the solution of the linearized PBE adapting the RS tensor format. It is based on a localized splitting scheme for the highly singular solution and right-hand side in the PBE, by using the RS tensor decomposition of the discretized multicentered Dirac delta introduced in [35]. This approach requires only a simple modification (regularization) in the right-hand side of the PBE in the solute region, but it does not change the interface conditions and, hence, the FEM system matrix. The most singular component in the potential is recovered explicitly as the short-range part in the RS tensor splitting of the free space potential on the grid. The computational advantages are due to the localization of the modified right-hand side within the molecular region and automatic maintaining of the continuity in the Cauchy data on the interface. The main computational benefits are the following:

- The new computational scheme only requires solving a single system of the FDM/FEM equations for the smooth long-range (regularized) part u^r of the collective potential, with the modified right-hand side that is represented by a low-rank RS-tensor with controllable precision (see sections 3.2 and 4.2 for details).
- The "singular" short-range part of the solution is computed directly (without solving the differential equation) by 1D tensor operations, and the approximation error is controlled by the rank parameter.
- The total target potential is then obtained by summing up the PBE solution for the regularized long-range part u^r and the precomputed tensor representation for the short-range contribution.
- The approach needs only a modification of the right-hand side (but not the interface and boundary conditions) and thus any favorable PBE software based on FEM/FDM discretization methods (construction of the stiffness matrix and error analysis) and the respective solver for the linear system of equations can be applied to calculate the regularized part u^r. In numerical tests, we compute the regularized solution u^r by using the adaptive Poisson–Boltzmann software (APBS) software package (version 1.5-linux64), employing the multigrid (PMG) accelerated FDM [24, 2, 25], with the modified right-hand side as input data.
- The electrostatic interaction energy can be calculated directly from the computed long-range potential u^r at each atomic position. This is a property which is hard to achieve by the commonly used regularization schemes because of their inability to decouple the short- and long-range components of the potential only in the molecular region.

As numerical illustrations, we compute the free space electrostatic potentials of biomolecules using the RS tensor format in the framework of our splitting scheme, and compare them with the solutions calculated by traditional FEM/FDM methods for the PBE, i.e., the APBS software package [24, 2], as well as the MATLAB program for biomolecular electrostatic calculations (MPBEC) [54]. An example for the linearized PBE is included.

The rest of the paper is organized as follows. In section 2, we formulate the linearized PBE and outline the computational difficulties for numerical treatment of this elliptic equation in 3D due to the highly singular input data. Section 3 describes the principles of the rank-structured tensor approximation to the long-range electrostatic potential and sketches the RS tensor decomposition techniques [7] for the free

space electrostatic potential of many particle systems. In section 4, we explain how the application of the RS tensor format leads to the new regularization scheme for solving the PBE. Section 5 presents the numerical tests illustrating the benefits of the proposed method, and comparisons with the solutions obtained by a standard FDM-based PBE solver are provided. Finally, Appendix A provides a short overview of existing solution decomposition schemes for the PBE problem.

2. Problem setting for the linearized PBE. We describe the presented new RS splitting scheme for solutions of elliptic PDEs with jumping coefficients in the presence of singular source terms that models the electrostatic potential in biomolecular systems (the PBE). In this section we discuss the corresponding problem setting.

We consider a solvated biomolecular system modeled by dielectrically separated domains with singular Coulomb potentials distributed in the molecular region. For schematic representation, we consider the system occupying a rectangular domain Ω with boundary $\partial\Omega$ (see Figure 1), where the solute (molecule) region is denoted by Ω_m and the solvent region by Ω_s such that

$$\overline{\Omega} = \overline{\Omega}_m \cup \overline{\Omega}_s.$$

In this case, the linearized (dimensionless) PBE takes the form (see [46])

$$-\nabla \cdot (\epsilon \nabla u) + \bar{\kappa}^2 u = \rho_f := \sum_{k=1}^{N} z_k \delta(\|\bar{x} - \bar{x}_k\|) \quad \text{in} \quad \Omega, \tag{2.1}$$

where ρ_f is the scaled singular charge distribution supported at points $\bar{x}_k$ in Ω_m, δ is the Dirac delta distribution, and $z_k \in \mathbb{R}$ denotes the charge located at the atomic center $\bar{x}_k$, while u denotes the electrostatic potential generated by these charges. Given the dielectric constants $\epsilon_m, \epsilon_s > 0$, the piecewise constant coefficients $\epsilon = \epsilon(\bar{x})$ and $\bar{\kappa} = \bar{\kappa}(\bar{x})$ are defined by

$$\epsilon(\bar{x}) = \begin{cases} \epsilon_m & \text{if } \bar{x} \in \Omega_m, \\ \epsilon_s & \text{if } \bar{x} \in \Omega_s, \end{cases} \qquad \bar{\kappa}(\bar{x}) = \begin{cases} 0 & \text{if } \bar{x} \in \Omega_m, \\ \bar{\kappa} & \text{if } \bar{x} \in \Omega_s. \end{cases} \tag{2.2}$$

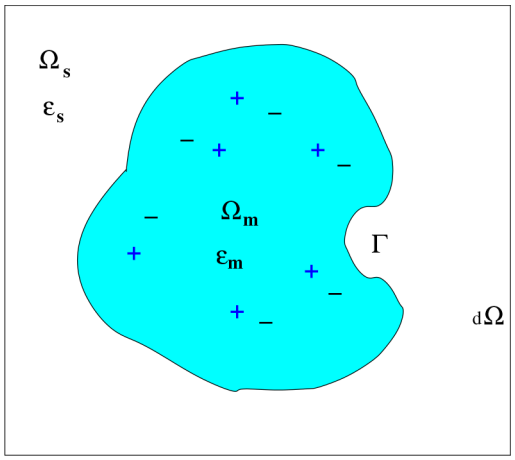

FIG. 1. *Solute and solvent regions in the computational domain for the PBE.*

Here $\epsilon = \epsilon_m = O(1) > 0$ and $\bar{\kappa} = 0$ in Ω_m, while in the solvent region, Ω_s, we have for the constant $\bar{\kappa} \geq 0$ and $\epsilon = \epsilon_s \geq \epsilon_m$ (in some cases the ratio ϵ_s/ϵ_m could be about several tens).

The interface conditions on the interior boundary $\Gamma = \partial\Omega_m$ arise from the dielectric theory (continuity of potentials and fluxes on the interface):

$$[u] = 0, \quad \left[\epsilon \frac{\partial u}{\partial n}\right] = 0 \quad \text{on} \quad \Gamma, \tag{2.3}$$

where $[f]_\Gamma = \lim_{t \longrightarrow 0} (f(\bar{x} + tn_\Gamma) - f(\bar{x} - tn_\Gamma))$ denotes the jump of f across the boundary. Here, n_Γ means the unit outward normal direction of the interface Γ.

The boundary conditions on the external boundary $\partial\Omega$, approximating the asymptotic at infinity, $|u(\bar{x})| \to \infty$ as $\|\bar{x}\| \to \infty$, can be specified depending on the particular problem setting. The simple homogeneous Dirichlet boundary conditions

$$u_{|\partial\Omega} = 0$$

can be utilized. In our numerical tests based on traditional FDM discretization in a bounded domain, we apply the more accurate inhomogeneous Dirichlet boundary conditions taking the form

$$u(\bar{x})_{|\partial\Omega} = \frac{\epsilon_m}{4\pi\epsilon_s} \sum_{k=1}^{N} \frac{z_k e^{-\bar{\kappa}(\|\bar{x}-\bar{x}_k\|-R_k)}}{\|\bar{x}-\bar{x}_k\|(1+\bar{\kappa}R_k)}, \quad \bar{x} \in \partial\Omega, \tag{2.4}$$

where the constants $R_k \geq 0$ denote the atomic radii. This condition is often used in the literature for the practical solution of the PBE via FEM/FDM methods applied on a bounded domain (also in case $\bar{\kappa} = 0$). Such an approximation can be motivated by the fact that the function in (2.4) represents the dominating term in the exact solution of the PBE in an unbounded domain, in the case when the molecular region Ω_m is the union of balls of radii R_k centered at $\bar{x}_k$; see details in [30]. This condition is exact in the case of free space potential $\epsilon_s = \epsilon_m$ and $\bar{\kappa} = 0$.

The source term in the right-hand side of the PBE (2.1) is strongly singular so that the low regularity of the potential u does not allow the direct application of the standard FEM approximation techniques with H^1 finite elements. To overcome the computational difficulties caused by lack of regularity, several solution decomposition (regularization) techniques have been suggested in the literature. In Appendix A, we sketch the most popular existing decomposition methods considered in the literature and applied in the framework of differential formulations of models describing the biomolecular electrostatics.

In what follows, we describe the RS tensor format developed in [7]. In the present paper this format constitutes the main numerical tool to locally modify the singular right-hand side in PBE in order to increase the accuracy and efficiency of its FEM numerical approximation. The main idea of the new regularization approach for the PBE is the splitting of the electrostatic potential based on the RS tensor decomposition of the singular right-hand side [35] in the respective atomic volumes of the biomolecules, thereby eliminating discontinuities in the potential at the solute-solvent interface. A detailed description of the new solution decomposition technique is provided in sections 4.1 and 4.2.

3. Sketch of the rank-structured tensor approximation of electrostatic potentials. Tensor-structured numerical methods are now becoming popular in scientific computing due to their intrinsic property of reducing the grid-based solution

of the multidimensional problems to essentially "1D" computations. These methods evolved from bridging of the traditional rank-structured tensor formats of multilinear algebra [52, 16] with the nonlinear approximation theory based on a separable representation of multivariate functions and operators [22, 19, 33]. One of the ingredients in the development of tensor methods was the RHOSVD, one which allows one to reduce the rank of tensors in a canonical format by the C2T decomposition without the need to construct the full-size tensor [36]. Originally, it was used for the reduction of the ranks of canonical tensors when calculating 3D convolution integrals in computational quantum chemistry; see [34, 32] and the references therein.

Recently, tensor-based approaches were suggested as new methods for the calculation of multiparticle long-range interaction potentials. For a given nonlocal generating kernel $p(\|\bar{x}\|)$, $\bar{x} \in \mathbb{R}^3$, the calculation of the weighted sum of interaction potentials in an N-particle system, with the particle locations at $\bar{x}_\nu \in \mathbb{R}^3$, $\nu = 1, \ldots, N$,

$$P_0(\bar{x}) = \sum_{\nu=1}^{N} z_\nu \, p(\|\bar{x} - \bar{x}_\nu\|), \quad z_\nu \in \mathbb{R}, \quad \bar{x}_\nu, \bar{x} \in \Omega = [-b, b]^3, \tag{3.1}$$

is computationally demanding for large N. Since the generating radial function $p(\|\bar{x}\|)$ exhibits a slow polynomial decay in $1/\|\bar{x}\|$ as $\|\bar{x}\| \to \infty$, it follows that each individual term in (3.1) contributes essentially to the total potential at each point in the computational domain Ω. This predicts the $\mathcal{O}(N)$ complexity for a straightforward summation at every fixed space point $\bar{x} \in \mathbb{R}^3$. Moreover, in general, the radial function $p(\|\bar{x}\|)$ has a singularity or a cusp at the origin, $\bar{x} = 0$, making its accurate grid representation problematic. An efficient numerical scheme for the grid-based calculation of $P(\bar{x})$ in multiparticle systems can be constructed by using the RS tensor format [7]; see the overview in section 3.2.

3.1. Canonical tensor approximation of the Newton kernel. First, we recall the grid-based method for the low-rank canonical representation of a spherically symmetric kernel function $p(\|\bar{x}\|)$, $\bar{x} \in \mathbb{R}^d$ for $d = 2, 3, \ldots$, by its projection onto the set of piecewise constant basis functions; see [8] for the case of the Newton kernel $p(\|\bar{x}\|) = \frac{1}{\|\bar{x}\|}$ for $\bar{x} \in \mathbb{R}^3$. A single reference potential like $1/\|\bar{x}\|$ can be represented on a fine 3D $n \times n \times n$ Cartesian grid as a low-rank canonical tensor [22, 8, 32].

