RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2026, **16**, 11361

Proton-driven many–body interactions and structural organization in He_nH^+ clusters

María Judit Montes de Oca-Estévez, [a] Javier Hernández-Rojas *[b] and Rita Prosmiti *[a]

We present a systematic description of the interactions governing He_nH^+ clusters, derived from gold-standard *ab initio* data. Two-, three-, and four-body potential contributions generated using CCSD(T)/CBS calculations and a kernel-based machine learning approach are integrated within a many-body expansion formalism. Benchmark tests show that inclusion of up to four-body terms is required to achieve consistent accuracy for larger clusters. Using the validated potential, global minimum searches reveal that structural organization proceeds through sequential binding of He atoms to a rigid linear HeH^+He core, with formation of well-defined solvation motifs, while the more weakly bound He atoms display increasing degrees of delocalization. Zero-point-corrected energetics reproduce the experimentally reported stability patterns for $n \leq 13$ clusters, while the onset of pronounced He delocalization, associated with microscopic superfluid-like behavior, introduces nontrivial quantum effects that influence stability trends for larger clusters. From a computational perspective, these results establish the central importance of higher-order many-body effects and quantum contributions in accurately describing proton microsolvation in He, while providing a reliable and comprehensive framework for interpreting experimental stability trends in proton-bound noble-gas clusters.

Received 21st January 2026
Accepted 18th February 2026

DOI: 10.1039/d6ra00552g

rsc.li/rsc-advances

1 Introduction

Hydrogen and helium constitute the vast majority of observable matter in the universe,[1,2] yet their chemical behaviors differ markedly. While hydrogen is a key component of most molecular species, helium was long regarded as chemically inert due to its high ionization potential and extremely low polarizability. The astronomical detection of ArH^+ (ref. 3 and 4) and HeH^+ (ref. 5) has challenged this view, demonstrating that noble gases can form stable ions under suitable conditions and motivating renewed interest in noble-gas chemistry.[6–18] Helium-containing complexes are particularly relevant because of helium's cosmological abundance and its role as a weakly interacting quantum solvent.

Among these systems, proton-bound complexes of the form $He_nH_m^+$(n, $m \geq 1$) provide a unique platform for studying microsolvation of H^+, H_2^+, and H_3^+ – species of central importance in the interstellar medium (ISM).[19–22] Experimentally, their characterization has relied primarily on high-resolution mass spectrometry. Early studies by Kojima *et al.*[23] identified He_nH^+ (up to $n = 14$) and $He_nH_3^+$ (up to $n = 13$), revealing distinct magic numbers at $n = 13$ and $n = 10$–11, respectively, while more recent ultrahigh-resolution spectroscopy measurements extended these data up to $n = 30$, quantifying binding energies and uncovering $He_nH_2^+$ clusters for the first time.[24] Additional studies reported enhanced stabilities at $n = 2$, 6, 11, and 13,[25,26] and identified signatures of microscopic superfluidity in the intermediate-size clusters.[25] Moreover, cryogenic IR spectroscopy has provided structural fingerprints for small complexes ($n = 3$–6).[27]

From the theoretical side, the strongly bound and highly symmetric He_2H^+ molecular ion is known to act as the structural core of He_nH^+ clusters, as is the case for the Ar_nH^+ clusters.[14,28] Several attempts have been made to calculate the equilibrium structures and relative energies of He_nH^+ for $n \geq 2$.[6,17,29–39] As n increases, additional He atoms interact only through very weak dispersive forces, and previous studies have reported multiple low-symmetry minimum-energy structures.[25,29,30,34] However, the observed progressive energetic decrease in the calculated evaporation energies between $n = 3$–6, suggests that He_6H^+ has a remarkable stability. The latest study[25] has also incorporated experimental information obtained *via* cryogenic ion-trap mass spectrometry up to $n = 7$, which provides complementary evidence on stability patterns and supports the theoretical characterization.

