The density-based many-body expansion for polypeptides and proteins

Johannes R. Vornweg[1], Toni M. Maier [2], Christoph R. Jacob[3]

Technische Universität Braunschweig, Institute of Physical and Theoretical Chemistry,
Gaußstraße 17, 38106 Braunschweig, Germany

Supporting Information

Date: April 1, 2025

[1]ORCID: 0000-0002-1457-0173
[2]ORCID: 0000-0001-5571-7576
[3]ORCID: 0000-0002-6227-8476; E-Mail: c.jacob@tu-braunschweig.de

S1 Explicit equations for density-based correction

S1.1 db-MBE

For arbitrary orders, the density-based energy correction $\Delta E^{(n)}_{\text{db-eb}}$ appearing in Eq. (10) is given by [cf. Eq. (11)],

$$\begin{aligned}\Delta E^{(n)}_{\text{db-eb}}&\Big[\{\rho_{i_1},\rho_{i_1i_2},\dots\}\Big]\\ &= \Big(V_{\text{nuc}}[\rho^{(n)}_{\text{tot}}]-V^{(n)}_{\text{nuc}}\Big)+\Big(J[\rho^{(n)}_{\text{tot}}]-J^{(n)}\Big)+\Big(E_{\text{NN}}-E^{(n)}_{\text{NN}}\Big)\\ &\quad+T_{\text{s}}^{\text{nadd},(n)}\big[\rho_{i_1},\rho_{i_1i_2},\dots\big]+E_{\text{xc}}^{\text{nadd},(n)}\big[\rho_{i_1},\rho_{i_1i_2},\dots\big],\end{aligned}\tag{S1}$$

with the n-body nonadditive kinetic and exchange-correlation energy functionals defined as

$$T_{\text{s}}^{\text{nadd},(n)}\Big[\{\rho_{i_1},\rho_{i_1i_2},\dots\}\Big]=T_{\text{s}}\big[\rho^{(n)}_{\text{tot}}\big]-T^{(n)}_{\text{s}}\tag{S2}$$

$$E_{\text{xc}}^{\text{nadd},(n)}\Big[\{\rho_{i_1},\rho_{i_1i_2},\dots\}\Big]=E_{\text{xc}}\big[\rho^{(n)}_{\text{tot}}\big]-E^{(n)}_{\text{xc}}.\tag{S3}$$

Here, $V^{(n)}_{\text{nuc}}$, $J^{(n)}$, $E^{(n)}_{\text{NN}}$, $T^{(n)}_{\text{s}}$, and $E^{(n)}_{\text{xc}}$ are the MBEs of the individual contributions to the KS-DFT total energy functional, which are defined in analogy to Eq. (1), i.e.,

$$X^{(n)}=X^{(1)}+\sum_{m=2}^{n}\Delta X^{(m)},\tag{S4}$$

with

$$\Delta X^{(n)}=\sum_{i_1=1}^{N}\sum_{i_2=i_1+1}^{N}\cdots\sum_{i_n=i_{n-1}+1}^{N}X_{i_1i_2\dots i_n}-\sum_{m=1}^{n-1}\frac{N^{n-m}}{(n-m)!}\cdot\Delta X^{(m)},\tag{S5}$$

Note that a correction due to the nuclear repulsion energy only appears at first order, because $E^{(n)}_{\text{NN}}=E_{\text{NN}}$ for $n\geq 2$.

At first order, the density-based correction is given by

$$\begin{aligned}
&\Delta E^{(1)}_{\text{db-eb}}\Big[\{\rho_{i_1}\}_{i_1=1,\dots,N}\Big] \\
&\quad= \sum_{i_1=1}^{N}\sum_{j_1=i_1+1}^{N} E_{\text{NN}}^{(i_1 j_1)} + \sum_{i_1=1}^{N}\sum_{\substack{j_1=1\\ j_1\neq i_1}}^{N}\int \rho_{i_1}(\boldsymbol{r}) v_{\text{nuc}}^{(j_1)}(\boldsymbol{r})\,\mathrm{d}^3 r \\
&\qquad+ \sum_{i_1=1}^{N}\sum_{j_1=i_1+1}^{N}\iint \frac{\rho_{i_1}(\boldsymbol{r})\rho_{j_1}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}^3 r \mathrm{d}^3 r' \\
&\qquad+ E_{\text{xc}}^{\text{nadd}}\Big[\{\rho_{i_1}\}_{i_1=1,\dots,N}\Big] + T_{\text{s}}^{\text{nadd}}\Big[\{\rho_{i_1}\}_{i_1=1,\dots,N}\Big],
\end{aligned} \tag{S6}$$