In the computational domain $\Omega = [-b, b]^3$, let us introduce the uniform $n \times n \times n$ rectangular Cartesian grid Ω_n with mesh size $h = 2b/n$ (n even). Let $\{\psi_{\mathbf{i}}\}$ be a set of tensor-product piecewise constant basis functions, $\psi_{\mathbf{i}}(\bar{x}) = \prod_{\ell=1}^{3} \psi_{i_\ell}^{(\ell)}(\bar{x}_\ell)$, for the 3-tuple index $\mathbf{i} = (i_1, i_2, i_3)$, $i_\ell \in I_\ell = \{1, \ldots, n\}$, $\ell = 1, 2, 3$. The generating kernel $p(\|\bar{x}\|)$ is discretized by its projection onto the basis set $\{\psi_{\mathbf{i}}\}$ in the form of a third order tensor of size $n \times n \times n$, defined entrywise as

$$\mathbf{P} := [p_{\mathbf{i}}] \in \mathbb{R}^{n \times n \times n}, \quad p_{\mathbf{i}} = \int_{\mathbb{R}^3} \psi_{\mathbf{i}}(\bar{x}) p(\|\bar{x}\|) \, \mathrm{d}\bar{x}. \tag{3.2}$$

Then using the Laplace–Gauss transform and sinc-quadratures, the third order tensor $\mathbf{P}$ can be approximated by the R-term canonical representation (see [22, 8, 31] for details),

$$\mathbf{P} \approx \mathbf{P}_R = \sum_{k=1}^{R} \mathbf{p}_k^{(1)} \otimes \mathbf{p}_k^{(2)} \otimes \mathbf{p}_k^{(3)} \in \mathbb{R}^{n \times n \times n}, \tag{3.3}$$

where $\mathbf{p}_k^{(\ell)} \in \mathbb{R}^n$.

Note that the reference tensor for summation of the potentials is generated in a larger computational domain, $\widetilde{\mathbf{P}}_R \in \mathbb{R}^{2n\times 2n\times 2n}$, which is necessary for application of the shift-and-windowing transforms $\mathcal{W}_\nu$; see section 3.2 and [32] for more details.

The canonical tensor representation of the Newton kernel (3.3) has been successfully applied in computation of multidimensional operators in quantum chemistry [36, 32]. It was shown that calculations using the grid-based tensor approximations exhibit the same high accuracy level as the analytically based computations for these operators. For the recent assembled tensor summation method for charged particles on rectangular finite lattices [31] it was proven that the canonical tensor rank of the collective electrostatic potential of large many-particle lattices equals to the rank of a single generating Newton kernel.

3.2. Short description of the RS tensor format. The RS canonical tensor format was introduced in [7] for calculation of the free space collective electrostatic potential of a many-particle system of general type.

DEFINITION 3.1 (RS canonical tensors [7]). *Given a reference tensor* $\mathbf{A}_0$ *supported by a small box such that* $\mathrm{rank}(\mathbf{A}_0) \le R_0$, *the separation parameter* $\gamma \in \mathbb{N}$, *and a set of distinct points* $x_\nu \in \mathbb{R}^d$, $\nu = 1,\dots,N$, *the RS canonical tensor format specifies the class of d-tensors* $\mathbf{A} \in \mathbb{R}^{n_1\times\cdots\times n_d}$, *which can be represented as a sum of a rank-*R_L *canonical tensor*

$$\mathbf{A}_{R_L} = \sum_{k=1}^{R_L} \xi_k \mathbf{a}_k^{(1)} \otimes \cdots \otimes \mathbf{a}_k^{(d)} \in \mathbb{R}^{n_1\times\cdots\times n_d} \tag{3.4}$$

and a cumulated canonical tensor

$$\widehat{\mathbf{A}}_S = \sum_{\nu=1}^{N} c_\nu \mathbf{A}_\nu, \tag{3.5}$$

generated by replication of the reference tensor $\mathbf{A}_0$ *to the points* x_ν. *Then the RS canonical tensor is represented in the form*

$$\mathbf{A} = \mathbf{A}_{R_L} + \widehat{\mathbf{A}}_S = \sum_{k=1}^{R_L} \xi_k \mathbf{a}_k^{(1)} \otimes \cdots \otimes \mathbf{a}_k^{(d)} + \sum_{\nu=1}^{N} c_\nu \mathbf{A}_\nu, \tag{3.6}$$

where diam$(\mathrm{supp}\mathbf{A}_0) \le 2\gamma$ *in the index size.*

In our application, we have $\mathbf{A}_{R_L} = \mathbf{P}_{R_L}$, where $\mathbf{P}_{R_L}$ is defined in (3.10), while the short-range part is specified by $\widehat{\mathbf{A}}_S = \mathbf{P}_s$.

Lemma 3.9 in [7] presents the storage cost of the RS canonical tensor $\mathbf{A}$ in (3.6) as follows:

$$\mathrm{mem}(\mathbf{A}) \le dR_L n + (d+1)N + dR_0\gamma.$$

Given $\mathbf{i} \in \mathcal{I} = I_1 \times \cdots \times I_d$, denote by $\overline{\mathbf{a}}_{i_\ell}^{(\ell)} \in {\mathbb{R}_{\mathbb{L}}}^{1\times R}$ the row-vector with index i_ℓ in the side matrix $A^{(\ell)} \in \mathbb{R}^{n_\ell\times R_L}$ of $\mathbf{A}$, and let $\xi = (\xi_1,\dots,\xi_d)$. Then the $\mathbf{i}$th entry of the RS canonical tensor $\mathbf{A} = [a_{\mathbf{i}}]$ can be calculated as a sum of long- and short-range contributions by

$$a_{\mathbf{i}} = \left(\odot_{\ell=1}^{d} \overline{\mathbf{a}}_{i_\ell}^{(\ell)}\right) \xi^T + \sum_{\nu\in\mathcal{L}(\mathbf{i})} c_\nu \mathbf{A}_\nu(\mathbf{i}),$$

at the expense $O(dR_L + 2d\gamma R_0)$. Here, $\mathcal{L}(\mathbf{i}) := \{\nu \in \{1,\dots,N\} : \mathbf{i} \in \mathrm{supp}\mathbf{A}_\nu\}$ is the set of indexes which label all the short-range tensors $\mathbf{A}_\nu$ that include the grid point $\mathbf{i}$ within their effective support [7].

The RS tensor can be represented in a Tucker tensor format as well; see [7]. RS tensors are efficient in many applications, for example, for modeling of the electrostatics of many-particle systems of general type or for modeling of scattered multidimensional data by using radial basis functions.

In what follows, we demonstrate how the RS tensor decomposition applies to the construction of low-complexity rank-structured tensor representation of the electrostatic potential of a many-particle system, discretized on a fine 3D Cartesian grid.

First, we recall the main ingredients for modeling of the long-range interaction potential in multiparticle systems of general type. The approach is based on the partitioning of the reference tensor representation of the Newton kernel into long- and short-range parts with a following assembling of the collective electrostatic potential of a molecular system in a special way. According to the tensor canonical representation of the Newton kernel (3.3) as a sum of Gaussians, one can distinguish their supports as the short- and long-range parts,

$$\mathbf{P}_R = \mathbf{P}_{R_s} + \mathbf{P}_{R_l},$$

where

$$\mathbf{P}_{R_s} = \sum_{k\in\mathcal{K}_s} \mathbf{p}_k^{(1)} \otimes \mathbf{p}_k^{(2)} \otimes \mathbf{p}_k^{(3)}, \quad \mathbf{P}_{R_l} = \sum_{k\in\mathcal{K}_l} \mathbf{p}_k^{(1)} \otimes \mathbf{p}_k^{(2)} \otimes \mathbf{p}_k^{(3)}. \tag{3.7}$$

Here, $\mathcal{K}_l := \{k|k = 0, 1, \ldots, R_l\}$ and $\mathcal{K}_s := \{k|k = R_l + 1, \ldots, M\}$ are the sets of indexes for the long- and short-range canonical vectors. Then the optimal splitting (3.7) is applied to the reference canonical tensor $\mathbf{P}_R$ and to its accompanying version $\widetilde{\mathbf{P}}_R = [\widetilde{p}_R(i_1, i_2, i_3)]$, $i_\ell \in \widetilde{I}_\ell$, $\ell = 1, 2, 3$, such that

$$\widetilde{\mathbf{P}}_R = \widetilde{\mathbf{P}}_{R_s} + \widetilde{\mathbf{P}}_{R_l} \in \mathbb{R}^{2n\times 2n\times 2n}.$$

In this way, the generating Newton kernel, discretized on $2n \times 2n \times 2n$ grid, is placed in the origin of the 3D box of twice larger size than the computational domain Ω. For every particle in an N-particle system the rank-1 shift-and-windowing operator [31]

$$\mathcal{W}_\nu = \mathcal{W}_\nu^{(1)} \otimes \mathcal{W}_\nu^{(2)} \otimes \mathcal{W}_\nu^{(3)}, \quad \nu = 1, \ldots, N,$$

is applied for shifting the generating Newton kernel according to the number of grid points (in the sets of univariate indexes $i_\ell \in \widetilde{I}_\ell$, $\ell = 1, 2, 3$) corresponding to the x-, y-, and z-coordinates of the corresponding particle and then windowing (restriction) of the replicated tensor to the computational domain of grid size $n \times n \times n$.

Then the total electrostatic potential $P_0(\bar{x})$ in (3.1) is represented by a projected tensor $\mathbf{P}_0 \in \mathbb{R}^{n\times n\times n}$ that can be constructed by a direct sum of shift-and-windowing transforms of the reference tensor $\widetilde{\mathbf{P}}_R$ (see also [32] for more details),

$$\mathbf{P}_0 = \sum_{\nu=1}^{N} z_\nu \, \mathcal{W}_\nu(\widetilde{\mathbf{P}}_R) = \sum_{\nu=1}^{N} z_\nu \, \mathcal{W}_\nu(\widetilde{\mathbf{P}}_{R_s} + \widetilde{\mathbf{P}}_{R_l}) =: \mathbf{P}_s + \mathbf{P}_l. \tag{3.8}$$

Thus, the shift-and-windowing transform $\mathcal{W}_\nu$ maps a reference tensor $\widetilde{\mathbf{P}}_R \in \mathbb{R}^{2n\times 2n\times 2n}$ onto its subtensor of smaller size $n \times n \times n$, obtained by first shifting the center of the reference tensor $\widetilde{\mathbf{P}}_R$ to the grid point x_ν and then restricting (windowing) the result onto the computational grid Ω_n. However, the tensor representation (3.8) is nonefficient as the ranks are growing linearly in the number of particles and remain nonreducible in both canonical and Tucker tensor formats.

This problem is solved in [7] by considering the global tensor decomposition of only the "long-range part" in the tensor $\mathbf{P}_0$, defined by

$$\mathbf{P}_l = \sum_{\nu=1}^{N} z_\nu \, \mathcal{W}_\nu(\widetilde{\mathbf{P}}_{R_l}) = \sum_{\nu=1}^{N} z_\nu \, \mathcal{W}_\nu \left(\sum_{k \in \mathcal{K}_l} \widetilde{\mathbf{p}}_k^{(1)} \otimes \widetilde{\mathbf{p}}_k^{(2)} \otimes \widetilde{\mathbf{p}}_k^{(3)} \right). \tag{3.9}$$

In [7] it was proven that the canonical rank R_L of the long-term sum tensor $\mathbf{P}_l$ (the number of rank-1 terms in $\mathbf{P}_l$ is about NR_l) only logarithmically depends on the number of particles N involved in the summation. It was also shown in [36] that the rank reduction

$$\mathbf{P}_l \mapsto \mathbf{P}_{R_L} \tag{3.10}$$

can be efficiently implemented by using the canonical-to-Tucker (C2T) and Tucker-to-canonical (T2C) algorithms, where the RHOSVD decomposition is a key ingredient [36]. As for the short-range part of the collective potential, in the RS format it is represented by a single small size tensor supplemented by a list of the particles' coordinates.

The RHOSVD-based C2T algorithm provides the reduction of the canonical rank due to exponentially fast decay of the singular values in the singular value decomposition of the side matrices for the canonical tensor $\mathbf{P}_l$; see details and the illustrating figures in [7]. Combination of the C2T and T2C algorithms does not provide the best canonical approximation; however, we observe the exponentially fast convergence in the Tucker rank. This guarantees rather low canonical rank with an upper bound equal to the square of the Tucker rank. Note that the C2T and T2C algorithms have been used for reducing the ranks of the canonical tensors in many applications of tensor numerical methods in quantum chemistry; see [32].

We summarize the tensor-based computational scheme which is the main ingredient in the construction of our the RS splitting of the solution of the PBE.

SUMMARY 3.2. *In the case of multiparticle systems, we need the low-rank tensor decomposition of the large sum of long-range canonical tensors (precomputed using the single reference tensor for the Newton kernel), resulting in the cumulative canonical tensor with large initial rank proportional to the number of single summands. The construction of this tensor sum is done by translation (replication and summation) of the reference tensor (centered at the origin) to the corresponding atomic centers (shift-and-windowing transform). Then we perform the rank reduction in the resultant cumulative canonical tensor with controllable precision. This rank-reduction algorithm includes two steps (the so-called can-to-Tucker-to-can transform* [36]*:*

(A) *Compute the low-rank Tucker decomposition (with controllable accuracy) of the initial canonical tensor with rather large canonical rank (in our case the initial rank is proportional to the number of particles).*

(B) *Transform the low-rank Tucker tensor to the canonical one without lost of accuracy, so that the canonical rank becomes only slightly larger than the Tucker rank (this is done by transformation of the small size Tucker core tensor to the canonical format without loss of accuracy).*

The approach is justified by the fact that for the considered $3D$ *tensor (obtained from the discretized Newton kernel) it is proven in the previous works* [22, 19, 33, 32] *that the Tucker approximation error decays exponentially fast in the Tucker rank. Due to this favorable feature the whole regularization scheme can be implemented with controllable precision by tuning the rank parameters. In application to multiparticle electrostatics the effective canonical rank is almost independent on the number of particles.*

Next, we illustrate the performance of the canonical RS tensor format in calculating the collective free space electrostatic potential of a model molecular system with 782 atoms; see also [7]. Summation is performed using the canonical tensor representing the reference Newton potential computed on the $n \times n \times n$ 3D Cartesian grid with $n = 257$ and the canonical rank $R = 29$. The RS tensor construction (for a single Newton kernel) is performed with $R_l = 14$, $R_s = 15$, simply by dividing the canonical vectors into two groups, that is, from every 29 vectors of the reference tensor and for every space dimension (x, y, z), 15 sharp Gaussians are separated as the short-range part and 14 smoother Gaussians as the long-range part.[1] Then their contributions to the collective sum of Newton potentials are calculated separately.