Despite decades of experimental and theoretical work, a quantitative and transferable description of the interactions governing He_nH^+ clusters remains a central challenge. Accurate modeling requires treating subtle many-body effects and long–range interactions, along with the sizable quantum

[a]*Institute of Fundamental Physics, CSIC (IFF-CSIC), Serrano 123, 28006 Madrid, Spain. E-mail: rita@iff.csic.es*

[b]*Departamento de Física and IUdEA, Universidad de La Laguna, 38200 Tenerife, Spain*

contributions arising from the light and weakly bound He atoms. Existing potential energy surfaces (PESs) provide valuable insights for small complexes,[10,11,15] but they are not designed to scale reliably to larger cluster sizes. Moreover, recent analyses of related protonated noble-gas clusters[14] show that interaction terms up to the four-body (4B) level are necessary to capture the energetics with quantitative fidelity.

The objective of this work is to establish a systematic and size-consistent description of the interactions in He_nH^+ clusters. To this end, we develop a high-accuracy potential energy surface for the He_3H^+ tetramer based on single and double excitation coupled cluster with perturbative triples at complete basis set limit (CCSD(T)/CBS) calculations and a reproducing kernel Hilbert space (RKHS) representation. This PES is then integrated within a many-body expansion (2B, 3B, and 4B) to build scalable interaction models applicable to clusters up to $n = 34$. Extensive benchmark calculations against additional CCSD(T)/CBS reference data demonstrate that accurate modeling requires inclusion of 4B terms, whereas lower-order truncations lead to substantial errors. Finally, the validated 4B potential is employed to characterize the global minima, microsolvation motifs, and size-dependent stabilities of He_nH^+ clusters. Our results reproduce known magic numbers and predict new stability patterns at larger sizes, offering a comprehensive framework for understanding proton microsolvation in He.

2 Computational details

2.1 Representation of intermolecular interactions

To model the complex interactions in He_nH^+ clusters, we employed the many-body expansion (MBE) formalism.[40] This method provides a systematic general decomposition of the total energy ($E_\mathcal{N}$) of a system with $\mathcal{N}$-bodies or monomers into 1-body (1B), 2-body (2B), …, $\mathcal{N}$-body ($\mathcal{N}$B) contributions,

$$E_\mathrm{N} = \sum_\mathrm{N} E_i^{(1\mathrm{B})} + \sum_\mathrm{N} E_{ij}^{2\mathrm{B}} + \sum_\mathrm{N} E_{ijk}^{3\mathrm{B}} + \cdots + E^{(\mathcal{N}\mathrm{B})} \quad (1)$$

Here, $E_i^{1\mathrm{B}}$ is the monomer (1B) energy, $E_{ij}^{2\mathrm{B}}$ the two-body (2B) interactions between monomers i and j, $E_{ijk}^{(3\mathrm{B})}$ the three-body (3B) interaction among i, j and k, and so on up to the $\mathcal{N}$-body ($\mathcal{N}$B) term. Each m-body contribution isolates the interaction arising from groups of m monomers, enabling a controlled, hierarchical description of many-body effects.

In practice, the MBE is truncated at a finite order, and in this work, we evaluate the predictive performance of the analytical models V truncated at the 2B, 3B and 4B terms for energetics of He_nH^+ clusters, retaining accordingly up to:

$$\mathrm{V}_{2\mathrm{B}}\left(\overrightarrow{R_i}, \overrightarrow{R_{il}}\right) = \sum_{i=1}^{n} \mathrm{V}_{\mathrm{HeH}^+}\left(\overrightarrow{R_i}\right) + \sum_{i,l=1,l>i}^{n} \mathrm{V}_{\mathrm{HeHe}}\left(\overrightarrow{R_{il}}\right) \quad (2)$$

$$\mathrm{V}_{3\mathrm{B}}\left(\overrightarrow{R_i}, \overrightarrow{R_{il}}, r\right) = \sum_{i=1}^{n} \mathrm{V}_{\mathrm{He_2H}^+}\left(\overrightarrow{R_i}, r\right) + \sum_{i,l=1,l>i}^{n} \mathrm{V}_{\mathrm{HeHe}}\left(\overrightarrow{R_{il}}\right) \quad (3)$$

$$\mathrm{V}_{4\mathrm{B}}\left(\overrightarrow{R_i}, \overrightarrow{R_{il}}, R_{12}\right) = \sum_{i=1}^{n} \mathrm{V}_{\mathrm{He_3H}^+}\left(\overrightarrow{R_i}, R_{12}\right) + \sum_{i,l=1,l>i}^{n} \mathrm{V}_{\mathrm{HeHe}}\left(\overrightarrow{R_{il}}\right) \quad (4)$$