which is the interaction energy expression also appearing in subsystem DFT [1].

At second order, the density-based correction is given by

$$\begin{aligned}
&\Delta E^{(2)}_{\text{db-eb}}\Big[\{\rho_{i_1}\}_{i=1,\dots,N}, \{\rho_{i_1 i_2}\}_{i_1=1,\dots,N;\, i_2=i_1,\dots,N}\Big] \\
&\quad= \Big(V_{\text{nuc}}[\rho^{(2)}_{\text{tot}}] - V^{(2)}_{\text{nuc}}\Big) + \Big(J[\rho^{(2)}_{\text{tot}}] - J^{(2)}\Big) \\
&\qquad+ E_{\text{xc}}^{\text{nadd},(2)}\Big[\{\rho_{i_1}\}_{i=1,\dots,N}, \{\rho_{i_1 i_2}\}_{i_1=1,\dots,N;\, i_2=i_1,\dots,N}\Big] \\
&\qquad+ T_{\text{s}}^{\text{nadd},(2)}\Big[\{\rho_{i_1}\}_{i=1,\dots,N}, \{\rho_{i_1 i_2}\}_{i_1=1,\dots,N;\, i_2=i_1,\dots,N}\Big].
\end{aligned} \tag{S7}$$

Here, the electron–nuclei interaction contribution is calculated as

$$\begin{aligned}
V_{\text{nuc}}[\rho^{(2)}_{\text{tot}}] - V^{(2)}_{\text{nuc}} = &\int \rho^{(1)}_{\text{tot}}(\boldsymbol{r})\, v^{\text{tot}}_{\text{nuc}}(\boldsymbol{r})\,\mathrm{d}^3 r + \int \Delta\rho^{(2)}_{\text{tot}}(\boldsymbol{r})\, v^{\text{tot}}_{\text{nuc}}(\boldsymbol{r})\,\mathrm{d}^3 r \\
&- \Bigg(\sum_{i_1=1}^{N}\int \rho_{i_1}(\boldsymbol{r})\, v^{(i_1)}_{\text{nuc}}(\boldsymbol{r})\,\mathrm{d}^3 r \\
&\qquad+ \sum_{i_1=1}^{N}\sum_{i_2=i_1+1}^{N}\Bigg[\int \rho_{i_1 i_2}(\boldsymbol{r})\, v^{(i_1 i_2)}_{\text{nuc}}(\boldsymbol{r})\,\mathrm{d}^3 r \\
&\qquad\qquad- \int \rho_{i_1}(\boldsymbol{r})\, v^{(i_1)}_{\text{nuc}}(\boldsymbol{r})\,\mathrm{d}^3 r - \int \rho_{i_2}(\boldsymbol{r})\, v^{(i_2)}_{\text{nuc}}(\boldsymbol{r})\,\mathrm{d}^3 r\Bigg]\Bigg)
\end{aligned} \tag{S8}$$