To reduce the rank of the long-range collective sum, the C2T (and T2C) transforms using RHOSVD [36] are used, with a choice of the truncation threshold 10^{-8}. We notice that in our scheme the RS splitting of the collective free space electrostatic potential is precomputed by using the RHOSVD algorithm in a preprocessing step before setting up the modified PBE. The numerical cost of RHOSDV scales linear in the number of particles N, $O(n^2 N)$, and it can be reduced to $O(nN)$ for larger values on n and N. In the presented example the resulting canonical rank decreases from $NR_l = 10948$ to $R_L = 382$, and it only logarithmically depends on the number of particles N; see [7].

The left panel in Figure 2 shows the cross section of the collective electrostatic potential of a molecular system at the middle of a z-plane, while the right panel shows the cross section of only the short-range part of the collective potential $\widehat{\mathbf{A}}_S$. Notice that in the left panel showing the total potential sum, the potential at the point of the plane with $(x, y) = (191, 143)$ equals to ~ 0.057 units, while for the short range (right panel) the sum equals to $2.8 \cdot 10^{-6}$. The left panel in Figure 3 presents the cross section of the low-rank long-range part of the collective potential at the same plane, while the right panel in this figure shows the error of the rank reduction (the same truncation threshold 10^{-8} as chosen above).

4. Regularization of PBE by using RS tensor format. In what follows, we describe the new approach for the construction of computationally effective boundary/interface conditions and source terms in the PBE describing the electrostatic potential of a biomolecule in gas phase and in solvent by solving the FEM/FDM discretization of the regularized PBE. The main advantage of our approach is due to complete avoidance of the direct FEM approximation (interpolation) of the highly singular right-hand sides in the traditional formulation of the PBE and, at the same time, preventing the modification of the stiffness matrix and/or the continuity conditions across the interface in the chosen FEM discretization scheme for elliptic PDE with jumping coefficients.

4.1. Across-the-interface tensor-based regularization scheme. The traditional numerical approaches for solving the PBE are based on either multigrid [46] or domain decomposition [10] methods which may be combined with boundary element methods [56, 20, 11] and [42]. We refer to [45, 23, 9] concerning the interface-based FDM solver; see also references therein.

The practically useful solution methods for the PBE are based on regularization schemes aiming at removing the singular component from the potentials in the

[1] Alternatively, separation of the reference tensor into the short- and long-range Gaussians can be performed by a chosen ε-truncation for the given interatomic distance.

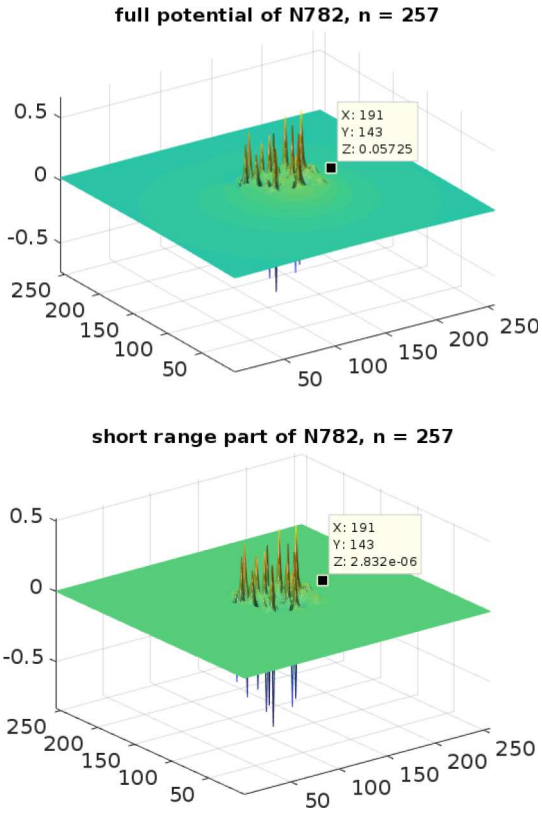

FIG. 2. *The full free space electrostatic potential of a system with* 782 *particles (left) and the sum of their short-range contributions (right).*

governing equation. Besides the regularization techniques discussed in Appendix A, here we consider one of the most commonly used approaches based on the additive splitting of the potential in the molecular region Ω_m; see, for example, [46]. To that end, we first discuss the recent version of additive splitting techniques introduced in [6], based on the application of the RS tensor format,

$$u = u^r + u_0^m, \quad \text{where} \quad u_0^m = 0 \quad \text{in} \quad \Omega_s, \tag{4.1}$$

and where the singular component u_0^m satisfies the following Poisson equation (PE) in Ω_m:

$$-\epsilon_m \Delta u_0^m = \rho_f \quad \text{in} \quad \Omega_m; \quad u_0^m = 0 \quad \text{on} \quad \Gamma. \tag{4.2}$$

In this case, (2.1) can be transformed to an equation for the regular potential u^r:

$$-\nabla \cdot (\epsilon \nabla u^r) + \bar{\kappa}^2 u^r = 0 \quad \text{in} \quad \Omega, \tag{4.3}$$

$$[u^r] = 0, \quad \left[\epsilon \frac{\partial u^r}{\partial n}\right] = -\epsilon_m \frac{\partial u_0^m}{\partial n} \quad \text{on} \quad \Gamma.$$

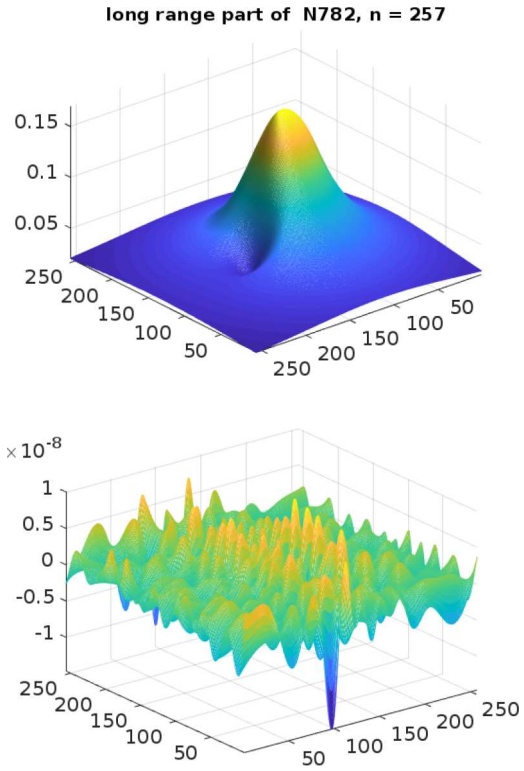

FIG. 3. *The low-rank tensor representation of the long-range part in the electrostatic potential of 782 charged particles (left) and the error of the canonical rank reduction.*

To facilitate the solution of (4.2) with highly singular data in the right-hand side, the singular potential U in the free space was utilized (see [7]),

$$-\epsilon_m \Delta U = \rho_f \quad \text{in} \quad \mathbb{R}^3, \quad |U(\bar{x})| \to 0, \quad |\bar{x}| \to \infty, \tag{4.4}$$

which can be written in the explicit form

$$U(\bar{x}) = \frac{1}{4\pi\epsilon_m} \sum_{k=1}^{N} \frac{z_k}{\|\bar{x} - \bar{x}_k\|}.$$

Introduce the characteristic (indicator) function, $\chi_{[\Omega_m]}(\bar{x})$, $\bar{x} \in \Omega$, of the domain $\Omega_m \subset \Omega$ by

$$\chi_{[\Omega_m]}(\bar{x}) = \begin{cases} 1 & \text{if } \bar{x} \in \overline{\Omega}_m, \\ 0 & \text{if } \bar{x} \in \Omega_s = \Omega \setminus \overline{\Omega}_m. \end{cases} \tag{4.5}$$

Then the restriction of U onto Ω_m can be calculated by

$$u^m = \chi_{[\Omega_m]} U,$$

implying the decomposition

$$u_0^m = u^m + u^{harm},$$

where the harmonic function u^{harm} compensates the discontinuity of u^m on Γ,

$$\Delta u^{harm} = 0 \quad \text{in} \quad \Omega_m; \quad u^{harm} = -u^m = -U \quad \text{on} \quad \Gamma.$$

The advantage of this formulation is twofold:

1. the absence of singularities in the solution u^r, and
2. the localization of the solution splitting only in the domain Ω_m.

Grid representation of the free space singular potential U, which may include a sum of hundreds or even thousands of single Newton kernels in three dimensions, leads to a challenging computational problem. In our approach it can be represented on large tensor grids in Ω with controlled precision by using the RS tensor format [7] characterized by the separability constant $\gamma > 0$ which in our application can be associated with the van der waals interatomic distance; see section 3.2. The long-range component in the formatted parametrization remains smooth and allows global low-rank representation in Ω. We conclude with the following.

PROPOSITION 4.1. *Let the effective support of the short-range components in the reference potential $\mathbf{P}_R$ be chosen not larger than $\gamma/2$. Then the interface conditions in the regularized formulation of the PBE in* (4.3) *depend only on the low-rank long-range component in the free space electrostatic potential of the atomic system. The numerical cost to build up the interface conditions on Γ in* (4.3) *does not depend on the number of particles N.*

The regularization $u = u^r + u_0^m$ like in (4.2)–(4.3) benefits from the local-global separability in the low-rank RS tensor representation of the free space electrostatic potential. Notice that here, we describe the splitting scheme in (4.2)–(4.3) just for illustration of the applicability of the RS tensor format for the solution of the PBE. This scheme requires modification of the interface and boundary conditions that is equivalent to a change of the system matrix which leads to a complicated implementation scheme. To avoid this nontrivial task, in what follows, we introduce an alternative approach, which avoids the additional computation of the auxiliary harmonic function u^{harm} in the rather complicated domain Ω_m as well as the modification of the interface conditions.

4.2. The localized RS tensor splitting scheme. In this section, we present the new splitting scheme which is based on the RS representation of the Dirac δ-distribution [35], which composes the highly singular right-hand side in the target PBE (2.1) or PE (4.2). First, we consider the PE as the proof of concept and validate the numerical results in section 5. We also present numerical tests for the linear PBE (LPBE). Following [35], we modify the right-hand side ρ_f in such a way that the short-range part in the solution u can be precomputed independently by the direct tensor decomposition of the free space potential, and the initial elliptic equation applies only to the long-range part of the total potential. The latter is a smooth function, hence the FDM/FEM approximation error can be reduced dramatically even on relatively coarse grids in three dimensions.

To fix the idea, we consider the simplest case of the single atom with unit charge located at the origin, such that the exact electrostatic potential reads $u(\bar{x}) = \frac{1}{\|x\|}$, $x \in \mathbb{R}^3$. Recall that the Newton kernel (3.3) discretized by the R-term sum of Gaussian type functions living on the $n \times n \times n$ tensor grid Ω_n is represented by a sum of short- and long-range tensors,

$$\frac{1}{\|x\|} \rightsquigarrow \mathbf{P}_R = \mathbf{P}_{R_s} + \mathbf{P}_{R_l} \in \mathbb{R}^{n\times n\times n},$$

where $\mathbf{P}_{R_s}$ and $\mathbf{P}_{R_l}$ are defined in (3.7).

Let us formally discretize the exact equation for the Newton potential, $u(\bar{x}) = \frac{1}{\|x\|}$,

$$-\Delta \frac{1}{\|x\|} = 4\pi\, \delta(\bar{x}),$$

by using the FDM/FEM Laplacian matrix A_Δ instead of Δ and via substitution of the canonical tensor decomposition $\mathbf{P}_R$ instead of $u(\bar{x}) = \frac{1}{\|x\|}$. This leads to the grid representation of the discretized Dirac delta [35]

$$\delta(\bar{x}) \rightsquigarrow \boldsymbol{\delta}_h := -\frac{1}{4\pi} A_\Delta \mathbf{P}_R$$

that will be applied in the framework of our discretization scheme.