This decomposition simplifies the analysis of complex cluster, while systematically incorporating higher-order interactions. Here, $\vec{R}_i$ denotes the vector connecting the proton to the i-th He atom (with angular coordinates θ_i, ϕ_i), while r and $R_{12} = r_\mathrm{e}$ represent the HeH^+ and HeH^+He bond vectors in eqn (3) and (4), respectively (see Fig. 1), R_{il} vectors connect the i-th and l-th He atoms in each He_nH^+ cluster. In this sense, the $\mathrm{V}_{\mathrm{HeH}^+}(R)$, $V_{\mathrm{He_2H}^+}(R, r)$ and $V_{\mathrm{He_3H}^+}(R, r_\mathrm{e})$ are the corresponding 2B, 3B and 4B CCSD(T)/CBS potentials for the HeHe,[41] He_2H^+,[10] and He_3H^+ systems, respectively, as developed in the present work and employed in the MBE approach.

2.2 Electronic structure computations and kernel representation

For the construction of the 4B He_3H^+ PES, *ab initio* electronic structure calculations were performed using the MOLPRO 2022 program,[42] while the DENEB software[43] package was employed to generate and organize all input and output data files, respectively. We carried out CCSD(T) calculations employing the AV5Z and AV6Z basis sets.[44,45] In turn, correlation energies were extrapolated at the CBS limit utilizing the two-point single inverse power function first introduced by Schwartz,[46] $E_n^{\mathrm{corr}} = E_{\mathrm{CBS}}^{\mathrm{corr}} + \frac{A}{n^3}$, with $n = 5$ and 6, in order to obtain the CCSD(T)/CBS[56] energies, $E_{\mathrm{CBS}} = E_{\mathrm{HF/AV6Z}} + E_{\mathrm{CBS}}^{\mathrm{corr}}$.

The configuration space is sampled using the R and θ coordinates with $R = 1.30$–8.0 Å and $\theta = 0$–$90°$ for $r_\mathrm{e} = 1.8496$ Å, the

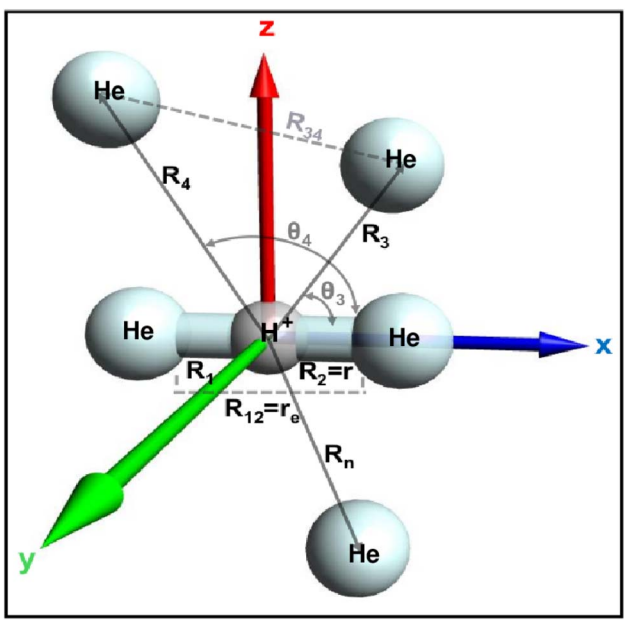

Fig. 1 Coordinate system used for the He_nH^+ clusters in the V_{2B}, V_{3B}, and V_{4B} potential approaches.

equilibrium $He–H^+–He$ bondlength from the CCSD(T)/AV6Z geometry optimizations. In total, 1900 grid points were considered, with the interaction energies defined as $\Delta E = E_{He_3H^+} - E_{He_2H^+} - E_{He}$. The energy of the total dissociation $He_3H^+ \rightarrow He + He + He + H^+$ corresponds to zero (see Table S1 in the SI), whereas $E_{He_3H^+}$,$E_{He_2H^+}$, and E_{He} are the CCSD(T)/CBS[56] energies of He_3H^+, He_2H^+, and He at each configuration, respectively. The basis set superposition error (BSSE)[47] is rendered negligible upon extrapolation of the interaction energies to the complete basis set (CBS) limit.