and the Coulomb interaction is calculated as

$$
\begin{aligned}
J[\rho_{\mathrm{tot}}^{(2)}] - J^{(2)} = \frac{1}{2} \iint \frac{\rho_{\mathrm{tot}}^{(1)}(\boldsymbol{r}) \rho_{\mathrm{tot}}^{(1)}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}^3 r \mathrm{d}^3 r' \\
+ \iint \frac{\rho_{\mathrm{tot}}^{(1)}(\boldsymbol{r}) \Delta\rho_{\mathrm{tot}}^{(2)}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}^3 r \mathrm{d}^3 r' + \frac{1}{2} \iint \frac{\Delta\rho_{\mathrm{tot}}^{(2)}(\boldsymbol{r}) \Delta\rho_{\mathrm{tot}}^{(2)}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}^3 r \mathrm{d}^3 r'. \\
- \Bigg(\sum_{i_1=1}^{N} \frac{1}{2} \iint \frac{\rho_{i_1}(\boldsymbol{r}) \rho_{i_1}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}^3 r \mathrm{d}^3 r' \\
+ \sum_{i_1=1}^{N} \sum_{i_2=i_1+1}^{N} \Bigg[\frac{1}{2} \iint \frac{\rho_{i_1 i_2}(\boldsymbol{r}) \rho_{i_1 i_2}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}^3 r \mathrm{d}^3 r' \\
- \frac{1}{2} \iint \frac{\rho_{i_1}(\boldsymbol{r}) \rho_{i_1}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}^3 r \mathrm{d}^3 r' \\
- \frac{1}{2} \iint \frac{\rho_{i_2}(\boldsymbol{r}) \rho_{i_2}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}^3 r \mathrm{d}^3 r' \Bigg] \Bigg)
\end{aligned}
\tag{S9}
$$

and the second-order nonadditive kinetic and exchange–correlation functionals are given by

$$
\begin{aligned}
&X^{\mathrm{nadd},(2)} \Big[\{\rho_{i_1}\}_{i=1,\ldots,N}, \{\rho_{i_1 i_2}\}_{i_1=1,\ldots,N;\, i_2=i_1,\ldots,N} \Big] \\
&\quad = X\big[\rho_{\mathrm{tot}}^{(2)}\big] - \Bigg(\sum_{i_1=1}^{N} \sum_{i_2=i_1+1}^{N} X\big[\rho_{i_1 i_2}\big] - (N-1) \sum_{i_1=1}^{N} X\big[\rho_{i_1}\big] \Bigg).
\end{aligned}
\tag{S10}
$$

S1.2 db-MFCC

For db-MFCC and eb-MFCC-MBE(2), the density-based correction can be evaluated using Eq. (S1), where for the many-body expansions of the individual energy terms, the more general definition including the contributions of the cap is used.

At first order (i.e., for db-MFCC), the density-based correction is given by

$$
\begin{aligned}
&\Delta E^{(1)}_{\text{db-eb}}\Big[\{\rho^{\text{f}}_{i_1}\}_{i_1=1,\dots,N},\{\rho^{\text{c}}_{[k_1,k_1+1]}\}_{k_1=1,\dots,N-1}\Big] \\
&\quad= E^{\text{tot}}_{\text{NN}} - \sum_{i_1=1}^{N}\sum_{j_1=i_1+1}^{N} E^{\text{f},(i_1)}_{\text{NN}} + \sum_{k_1=1}^{N}\sum_{k_1=1}^{N-1} E^{\text{c},(k_1)}_{\text{NN}} \\
&\qquad+ \sum_{i_1=1}^{N}\sum_{\substack{j_1=1\\ j_1\neq i_1}}^{N}\int \rho^{\text{f}}_{i_1}(\boldsymbol{r})v^{\text{f},(j_1)}_{\text{nuc}}(\boldsymbol{r})\,\mathrm{d}^3r - \sum_{i_1=1}^{N}\sum_{k_1}^{N-1}\int \rho^{\text{f}}_{i_1}(\boldsymbol{r})v^{\text{c},[k_1,k_1+1])}_{\text{nuc}}(\boldsymbol{r})\,\mathrm{d}^3r \\
&\qquad\quad+ \sum_{k_1=1}^{N-1}\sum_{\substack{l_1=1\\ l_1\neq k_1}}^{N-1}\int \rho^{\text{c}}_{[k_1,k_1+1]}(\boldsymbol{r})v^{\text{c},([l_1,l_1+1])}_{\text{nuc}}(\boldsymbol{r})\,\mathrm{d}^3r \\
&\qquad+ \sum_{i_1=1}^{N}\sum_{j_1=i_1+1}^{N}\iint \frac{\rho^{\text{f}}_{i_1}(\boldsymbol{r})\rho^{\text{f}}_{j_1}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}^3r\mathrm{d}^3r' - \sum_{i_1=1}^{N}\sum_{k_1=1}^{N-1}\iint \frac{\rho^{\text{f}}_{i_1}(\boldsymbol{r})\rho^{\text{c}}_{[k_1,k_1+1]}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}^3r\mathrm{d}^3r' \\
&\qquad\quad+ \sum_{k_1=1}^{N-1}\sum_{l_1=k+1}^{N-1}\iint \frac{\rho^{\text{c}}_{[k_1,k_1+1]}(\boldsymbol{r})\rho^{\text{c}}_{[l_1,l_1+1]}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}^3r\mathrm{d}^3r' \\
&\qquad+ E^{\text{nadd}}_{\text{xc}}\Big[\{\rho^{\text{f}}_{i_1}\}_{i_1=1,\dots,N},\{\rho^{\text{c}}_{[k_1,k_1+1]}\}_{k_1=1,\dots,N-1}\Big] \\
&\qquad+ T^{\text{nadd}}_{\text{s}}\Big[\{\rho^{\text{f}}_{i_1}\}_{i_1=1,\dots,N},\{\rho^{\text{c}}_{[k_1,k_1+1]}\}_{k_1=1,\dots,N-1}\Big].
\end{aligned}
\tag{S11}
$$