We recall that the 3D finite difference Laplacian matrix A_Δ, defined on the uniform rectangular grid, takes the form

$$A_\Delta = \Delta_1 \otimes I_2 \otimes I_3 + I_1 \otimes \Delta_2 \otimes I_3 + I_1 \otimes I_2 \otimes \Delta_3, \tag{4.6}$$

where $-\Delta_\ell = h_\ell^{-2}\mathrm{tridiag}\{1, -2, 1\} \in \mathbb{R}^{n_\ell \times n_\ell}$, $\ell = 1, 2, 3$, denotes the discrete univariate Laplacian, such that the Kronecker rank of A_Δ equals to 3. Here I_ℓ, $\ell = 1, 2, 3$, is the identity matrix in the corresponding single dimension.

Now we are in a position to describe the RS tensor–based splitting scheme. To that end, we use the splitting of the discretized δ-distribution into short- and long-range components in the form [35]

$$\boldsymbol{\delta}_h = \boldsymbol{\delta}_s + \boldsymbol{\delta}_l, \tag{4.7}$$

where

$$\boldsymbol{\delta}_s := -\frac{1}{4\pi} A_\Delta \mathbf{P}_{R_s} \quad \text{and} \quad \boldsymbol{\delta}_l := -\frac{1}{4\pi} A_\Delta \mathbf{P}_{R_l}. \tag{4.8}$$

We observe that by construction, the short-range part vanishes on the interface Γ, hence it satisfies the discrete PE in Ω_m with the respective right-hand side in the form $\boldsymbol{\delta}_s$ and zero boundary conditions on Γ. Then we deduce that this equation can be subtracted from the full discrete linear system, such that the long-range component of the solution, $\mathbf{P}_{R_l}$, will satisfy the same linear system of equations (same interface conditions), but with a modified right-hand side corresponding to the weighted sum of the long-range tensors $\boldsymbol{\delta}_l$ only (see Lemmas 3.1 and 3.2 in [35]). In the simple example of the single charge, we arrive at the particular discrete PE for the long-range part in the full potential $\mathbf{P}_R$, $\mathbf{U}_l = \mathbf{P}_{R_l}$,

$$-A_\Delta \mathbf{U}_l = \boldsymbol{\delta}_l, \tag{4.9}$$

which can be solved by an appropriate method.

Figure 4 illustrates the modified right-hand side representing the long-range part of the discrete Dirac delta $\boldsymbol{\delta}_l$. It it worth noting that the FEM approximation theory can be applied to this formulation since the input data (i.e., the right-hand side) are regular enough in contrast to the initial formulation with the highly singular Dirac delta distribution in the right-hand side.

This scheme can be easily extended to the case of many-atomic systems just by additive representation of the short- and long-range parts in the total free space potential,

$$-A_\Delta \mathbf{P}_{R_l} = \boldsymbol{\delta}_{R_L}, \tag{4.10}$$

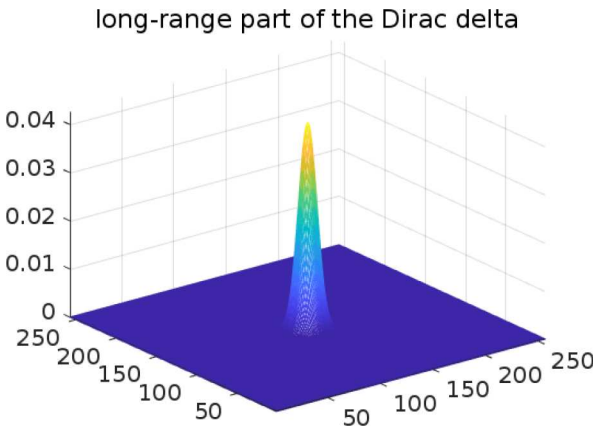

FIG. 4. *The long-range part of the Dirac delta* $\boldsymbol{\delta}_l$ *on an* $n^{\otimes 3}$ *3D grid,* $n = 257$.

where we suppose that R_L is the rank of the long-range part $\mathbf{P}_{R_l}$ of the corresponding RS tensor of type (3.6), and $\boldsymbol{\delta}_{R_L}$ is calculated as shown in (4.12). This decomposition scheme remains applicable to the LPBE; see section 4.3.

We summarize benefits of the aforementioned solution decomposition scheme.

- Most important is that due to efficient splitting of the short- and long-range parts in the target tensor representing both the single Newton kernel and the total free space potential, the singular component (short-range part) does not contribute to the jump condition at the interface.
- A remarkable advantage is that the long-range part in the RS tensor decomposition of the Dirac delta distribution [35] vanishes at the interface and, hence, the modified right-hand side generated by this long-range component remains localized in the "linear" solute region Ω_m.
- The boundary conditions on $\partial\Omega$ are obtained from the long-range part in the tensor representation of the collective electrostatic potential which reduces the computational costs involved, in contrast to solving the equation with some analytical function used to define the boundary condition.
- Only a single system of algebraic equations discretizing the elliptic PBE is solved for the smooth long-range (i.e., regularized) part of the collective potential discretized with controllable precision on a relatively coarse grid, which is then added to the directly precomputed (avoiding PDE solutions) low-rank tensor representation for the short-range contribution.

Next, we briefly comment on the general FEM approximation for the Laplacian. We recall the tensor-based scheme for evaluation of the Laplace operator in a separable basis set [32] applied for calculation of the kinetic energy part in the Fock operator.

Let the problem be posed in the finite volume box $\Omega = [-b, b]^3 \in \mathbb{R}^3$, subject to the homogeneous Dirichlet boundary conditions on $\partial\Omega$. For given discretization parameter $n \in \mathbb{N}$, the equidistant $n \times n \times n$ tensor grid $\omega_{\mathbf{3},N} = \{x_{\mathbf{i}}\}$, $\mathbf{i} \in \mathcal{I} := \{1, \ldots, n\}^3$, is used, with the mesh-size $h = 2b/(n-1)$. Then the Kronecker rank-3 tensor representation of the respective Galerkin FEM stiffness matrix is given by (4.6).

Notice that the MATLAB representation of the matrix A_Δ (say, the FD matrix) can be easily described in terms of `kron` operations as follows:

$$\frac{1}{h^2} A_\Delta = \texttt{kron}(\texttt{kron}(\Delta_1, I), I) + \texttt{kron}(\texttt{kron}(I, \Delta_1, I) + \texttt{kron}(\texttt{kron}(I, I), \Delta_1), \tag{4.11}$$

applied to a long vector of size n^3 representing the Newton potential.

Then the rank-structured calculation of the "collective" right-hand side $\boldsymbol{\delta}_{R_L}$ in (4.10) is reduced to 1D operations,

(4.12)
$$-\boldsymbol{\delta}_{R_L} = \sum_{k=1}^{R_L} \xi_k \left(\Delta_1 \mathbf{a}_k^{(1)} \otimes \mathbf{a}_k^{(2)} \otimes \mathbf{a}_k^{(3)} + \mathbf{a}_k^{(1)} \otimes \Delta_1 \mathbf{a}_k^{(2)} \otimes \mathbf{a}_k^{(3)} + \mathbf{a}_k^{(1)} \otimes \mathbf{a}_k^{(2)} \otimes \Delta_1 \mathbf{a}_k^{(3)} \right),$$

where $\mathbf{a}_k^{(\ell)}$, $\ell = 1, 2, 3$, are the canonical vectors and R_L is the canonical rank of the long-range part of the collective electrostatic free space potential of a biomolecule computed in the RS tensor format (3.6).

The tensor ansatz (4.12) is proposed to be used as the right-hand side in (4.10), as well as in the case of LPBE, which we study in the numerical experiments. With a subsequent application of the canonical-to-full tensor transform and after reshaping the three-way tensor $\boldsymbol{\delta}_{R_L}$ into a long vector, the result can be used in a standard PBE iterative solver as the right-hand side for calculation of the long-range part in the solution. Another advantage of our scheme is that the short-range part of the solution in the PBE (2.1) is obtained for free, since it is merely incorporated as the set of short-range parts of the respective Newton potentials for every particle in a biomolecule. That corresponds to a set of tensors in the second term of the collective electrostatic potential in the RS tensor format (3.6).

4.3. Sketch of the computational scheme for linear regularized PBE. We summarize the main computational tasks involved in the presented tensor-based numerical scheme. Note that here, we compute the original potential $\psi(\bar{x}) = \kappa_B T u(\bar{x})/e_c$ by rescaling the LPBE in (A.1) by $\kappa_B T/e_c$. The linear regularized PBE (LRPBE) can be solved by the following steps.

First, we compute the regularized right-hand side and the short-range part of the solution by using tensor techniques [32, 7, 35] by the RS tensor decomposition of the collective Dirac delta (RSDD) algorithm.

Algorithm RSDD.

1. *Generate the low-rank canonical tensor representation for the Newton kernel using $n \times n \times n$ Cartesian, with complexity $O(n)$* [8].
2. *Calculate the collective free-space electrostatic potential $\mathbf{P}_0 = \mathbf{P}_s + \mathbf{P}_l$ (see* (3.8)*) of an N-particle molecule by using the RS canonical tensor format* [7] *and the multigrid $C2T$ transform* [36]. *The complexity is of the order of $O(Nn)$.*
3. *Transform the short-range tensor $\mathbf{P}_s$ to a full format $n \times n \times n$-tensor and reshape it to long vector of size n^3, which serves as the short-range part in the solution of PBE* [35].
4. *Compute the long-range part of the collective Dirac delta, $\boldsymbol{\delta}_{R_L}$, as in* (4.12)*, by applying the Laplace operator to the long-range free-space electrostatic potential $\mathbf{P}_l$; see* [35]. *The complexity is in $O(nR)$, where R is the rank of the tensor $\mathbf{P}_l$.*

The splitting scheme described in section 4.2 allows one to reduce the initial equation for the solution of the system with modified right-hand side by using the new splitting via RS representation of the discretized Dirac delta distribution. In this approach the problem is reduced to computation of the short-range part in the collective free space electrostatic potential of the system and to the subsequent solution of the PBE equation for the long-range part only by the simple modification of the

right-hand side. The advantage is that the PBE applies to the smooth part of the total potential and hence a controllable FDM/FEM approximation error on moderate size 3D grids can be achieved.

Next, the quantities obtained in algorithm RSDD are used for the solution of the LRPBE.

Algorithm LRPBE.

1. *Compute the regularized component* $\boldsymbol{\delta}_{R_L}$ *of the Dirac delta distribution in* (4.10) by the RSDD algorithm.
2. *Substitute* $\boldsymbol{\delta}_{R_L}$ *into the right-hand side of the PBE* (2.1) *to obtain the LRPBE*

$$-\nabla \cdot (\epsilon \nabla u^r(\bar{x})) + \bar{\kappa}^2(\bar{x}) u^r(\bar{x}) = \boldsymbol{\delta}_{R_L}, \quad \text{in } \Omega, \tag{4.13}$$

   subject to the Dirichlet boundary conditions from (2.4).
3. *Discretize the LRPBE* (4.13) *to obtain the following system of equations*

$$A\mathbf{u}^r = \mathbf{b}, \tag{4.14}$$

   which can be solved by any linear system solver.
4. *Insert the short-range component* $\mathbf{u}^s = \mathbf{P}_s$ *of the free space potential in* (3.8) *computed in algorithm RSDD. Obtain the final LPBE solution* $\mathbf{u}$ *by the sum*

$$\mathbf{u} = \mathbf{u}^r + \mathbf{u}^s.$$

This method can be combined with the reduced basis approach for PBE with parametric coefficients to further accelerate the numerical computations [39, 5, 38, 40]. This is because the modified model is affinely dependent on the parameter ($\bar{\kappa}$), thereby providing a natural off-line/on-line decomposition that can be used within the reduced basis method.

Finally, we notice that an important characterization of the protein molecule is given by the electrostatic solvation energy [46], which is the difference between the electrostatic free energy in the solvated state (described by the PBE) and the electrostatic free energy in the absence of solvent, i.e., E_N. Having at hand the free energy E_N, the electrostatic solvation energy can be computed in the framework of the regularized formulation of the PBE as described above.

5. Numerical tests. In this section, we first consider the free space electrostatic potential computed by the modified PE and the RS tensor format–based splitting scheme. We compare the results with those of the traditional PE for various biomolecules. In this case, the PBE can be reduced to the PE by considering zero ionic strength which implies that $\bar{\kappa}^2 = 0$, hence the Boltzmann distribution term in (2.1) is annihilated. Consequently, homogeneous dielectric constants of $\epsilon_m = \epsilon_s = 1$ are considered. Moreover, we consider a numerical example for the LPBE such that $\bar{\kappa}^2(\bar{x}) > 0$ and $\epsilon_s/\epsilon_m \approx 40$.

We compute the electrostatic potentials using $n \times n \times n$ 3D Cartesian grids, in a box $[-b, b]^3$ with equal step size $h = 2b/(n-1)$. Conventional computations by the PBE/PE solver are limited to $n = 257$, on a PC with 8GB RAM due to the storage needs of the order of $\mathcal{O}(n^3)$. Notice that since we solve the discrete elliptic system only for the smooth long-range part of the collective potential, our approach allows us to have satisfactory approximation accuracy of the solution already for moderate grid size.

In our numerical tests, we first precompute the modified right-hand-side $\mathbf{b}$ in (4.14) (to be used to compute the regularized part $\mathbf{u}^r$ of the potential), as well as the singular part $\mathbf{u}^s$ in the total potential $\mathbf{u} = \mathbf{u}^r + \mathbf{u}^s$ (both presented on the Cartesian grid). We control the numerical error in the precomputed right-hand side and in the singular part $\mathbf{u}^s$ by turning the rank parameter in the RS tensor decomposition. Given the modified right-hand side as input data, the FDM discretization of the PBE (construction of the stiffness matrix A in (4.14)) and solving the arising linear system of equations (4.14) for $\mathbf{u}^r$ (with the modified right-hand side) is performed by using the adaptive Poisson–Boltzmann software (APBS) package (version 1.5-linux64), employing the multigrid (PMG) accelerated FDM [2, 25] as well as MPBEC [54]. The FDM error estimate is determined by the results of traditional FEM approximation theory applied to the chosen FDM discretization scheme, and the corresponding contribution to the total error can be considered independently on the error analysis of our tensor-based RS splitting sketched in section 3.2.

In RS tensor calculations, first, the reference Newton kernel is generated using the $2n \times 2n \times 2n$ 3D Cartesian grid, with a given ε- accuracy. Usually, for grids with the univariate size $n \leq 1024$, and $\varepsilon \sim 10^{-6}$, the rank is $R \leq 30$. The result is the tensor representation of the Newton kernel at the origin of the computational box, in a form (3.3) of a sum of discretized Gaussians, which are split into short- and long-range parts (by sorting them according to chosen criterion). After replication (shifting-windowing) of both short-range and long-range parts of the reference Newton kernel, according to coordinates of the charges particles, the tensors are separately summed. The short-range part is transformed directly from the canonical tensor to a full size format and then reshaped to a vector of size n^3 which becomes actually a singular part of the solution. The sum of long-range parts undergoes the C2T and T2C transforms to reduce the canonical rank, where precision is controlled by the truncation threshold. Then the Kronecker tensor product form of the 3D Laplace operator (4.12) is applied to the result of canonical rank truncation to produce the canonical tensor representation of the long-range contribution $\boldsymbol{\delta}_{R_L}$ to the modified right-hand side of the PBE. This tensor is first transformed to a full size format and then reshaped to a vector of size n^3 to be applied in the conventional PBE solver as the input for the right-hand side. Ultimately, the obtained solution of regularized PBE is added to the already precomputed short-range part. The basic tensor operations are performed by using the programs from the MATLAB TESC package on tensor numerical methods developed in recent years by the second and third authors; see [32] for short descriptions and related references.

5.1. Validating the accuracy of RS splitting for the free space potential. We compare the accuracy of the free space electrostatic potential calculated by the traditional PE model and by the PE model modified by the RS tensor format for the approximation of the single Newton kernel. Figure 5 shows the single Newton kernel on the grid with $n = 129$ at the cross section of the volume box in the middle of the z-axis computed by the canonical tensor approximation obtained by sinc-quadratures and the corresponding errors when using the traditional PE model and by the modified PE model computed by the FDM solver. We notice that the solution of the modified PE model is of higher accuracy than that of the traditional PE model because it captures the singularities exactly; see the central region of Figure 5 (right).

In a similar vein, we consider the electrostatic potential of the acetazolamide molecule consisting of 18 atoms and determine the accuracy of the calculation of the

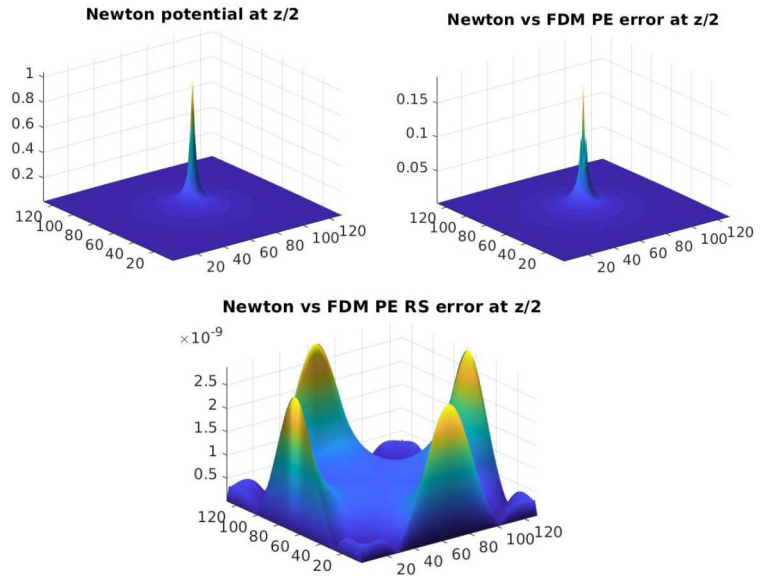

FIG. 5. *The Newton potential computed by the canonical tensor decomposition (left), by the error of its computation on the same grid by using the classical PE (middle), and by the RS-modified PE (right).*

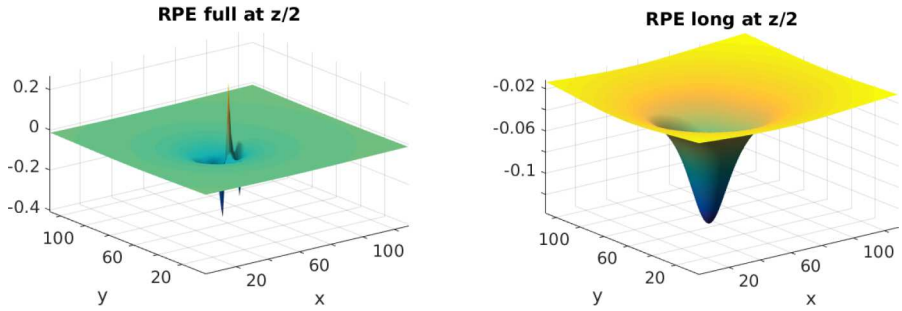

FIG. 6. *The total free space electrostatic potential (left) and its long-range component (right) computed by the regularized PE.*

free space potential obtained by traditional PE model versus the PE model modified by the RS tensor format. This molecule is used as a ligand in the human carbonic anhydrase (hca) protein-ligand complex for the calculation of the binding energy in the APBS package [25] and MPBEC [54].

Figure 6 shows the total free space electrostatic potential computed by the regularized PE and its long-range component, respectively, while Figure 7 shows the error between the exact Newton potential and the classical PE (left) and the regularized PE (right) visualized on an $n \times n$ grid surface with $n = 129$. It is shown that the RS-modified PE model provides more accurate solutions as compared with those obtained of the traditional PE due to the accurate treatment of the solution singularities by the RS tensor format.

We notice that the classical PE model does not capture accurately the singularities in the electrostatic potential due to the numerical errors introduced by the Dirac delta

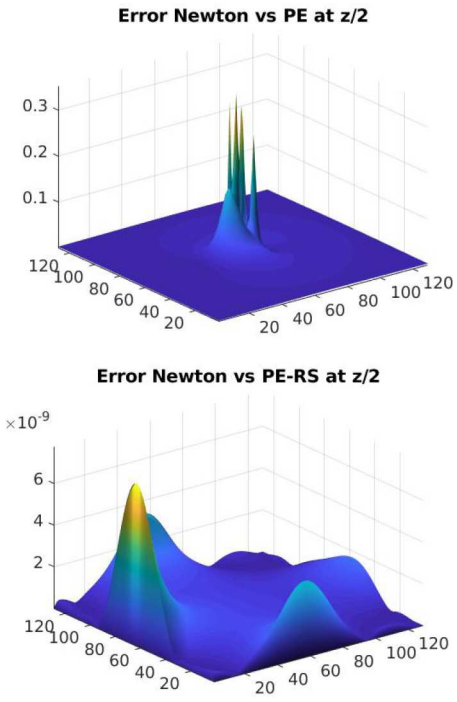

FIG. 7. *The error of the free space potential between the exact Newton potential sums and the classical PE solution (left) and between the Newton potential sums and the regularized PE solution (right).*

distribution and partly due to the smoothing effect caused by the spline interpolation of the charges onto the grid. The RS-modified PE model, on the other hand, is able to capture the singularities due to the independent treatment of the singularities by the RS tensor technique. This is demonstrated in Figure 8.

5.2. The regularized Poisson equation on a sequence of fine grids. Here, we illustrate the accuracy of the potential obtained by the RS-modified PE by calculating the free space electrostatic potential on a sequence of fine grids and compare it with the Newton (Coulomb) potential obtained by accurate canonical tensor calculations. We first consider the Coulomb potential case and show the absolute error for the finest Cartesian grid and the discrete L_2 norm of the error on a sequence of Cartesian grids.

Figure 9 shows the absolute error of $\mathcal{O}(10^{-11})$ obtained on a 257^3 Cartesian grid and 32 Å box length. Table 1 shows the decay of the discrete L_2 norm of the error on a sequence of refined grids.

Next, we consider the acetazolamide molecule with 18 atoms. The electrostatic potential is computed as in the previous case, employing the same grid properties. The absolute error is shown in Figure 10 and the error behavior with respect to mesh refinement is shown in Table 2.

Finally, we consider the protein Fasciculin 1, with 1228 atoms, an antiacetylcholinesterase toxin from green mamba snake venom [41]. The electrostatics potential is computed as in the previous case but with 60 Å box length and 257^3 as the minimum Cartesian grid size, due to a larger molecular size. The results shown in Figure 11 and Table 3 illustrate a similar trend of accuracy as in the previous test examples.

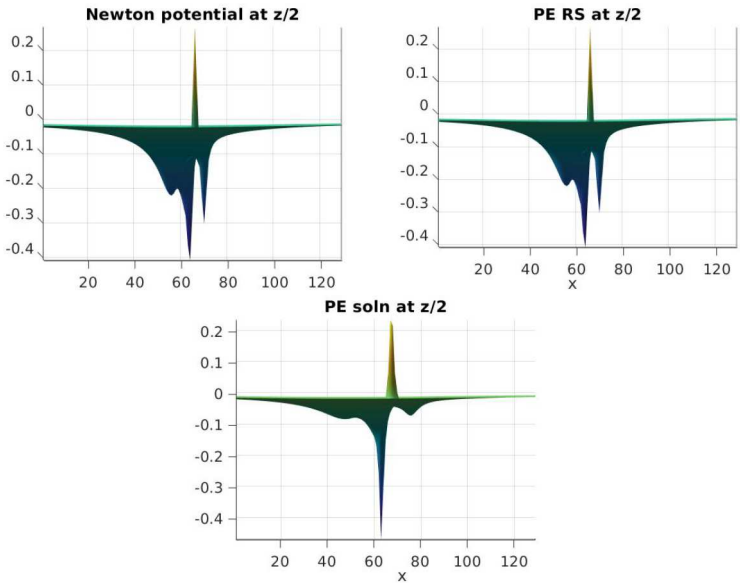

FIG. 8. *Demonstration of the solution singularities for the acetazolamide molecule captured by the canonical tensor approximation (left), by the RS-modified PE model (middle), and by the classical PE (right).*

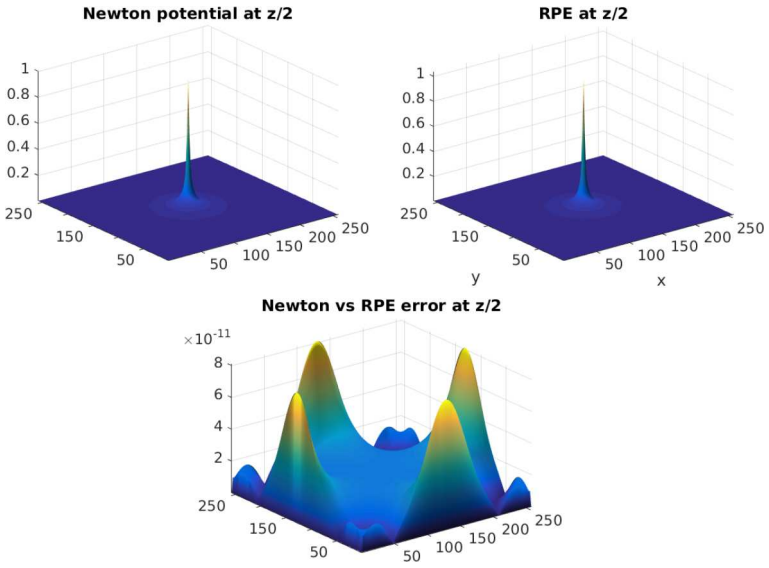

FIG. 9. *Absolute error between the solutions of the Newton potential and the RS-modified PE for the Born ion.*