The nonadditive exchange–correlation and kinetic energy functionals are now given by

$$
\begin{aligned}
&X^{\text{nadd}}\Big[\{\rho^{\text{f}}_{i_1}\}_{i_1=1,\dots,N},\{\rho^{\text{c}}_{[k_1,k_1+1]}\}_{k_1=1,\dots,N-1}\Big] \\
&\quad= X\big[\rho_{\text{tot}}\big] - \sum_{i_1=1}^{N} X\big[\rho^{\text{f}}_{i_1}\big] + \sum_{k_1=1}^{N-1} X\big[\rho^{\text{c}}_{[k_1,k_1+1]}\big]
\end{aligned}
\tag{S12}
$$

S1.3 db-MFCC-MBE(2)

At second order, the density-based correction is given by

$$\begin{aligned}
&\Delta E^{(2)}_{\text{db-eb}}\Big[\{\rho^{\text{f}}_{i_1}\},\{\rho^{\text{c}}_{[k_1,k_1+1]}\},\{\rho^{\text{ff}}_{i_1 i_2}\},\{\rho^{\text{fc}}_{i_1[k_1,k_1+1]}\},\{\rho^{\text{cc}}_{[k_1,k_1+1],[k_2,k_2+1]}\}\Big] \\
&\quad= \Big(V_{\text{nuc}}[\rho^{(2)}_{\text{tot}}] - V^{(2)}_{\text{nuc}}\Big) + \Big(J[\rho^{(2)}_{\text{tot}}] - J^{(2)}\Big) \\
&\qquad + E^{\text{nadd},(2)}_{\text{xc}}\Big[\{\rho^{\text{f}}_{i_1}\},\{\rho^{\text{c}}_{[k_1,k_1+1]}\},\{\rho^{\text{ff}}_{i_1 i_2}\},\{\rho^{\text{fc}}_{i_1[k_1,k_1+1]}\},\{\rho^{\text{cc}}_{[k_1,k_1+1],[k_2,k_2+1]}\}\Big] \\
&\qquad + T^{\text{nadd},(2)}_{\text{s}}\Big[\{\rho^{\text{f}}_{i_1}\},\{\rho^{\text{c}}_{[k_1,k_1+1]}\},\{\rho^{\text{ff}}_{i_1 i_2}\},\{\rho^{\text{fc}}_{i_1[k_1,k_1+1]}\},\{\rho^{\text{cc}}_{[k_1,k_1+1],[k_2,k_2+1]}\}\Big].
\end{aligned} \tag{S13}$$

All terms in this expression can be evaluated in full analogy to those discussed above for the case of the db-MBE, with the difference that in the many-body expansions of the individual energy terms, the more general expansions including fragments and cap molecules as well as fragment–fragment, fragment–cap, and cap–cap interactions need to be accounted for.