5.3. Accurate representation of the long-range electrostatic potential by the RS tensor. Here, we highlight the advantages of the RS tensor format in the low-rank approximation of the long-range component in the total potential sum. For this purpose, the RHOSVD within the multigrid C2T transform [36] is used which provides computation of the low-rank canonical/Tucker tensor representation of the

TABLE 1
The discrete L_2 norm of the error with respect to grid size for the Born ion.

n	97	129	257
Discrete L_2 norm	4.1176×10^{-7}	9.5516×10^{-8}	2.7975×10^{-9}

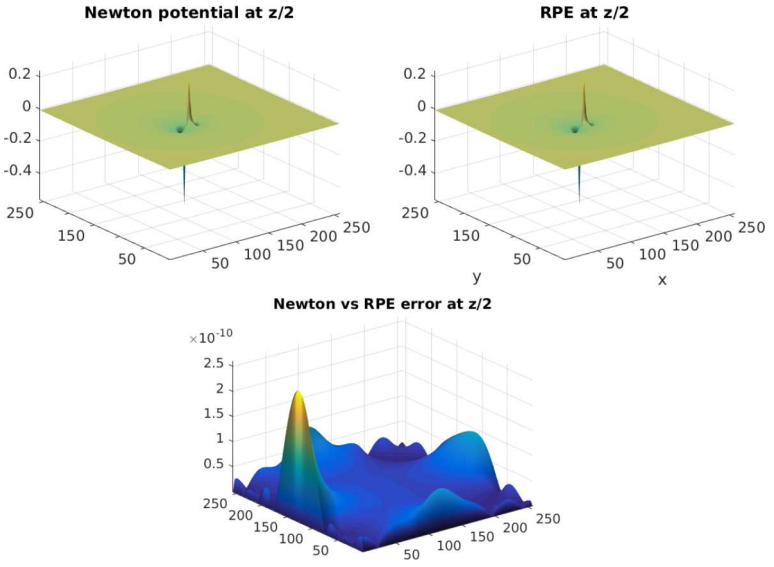

FIG. 10. *Absolute error between the solutions of the Newton potential sums and the RS-modified PE for the acetazolamide molecule.*

TABLE 2
The discrete L_2 norm of the error and the relative error with respect to grid size for the acetazolamide molecule.

n	97	129	257
Discrete L_2 norm	4.1176×10^{-7}	1.1936×10^{-7}	3.7003×10^{-9}

long-range part at the asymptotic cost of $\mathcal{O}(Nn)$. Here, N is the number of charges in the molecule while n represents the grid dimension in a single direction.

Figure 12 shows an error of $\mathcal{O}(10^{-5})$ of the RS tensor format approximation compared with the full size representation at various grid sizes. These data correspond to the long-range RS rank equal to 10, with ε-truncation threshold chosen as $\mathcal{O}(10^{-6})$ for the reference Newton kernel and $\mathcal{O}(10^{-7})$ for the C2T transform.

Figure 13 shows that the error is reduced by one order of magnitude, i.e., to $\mathcal{O}(10^{-6})$, if we take a stronger rank truncation criterion ε of an order less for both the Newton kernel, i.e., $\mathcal{O}(10^{-7})$, and for the C2T transform, $\mathcal{O}(10^{-8})$.

5.4. Numerical tests for LRPBE. In this section, we use the dielectric and kappa functions as defined in (A.3) for an inhomogeneous dielectric medium (i.e., $\epsilon_m = 2$, $\epsilon_s = 78.54$ and $\kappa^2 = 8\pi N_A e_c I/\epsilon_s$, where $N_A = 6.0221367 \times 10^{23}$, $e_c = 4.8032424 \times 10^{-10}$, and $I = 0.15M$) as the LRPBE parameters which are mapped on the grid using the smoothed molecular surface method, which is calculated using the Connolly approach [14]. The atomic charges are obtained from the acetazolamide molecule consisting of 18 atoms, which is used as a ligand in the human carbonic anhydrase (hca) protein-ligand complex for the calculation of the binding energy [25, 54]. Figure 14

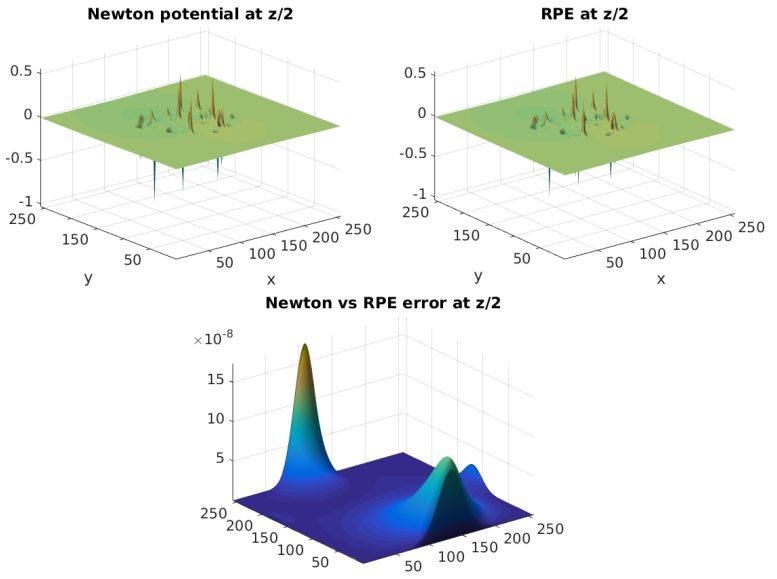

FIG. 11. *Absolute error between the solutions of the Newton potential sums and the RS-modified PE for the protein Fasciculin* 1.

TABLE 3
The discrete L_2 norm of the error and the relative error with respect to grid size for the protein Fasciculin 1.

n	129	193	257
Discrete L_2 norm	1.2919×10^{-6}	1.7395×10^{-7}	4.3060×10^{-8}

illustrates the full electrostatic potential computed for the acetazolamide molecule using the LRPBE (left) computed by using the RS tensor decomposition for the Dirac delta, and the long-range part of this potential (right) on an $n \times n \times n$ grid with $n = 129$ in the volume box (32 Å).

Figure 15 shows the difference between the potential from LPBE compared with that calculated by the new RS regularized scheme, for the case of 0.15 ionic strength. We observe that the computational error for the traditional regularization scheme indicated in Figure 7 (left) for the case of free space collective potential is almost of the same order of magnitude as that presented in Figure 15 for the case of linearized PBE.

Remark 5.1. We demonstrate the conspicuous difference between the RPE and the LPBE solutions in Figures 6 and 14, respectively, due to the effect of the inhomogeneous (or jumping) dielectric coefficient and the ionic strength in the latter. As a result of these coefficients, we can see that the short-range potential component of the LPBE in Figure 14 (left) is halved due to the change of ϵ_m from 1 in free space to 2 in the solvated state, while the long-range potential component in Figure 6 scales by approximately 1/80 in addition to the effect of the ionic strength as a damping coefficient.

5.4.1. Validation tests for the LRPBE. To validate the RS tensor–based LRPBE solver, we consider the analytical solution of the Born ball model similar to

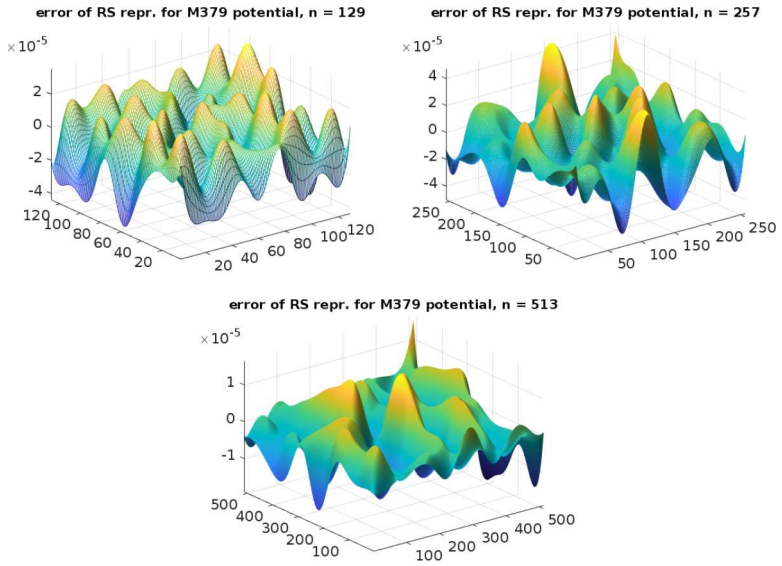

FIG. 12. *The error due to the low-rank approximation of the long-range component for the* 379 *atomic molecule for* $n = 129^3$ *(left),* $n = 257^3$ *(middle), and* $n = 513^3$ *(right) grids.*

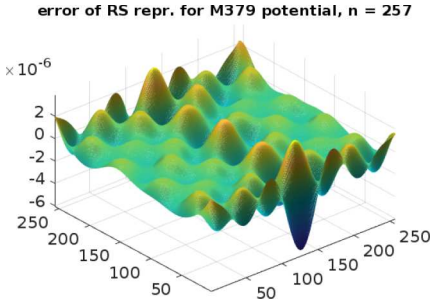

FIG. 13. *The error due to the low-rank approximation of the long-range component for the* 379 *atomic molecule for* $n = 257^3$ *grid at a lower tolerance.*

that in [58]. However, in our model, we use the centimeter-gram-second units instead of the SI units, in addition to the scaling differences for the constant 4π,

$$\psi(\bar{x}) = \begin{cases} \frac{\alpha z}{\epsilon_m d} + \frac{\alpha z}{R}\left(\frac{1}{\epsilon_s} - \frac{1}{\epsilon_m}\right) & \text{if} \quad \bar{x} \in \Omega_m, \\ \frac{\alpha z}{\epsilon_s d} & \text{if} \quad \bar{x} \in \Omega_s, \end{cases} \tag{5.1}$$

where $\Omega = \Omega_m \cup \Omega_s = [-6, 6]^3$, $d = \|\bar{x}\|$, $\alpha = 4\pi^2 e_c \times 10^8$, $\epsilon_m = 2$, $\epsilon_s = 78.54$, $z = 1$ is a unit charge, and $R = 3$ is the atomic radius of the Born ion. Note that the given values for ϵ_m and ϵ_s correspond to those of physiological processes, wherein the dielectric constant for biomolecules such as proteins is 2, while that of the surrounding ionic solution, whose major component is water, is 78.4. See [26] for details.

Table 4 illustrates the validation results for our LRPBE solver for the Born ball model on a sequence of fine grids. The discrete L_2 error demonstrates good convergence of our LRPBE model. Note that Editha, one of the computer clusters at

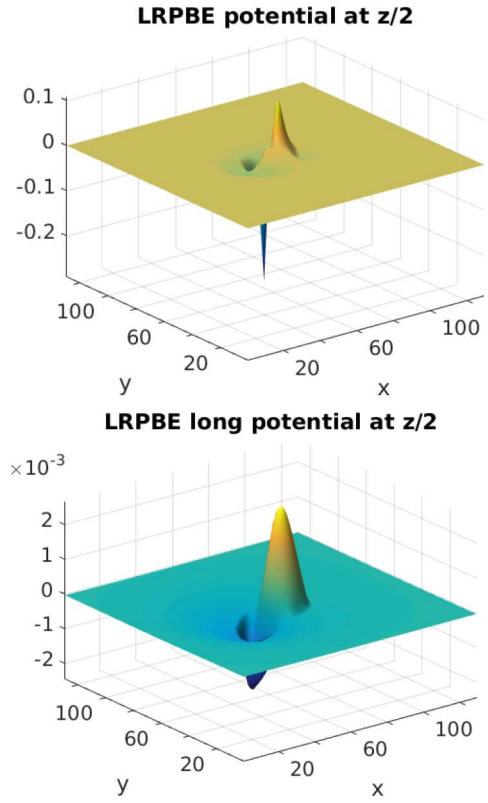

FIG. 14. *The full electrostatic potential (left) and its long-range component (right) computed by the LRPBE at* 0.15 *ionic strength for the acetazolamide molecule.*

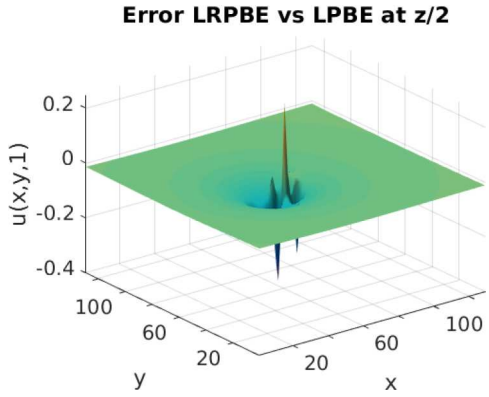

FIG. 15. *The difference between the potential from LPBE compared with that calculated by the new RS regularized scheme at* 0.15 *ionic strength.*

the MPI in Magdeburg, was used to carry out the last computation in Table 4 due to its large memory requirement, i.e., $56,623,104$ mesh points.

We point out that our numerical experiments indicate that the computational error is quite uniformly distributed over the computational domain depicting some small spikes at the atomic centers (this is expected); see Figures 12 and 13. Hence, the L_∞ error appears to be of the same order as the L_2 error.

TABLE 4
Discrete L_2 error between the solution to the LRPBE in (4.13) *and the analytical solution in* (5.1) *for a Born ion of unit charge and radius $R = 3$ on a sequence of fine grids.*

Mesh size h	Number of mesh points	Discrete L_2 error
$h_1 = 0.25$	110 592	5.1059e-02
$h_2 = h_1/2$	884 736	1.2504e-02
$h_3 = h_2/2$	7 077 888	3.0426e-03
$h_4 = h_3/2$	56 623 104	7.3218e-04

Notice that the computational runtimes for the classical LPBE and the LRPBE using our solvers are almost equal because the linear systems are solved by the same solver, that is, the aggregation-based algebraic multigrid (AGMG[2]) method [48, 49].

6. Conclusions. In this paper we demonstrate that the RS tensor format is gainfully applicable for the solution of the PBE for calculation of electrostatics in large molecular systems. The efficiency of the new tensor-based regularization scheme for the PBE is based on the exceptional properties of the grid-based RS tensor splitting of the Dirac delta distribution. The main computational benefits are due to the localization of the modified right-hand side within the molecular region and automatic maintaining of the continuity of the Cauchy data on the interface. Another advantage is that our computational scheme only includes solving a single system of algebraic equations for the smooth long-range (i.e., regularized) part of the collective potential discretized by the FDM. The total potential is obtained by adding this solution to the directly precomputed low-rank tensor representation for the short-range contribution.

The various numerical tests illustrate the main properties of the presented scheme. For example, it is clear from Figure 8 that the classical PE model does not accurately capture the solution singularities which emanate from the short-range component of the total target electrostatic potential in the numerical approximation. We emphasize that this problem can be efficiently circumvented by applying the RS tensor format as a solution decomposition technique in order to modify the PBE/PE. In the modified PBE/PE, the Dirac delta distribution is replaced by a smooth long-range function from (4.8). We thus only need to solve for the long-range electrostatic potential numerically and add this solution to the short-range component which is computed a priori using the canonical tensor approximation to the Newton kernel. The resultant total potential sum is of high accuracy as evident from Figures 9–11 and Tables 1–3. On the other hand, the computational error for the traditional regularization scheme indicated in Figure 7 (left) for the case of free space collective potential is almost of the same magnitude as that presented in Figure 15 for the case of linearized PBE.

Finally, we summarize that the regularization scheme presented in this paper has capabilities for various generalizations which can be effectively implemented with minor changes in the RS tensor decompositions. We notice the following directions:

- The possibility of PBE computations on much finer grids due to its good convergence behavior is also an advantage of the proposed approach. Tensor techniques practically can be applied to finer grids compared to traditional finite element approaches;

[2]AGMG implements an aggregation-based algebraic multigrid method, which solves algebraic systems of linear equations, and is expected to be efficient for large systems arising from the discretization of scalar second order elliptic PDEs [48].

- The regularization scheme remains verbatim in the case of the nonlinear PBE since it requires only the modification of input data for the right-hand side in the molecular region Ω_m, where the equation is linear, but this does not affect the nonlinearity domain Ω_s;
- Our approach allows the efficient calculation of electrostatics under multiple rotations of the biomolecule, which is a crucial problem in the numerical modeling of proteins.

Appendix A. Sketch of the existing analytical solution decomposition techniques. In this paper we present the new solution decomposition scheme for the elliptic PDE with piecewise constant coefficients on the example of linearized PBE; see the dimensionless formulation in (2.1). However, our approach also applies to the case of nonlinear PBE since the regularization procedure is completely localized in the "linear" molecular region Ω_m and does not change the equation on the interface and in the interior of Ω_m. Hence, in what follows, we sketch some commonly used regularization schemes for the case of nonlinear PBE, described in the literature. We discuss this issue for the equation in the form (A.1), where the specific physical constants are imposed because this is convenient for the comparison of our numerical results with some others presented in the literature.

The PBE models the dimensionless potential $u(\bar{x}) = e_c\psi(\bar{x})/\kappa_B T$, which is scaled by $e_c/\kappa_B T$, and $\psi(\bar{x})$ is the original electrostatic potential in centimeter-gram-second units at $\bar{x} = (x, y, z) \in \mathbb{R}^3$; see (2.1). It is given by

$$-\nabla \cdot (\epsilon(\bar{x})\nabla u(\bar{x})) + \bar{\kappa}^2(\bar{x}) \sinh(u(\bar{x})) = \frac{4\pi e_c^2}{\kappa_B T} \sum_{i=1}^{N_m} z_i \delta(\bar{x} - \bar{x}_i), \quad \Omega \in \mathbb{R}^3, \tag{A.1}$$

subject to

$$u(\bar{x}) = \frac{e_c^2}{\kappa_B T} \sum_{i=1}^{N_m} \frac{z_i e^{-\bar{\kappa}(d-a_i)}}{\epsilon_s(1+\bar{\kappa}a_i)d} \quad \text{on } \partial\Omega, \quad d = \|\bar{x} - \bar{x}_i\|, \tag{A.2}$$

where $\kappa_B T$, κ_B, T, and e_c are the thermal energy, the Boltzmann constant, the absolute temperature, and the electron charge, respectively, $\kappa^2 = 8\pi N_A e_c^2 I/1000\epsilon_s\kappa_B T$ is a function of the ionic strength $I = 1/2\sum_{j=1}^{m} c_j z_j^2$, where c_j and z_j are the charge and concentration of each ion, and N_A is the Avogadro constant. The sum of Dirac delta distributions, located at atomic centers $\bar{x}_i$, represents the molecular charge density, z_i are the point partial charges of the protein, ϵ_s is the solvent dielectric constant, a_i are the atomic radii, and N_m is the total number of point partial charges in the protein. The functions $\epsilon(\bar{x})$ and $\bar{\kappa}^2(\bar{x})$ are piecewise constant defined by

$$\epsilon(\bar{x}) = \begin{cases} \epsilon_m & \text{if } \bar{x} \in \Omega_m, \\ \epsilon_s & \text{if } \bar{x} \in \Omega_s, \end{cases} \qquad \bar{\kappa}(\bar{x}) = \begin{cases} 0 & \text{if } \bar{x} \in \Omega_m, \\ \sqrt{\epsilon_s}\kappa & \text{if } \bar{x} \in \Omega_s, \end{cases} \tag{A.3}$$

where Ω_m and Ω_s are the molecular and solvent regions, respectively, as shown in Figure 1.

Note that in (A.3), $\bar{\kappa}^2$ describes both the ion accessibility and the bulk ionic strength (or concentration) I. Consequently, we denote by $\bar{\kappa}(\bar{x})$ the position-dependent piecewise constant $\bar{\kappa}$ function in order to distinguish it from the scalar variable κ. The factor 1000 in the definition of $\bar{\kappa}$ is associated to the need of using the molarity in liters instead of cubic centimeters. The boundary condition (A.2) is the analytical solution for the linearized PBE for a spherical molecule. In fact, this form of potential

is analogous to the Yukawa-type potential, which is well known to decay exponentially with respect to distance. We refer to [26] for more details.

In order to overcome the difficulties arising from the singularities caused by the impulsive source term, several solution decomposition techniques have been suggested in the literature. In the following, we will discuss those techniques that form the state of the art for solving the PBE. Following [57], the first solution decomposition is generally given by

$$\begin{aligned} -\epsilon_m \Delta u(\bar{x}) &= C \sum_{i=1}^{N_m} z_i \delta(\bar{x} - \bar{x}_i), && \bar{x} \in \Omega_m, \\ -\epsilon_s \Delta u(\bar{x}) + \bar{\kappa}^2 \sinh(u(\bar{x})) &= 0, && \bar{x} \in \Omega_s, \\ u(s^+) = u(s^-), \quad \epsilon_s \frac{\partial u(s^+)}{\partial n(s)} &= \epsilon_m \frac{\partial u(s^-)}{\partial n(s)}, && s \in \Gamma, \\ u(s) &= g(s), && s \in \partial\Omega, \end{aligned} \tag{A.4}$$

where $C = 4\pi e_c^2/\kappa_B T$, s^+ and s^- represent the grid points in the vicinity of the interface Γ in the solvent and the molecular regions, respectively, while the Dirichlet boundary conditions $g(s)$ are defined by the right-hand side in (A.2). The solution $u(\bar{x})$ is decomposed as follows:

$$u(\bar{x}) = G(\bar{x}) + \tilde{\phi}(\bar{x}) + \tilde{\psi}(\bar{x}). \tag{A.5}$$

The corresponding components of $u(\bar{x})$ include the analytical solution $G(\bar{x})$ of the PE in the molecular domain,

$$G(\bar{x}) = \frac{C}{4\pi\epsilon_m} \sum_{i=1}^{N_m} \frac{z_i}{\|\bar{x} - \bar{x}_i\|}, \qquad \bar{x} \in \Omega_m, \tag{A.6}$$

the solution of the linear interface boundary value problem

$$\begin{aligned} \Delta\tilde{\phi}(\bar{x}) &= 0, && \bar{x} \in \Omega_m \cup \Omega_s, \\ \tilde{\phi}(s^+) = \tilde{\phi}(s^-), \quad \epsilon_s \frac{\partial \tilde{\phi}(s^+)}{\partial n(s)} &= \epsilon_m \frac{\partial \tilde{\phi}(s^-)}{\partial n(s)} + (\epsilon_m - \epsilon_s) \frac{\partial G(s)}{\partial n(s)}, && s \in \Gamma, \\ u(s) &= g(s) - G(s), && s \in \partial\Omega, \end{aligned} \tag{A.7}$$

and the solution of the nonlinear interface boundary value problem

$$\begin{aligned} \Delta\tilde{\psi}(\bar{x}) &= 0, && \bar{x} \in \Omega_m, \\ -\epsilon_s \Delta\tilde{\psi}(\bar{x}) + \bar{\kappa}^2 \sinh(\tilde{\psi}(\bar{x}) + \tilde{\phi}(\bar{x}) + G(\bar{x})) &= 0, && \bar{x} \in \Omega_s, \\ \tilde{\psi}(s^+) = \tilde{\psi}(s^-), \quad \epsilon_s \frac{\partial \tilde{\psi}(s^+)}{\partial n(s)} &= \epsilon_m \frac{\partial \tilde{\psi}(s^-)}{\partial n(s)}, && s \in \Gamma, \\ u(s) &= 0, && s \in \partial\Omega. \end{aligned} \tag{A.8}$$

Second, we recall the solution decomposition from [47] which takes the form

$$u(\bar{x}) = \hat{u}(\bar{x}) + \tilde{u}(\bar{x}).$$

The short-range part $\hat{u}(\bar{x})$ is given by

$$\hat{u}(\bar{x}) = \begin{cases} G(\bar{x}) + u^0(\bar{x}) & \text{if} \quad \bar{x} \in \Omega_m, \\ 0 & \text{if} \quad \bar{x} \in \Omega_s, \end{cases} \tag{A.9}$$

where $u^0(\bar{x})$ is a harmonic function which compensates for the discontinuity on the interface and satisfies the following conditions:

$$\begin{aligned} \Delta u^0(\bar{x}) &= 0 \qquad \text{if} \quad \bar{x} \in \Omega_m, \\ u^0(s) &= -G(s), \qquad\quad s \in \Gamma. \end{aligned} \tag{A.10}$$

The regular part $\tilde{u}(\bar{x})$ is represented by

$$\begin{aligned} -\nabla \cdot (\epsilon(\bar{x})\nabla\tilde{u}(\bar{x})) + \bar{\kappa}^2(\bar{x})\sinh(\tilde{u}(\bar{x})) &= 0, \\ [\tilde{u}(\bar{x})]_\Gamma = 0, \quad [\epsilon\nabla\tilde{u}(\bar{x})\cdot\mathbf{n}]_\Gamma &= -\epsilon_m\nabla(G(\bar{x}) + u^0(\bar{x}))\cdot\mathbf{n}|_\Gamma. \end{aligned} \tag{A.11}$$

Last, the solution decomposition in [12] is as follows: $u(\bar{x}) = G(\bar{x}) + u^r(\bar{x})$, where $G(\bar{x})$ is as in (A.6) and the regular part is given by

$$\begin{aligned} -\nabla\cdot(\epsilon\vec{\nabla}u^r) + \bar{\kappa}^2\sinh(u^r + G) &= \nabla\cdot((\epsilon-\epsilon_m)\vec{\nabla}G) \qquad \text{in } \Omega, \\ u^r &= g - G \qquad\qquad\qquad\quad \text{on } \partial\Omega. \end{aligned} \tag{A.12}$$

We can also further decompose (A.12) into the linear and nonlinear components so that $u^r(\bar{x}) = u^l(\bar{x}) + u^n(\bar{x})$, where

$$\begin{aligned} -\nabla\cdot(\epsilon\vec{\nabla}u^l) &= \nabla\cdot((\epsilon-\epsilon_m)\nabla G) \qquad \text{in } \Omega, \\ u^l &= 0 \qquad\qquad\qquad\quad \text{on } \partial\Omega, \end{aligned} \tag{A.13}$$

and

$$\begin{aligned} -\nabla\cdot(\epsilon\vec{\nabla}u^n) + \bar{\kappa}^2\sinh\left(u^n + u^l + G\right) &= 0 \qquad\quad \text{in } \Omega, \\ u^n &= g - G \quad \text{on } \partial\Omega. \end{aligned} \tag{A.14}$$

The fundamental idea in the above decomposition strategies is the pursuit of an efficient solution decomposition technique for the short- and long-range parts in a target tensor. However, all these techniques do not efficiently separate the long- and short-range components in each of the atomic volumes of the biomolecule. Rather, they split the Laplacian operator at the solute/solvent interface using the dielectric coefficient as a cutoff function, thereby creating the need to apply so-called interface (or jump) conditions at the interface in order to reduce the discontinuities.