A general implementation that covers both the conventional db-MBE case and the extension to the MFCC expansion can be achieved by assigning a prefactor of +1 to all fragments and a prefactor of −1 to all cap molecules, and carrying these prefactors through all the expressions.

S2 Plot of total absolute errors for protein test set

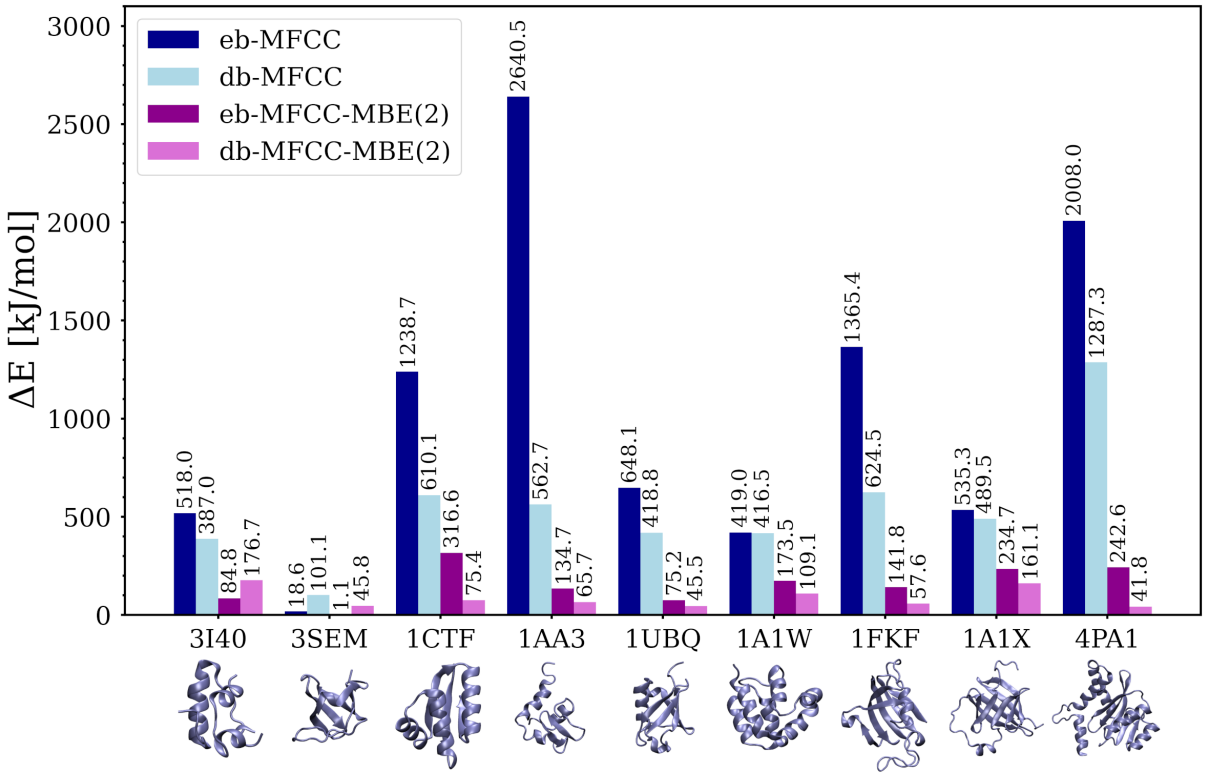

Figure S1: Comparison of the total absolute errors in the total energy (DFT/BP86/DZP) with the energy-based and the density-based MFCC and MFCC-MBE(2) schemes for our test set of small proteins. As reference, single-point calculations for the full proteins have been performed.

References

[1] Ch. R. Jacob and J. Neugebauer, *Subsystem density-functional theory*, WIREs Comput. Mol. Sci. **4**, 325 (2014).