REFERENCES

[1] N. A. Baker, *Poisson-Boltzmann methods for biomolecular electrostatics*, Methods Enzymology, 383 (2004), pp. 94–118.

[2] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, *Electrostatics of nanosystems: Application to microtubules and the ribosome*, Proc. Natl. Acad. Sci. USA, 98 (2001), pp. 10037–10041.

[3] V. Barone, M. Cossi, and J. Tomasi, *A new definition of cavities for the computation of solvation free energies by the polarizable continuum model*, J. Chem. Phys., 107 (1997), pp. 3210–3221.

[4] D. Bashford and D. A. Case, *Generalized born models of macromolecular solvation effects*, Annu. Rev. Phys. Chem., 51 (2000), pp. 129–152.

[5] P. Benner, L. Feng, C. Kweyu, and M. Stein, *Fast Solution of the Poisson-Boltzmann Equation with Nonaffine Parametrized Boundary Conditions Using the Reduced Basis Method*, arXiv:1705.08349, 2017.

[6] P. Benner, V. Khoromskaia, and B. N. Khoromskij, *Range-Separated Tensor Formats for Numerical Modeling of Many-Particle Interaction Potentials*, arXiv:1606.09218v3, 2016.

[7] P. Benner, V. Khoromskaia, and B. N. Khoromskij, *Range-separated tensor format for many-particle modeling*, SIAM J. Sci. Comput., 2 (2018), pp. A1034–A1062.

[8] C. Bertoglio and B. N. Khoromskij, *Low-rank quadrature-based tensor approximation of the Galerkin projected Newton/Yukawa kernels*, Comput. Phys. Commun., 183 (2012), pp. 904–912.
[9] H. Boschitsch and M. O. Fenley, *An adaptive Cartesian grid-based Poisson-Boltzmann solver: Energy and surface electrostatic properties*, in Computational Electrostatics for Biological Applications, W. Rocchia, M. Spagnuolo, eds., Springer, Cham, 2015, pp. 73–110.
[10] E. Cances, Y. Maday, and B. Stamm, *Domain decomposition for implicit solvation models*, J. Chem. Phys., 139 (2013), 054111.
[11] J. Chen and W. Geng, *On preconditioning the treecode-accelerated boundary integral (TABI) Poisson–Boltzmann solver*, J. Comput. Phys., 373 (2018), pp. 750–762.
[12] L. Chen, M. Holst, and J. Xu, *The finite element approximation of the nonlinear Poisson-Boltzmann equation*, SIAM J. Numer. Anal., 45 (2007), pp. 2298–2320.
[13] I.-L. Chern, J.-G. Liu, and W.-C. Wang, *Accurate evaluation of electrostatics for macromolecules in solution*, Methods Appl. Anal., 10 (2003), pp. 309–328.
[14] M. L. Connolly, *Solvent-accessible surfaces of proteins and nucleic acids*, Science, 221 (1983), pp. 709–713.
[15] M. E. Davis and A. McCammon, *Electrostatics in biomolecular structure and dynamics*, Chem. Rev., 90 (1990), pp. 509–521.
[16] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.
[17] M. Deserno and C. Holm, *How to mesh up Ewald sums.* I. *A theoretical and numerical comparison of various particle mesh routines*, J. Chem. Phys., 109 (1998), pp. 7678–7693.
[18] F. Fogolari, A. Brigo, and H. Molinari, *The Poisson-Boltzmann equation for biomolecular electrostatics: A tool for structural biology*, J. Mol. Recognit., 15 (2002), pp. 377–392.
[19] I. P. Gavrilyuk, W. Hackbusch, and B. N. Khoromskij, *Hierarchical tensor-product approximation to the inverse and related operators in high-dimensional elliptic problems*, Computing, 74 (2005), pp. 131–157.
[20] W. Geng and S. Zhao, *A two-component Matched Interface and Boundary (MIB) regularization for charge singularity in implicit solvation*, J. Comput. Phys., 351 (2017), pp. 25–39.
[21] M. K. Gilson, M. E. Davis, B. A. Luty, and J. A. McCammon, *Computation of electrostatic forces on solvated molecules using the Poisson-Boltzman equation*, J. Phys. Chem., 97 (993), pp. 3591–3600.
[22] W. Hackbusch and B. Khoromskij, *Low-rank Kronecker product approximation to multi-dimensional nonlocal operators. Part* I. *Separable approximation of multi-variate functions*, Computing, 76 (2006), pp. 177–202.
[23] R. C. Harris, T. Mackoy, and M. O. Fenley, *Problems of robustness in Poisson-Boltzmann binding free energies*, J. Chem. Theory Comput., 11 (2015), pp. 705–712.
[24] M. Holst, N. Baker, and F. Wang, *Adaptive multilevel finite element solution of the Poisson-Boltzmann equation: Algorithms and examples*, J. Comput. Chem., 21 (2000), pp. 1319–1342.
[25] M. Holst and F. Saied, *Multigrid solution of the Poisson-Boltzmann equation*, J. Comput. Chem., 14 (1993), pp. 105–113.
[26] M. J. Holst, *Multilevel Methods for the Poisson-Boltzmann Equation*. Ph.D. thesis, Numerical Computing Group, University of Illinois, Urbana-Champaign, 1994.
[27] B. Honig and A. Nicholls, *Classical electrostatics in biology and chemistry*, Sci. New Ser., 268 (1995), pp. 1144–1149.
[28] P. H. Hünenberger and J. A. McCammon, *Effect of artificial periodicity in simulations of biomolecules under Ewald boundary conditions: A continuum electrostatics study*, Biophys. Chemistry, 78 (1999), pp. 69–88.
[29] W. Im, D. Beglov, and B. Roux, *Continuum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation*, Comput. Phys. Commun., 111 (1998), pp. 59–75.
[30] J. D. Jackson, *Classical Electrodynamics*, 3rd ed., Wiley, New York, 1999.
[31] V. Khoromskaia and B. N. Khoromskij, *Grid-based lattice summation of electrostatic potentials by assembled rank-structured tensor approximation*, Comput. Phys. Commun., 185 (2014).
[32] V. Khoromskaia and B. N. Khoromskij, *Tensor Numerical Methods in Quantum Chemistry*, De Gruyter, Berlin, 2018.
[33] B. N. Khoromskij, *Structured rank-*$(r_1, \ldots, r_d)$ *decomposition of function-related operators in* $\mathbb{R}^d$, Comput. Methods Appl. Math., 6 (2006), pp. 194–220.
[34] B. N. Khoromskij, *Tensor Numerical Methods in Scientific Computing*, De Gruyter, Berlin, 2018.

[35] B. N. Khoromskij, *Range-separated tensor representation of the discretized multidimensional Dirac delta and elliptic operator inverse*, J. Comput. Phys., 401 (2020), 108998.
[36] B. N. Khoromskij and V. Khoromskaia, *Multigrid accelerated tensor approximation of function related multidimensional arrays*, SIAM J. Sci. Comput., 31 (2009), pp. 3002–3026.
[37] J. G. Kirkwood, *Theory of solutions of molecules containing widely separated charges with special application to zwitterions*, J. Chem. Phys., 2 (1934), pp. 351–361.
[38] C. Kweyu, L. Feng, M. Stein, and P. Benner, *Fast solution of the linearized Poisson–Boltzmann equation with nonaffine parametrized boundary conditions using the reduced basis method*, Comput. Vis. Sci., 23 (2020), pp. 1–19.
[39] C. Kweyu, M. Hess, L. Feng, M. Stein, and P. Benner, *Reduced basis method for Poisson-Boltzmann Equation*, in ECCOMAS Congress 2016, Volume 2, M. Papadrakakis, V. Papadopoulos, G. Stefanou, and V. Plevris, eds., Athens, 2016, pp. 4187–4195.
[40] C. M. Kweyu, *Fast Solution of the Poisson-Boltzmann Equation by the Reduced Basis Method and Range-Separated Canonical Tensor Format*, Ph.D. thesis, Otto-von-Guericke-Universität Magdeburg, Fakultät für Mathematik, 2019.
[41] M. Le Du, P. Marchot, P. Bougis, and J. Fontecilla-Camps, 1.9 *Angstrom resolution structure of fasciculine* 1*, an anti-acetylcholinesterase toxin from green mamba snake venom*, J. Biol. Chem., 267 (1992), pp. 22122–22130.
[42] E. Lindgren, A. J. A. J. Stace, Polack, E. Y. Maday, B. Stamm, and E. Besley, *An integral equation approach to calculate electrostatic interactions in many-body dielectric systems*, J. Comput. Phys., 371 (2018), pp. 712–731.
[43] F. Lipparini, B. Stamm, E. Cances, Y. Maday, and B. Mennucci, *Domain decomposition for implicit solvation models*, J. Chem. Theory. Comput., 9 (2013), pp. 3637–3648.
[44] B. Lu, X. Cheng, J. Huang, and J. A. McCammon, *An adaptive fast multipole boundary element method for Poisson-Boltzmann electrostatics*, J. Chem. Theory Comput., 5 (2009), pp. 1692–1699.
[45] B. Lu, D. Zhang, and A. McCammon, *Computation of electrostatic forces between solvated molecules determined by the Poisson-Boltzmann equation using a boundary element method*, J. Chem. Phys., 122 (2005), 214102.
[46] B. Z. Lu, Y. C. Zhou, M. J. Holst, and J. A. McCammon, *Recent progress in numerical methods for Poisson-Boltzmann equation in biophysical applications*, Commun. Comput. Phys., 3 (2008), pp. 973–1009.
[47] M. Mirzadeh, M. Theillard, A. Helgadottir, D. Boy, and F. Gibou, *An adaptive, finite difference solver for the nonlinear Poisson-Boltzmann equation with applications to biomolecular computations*, Commun. Comput. Phys., 13 (2013), pp. 150–173.
[48] Y. Notay, *An aggregation-based algebraic multigrid method*, Electron. Trans. Numer. Anal., 37 (2010), pp. 123–146.
[49] Y. Notay, *Aggregation-based algebraic multigrid for convection-diffusion equations*, SIAM J. Sci. Comput., 34 (2012), pp. A2288–A2316.
[50] E. L. Pollock and J. Glosli, *Comments on $P(3)m$, FMM and the Ewald method for large periodic Coulombic systems*, Comput. Phys. Commun., 95 (1996), pp. 93–110.
[51] P. Ren, J. Chun, D. G. Thomas, M. J. Schnieders, M. Marucho, J. Zhang, and N. A. Baker, *Biomolecular electrostatics and solvation: a computational perspective*, Quart. Rev. Biophys., 45 (2012), pp. 427–491.
[52] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis with Applications in the Chemical Sciences*, Wiley, New York, 2004.
[53] M. Stein, R. R. Gabdoulline, and R. C. Wade, *Cross-species analysis of the glycolitic pathway by comparison of molecular interaction fields*, Molecular Biosystems, 6 (2010), pp. 162–174.
[54] S. Vergara-Perez and M. Marucho, *MPBEC, a Matlab program for biomolecular electrostatic calculations*, Comput. Phys. Commun., 198 (2016), pp. 179–194.
[55] J. Wagoner and N. Baker, *Solvation forces on biomolecular structures: A comparison of explicit solvent and Poisson-Boltzmann models*, J. Comput. Chem., 25 (2004), pp. 1623–1629.
[56] C. Wang, J. Wang, Q. Cai, Z. Li, H. K. Zhao, and R. Luo, *Exploring accurate Poisson-Boltzmann methods for biomolecular simulations*, Comput. Theor. Chem., 1024 (2013), pp. 34–44.
[57] D. Xie, *New solution decomposition and minimization scheme for Poisson-Boltzmann equation in calculation of biomolecular electrostatics*, J. Comput. Phys., 275 (2014), pp. 294–309.
[58] D. Xie and J. Ying, *A new box iterative method for a class of nonlinear interface problems with application in solving Poisson-Boltzmann equation*, J. Comput. Appl. Math., 307 (2016), pp. 319–334.