The ML-PES was constructed using the reproducing kernel Hilbert space (RKHS) approach,[48] in which the potential energy function $V(x)$ is interpolated from a set of known *ab initio* energies V_i corresponding to molecular configurations x_i. By applying the representer theorem, a general functional relationship can be written as a linear combination of kernel functions, $\tilde{V}(x) \approx \tilde{V}_i = \sum_{i=1}^{\mathcal{N}} C_i K(x, x_i)$, where C_i are expansion coefficients and $K(x, x')$ is the reproducing kernel. The coefficients are obtained by solving the linear system $\mathbf{C} = \mathbf{K}^{-1}\mathbf{y}$, with $\mathbf{y} = (V_1, V_2, \ldots, V_N)$ the vector of input energies, $\mathbf{C} = (C_1, \ldots, C_N)$ the coefficient vector, and $\mathbf{K}$ the $\mathcal{N} \times \mathcal{N}$ kernel matrix.

For multidimensional systems, the full kernel is constructed as a direct product of one-dimensional (1D) kernels, $K(x, x') = \prod_{d=1}^{D} k_d(x_d, x_d')$, where D is the dimensionality of the coordinate space and k_d are the corresponding 1D kernels. Several RKHS kernels have been proposed for different applications; in this work we employ the $k_1^{n,m}$ kernel for the distance-like coordinates R, and the k_2 kernel for the angle-like coordinates θ. The angular coordinate is expressed through its reduced form $z = \cos\theta$. The grid contains N_R points along R and N_θ points along θ. The distance-like reproducing kernel $k_1^{n',m'}$ is given by $k_1^{n',m'}(\chi, \chi') = n'^2 \chi_>^{-(m'+1)} B(m'+1, n')_2F_1(-n'+1, m'+1; n'+m'+1; \chi_</\chi_>)$, where χ, $\chi' \equiv R$, $\chi_>$ and $\chi_<$ denote the larger and smaller of (χ, χ'), B is the beta function, and ${}^2F_1$ is the Gaussian hypergeometric function. The parameters n' and m' control the smoothness and asymptotic decay of the kernel; here we use $n' = 2$ and $m' = 3$, corresponding to the leading R^{-4} dispersion interaction between He and the $He–H^+–He$ molecular ion. The angle-like kernel is expanded in terms of Legendre polynomials $k_2(z, z') = \sum_\ell \frac{2\ell+1}{2} P_\ell(z) P_\ell(z')$, where P_ℓ denotes the Legendre polynomial of order $\ell = 0 - 18$.

The final expression of the ML-PES is given as

$$V(r_e; R, \theta) = \sum_{i=1}^{N_R} \sum_{j=1}^{N_R} C_{ij} k_1^{n',m'}(R_i, R) k_2(\theta_j, \theta), \qquad (5)$$

with the coefficients C_{ij} obtained by solving the corresponding linear equations.

3 Results and discussion

3.1 Building up a ML-PES for the He_3H^+

The CCSD(T)/CBS[56] dataset was split into 760 points for training and 1140 points for testing and validation. The training set was arranged on a two-dimensional grid consisting of 40 equidistant points along the radial coordinate R (1.30–8.00 Å) and 19 points along the angular coordinate θ (0–90°).

The best-performed RKHS PES model was identified using a hold-out cross-validation procedure applied to six different training-grid datasets of 200 (20 × 10), 380 (20 × 19), 300 (30 × 10), 570 (30 × 19), 400 (40 × 10), and 760 (40 × 19) points. These configurations varied systematically in the number of radial (20, 30, or 40) and angular (10 or 19) sampling points, thereby providing comprehensive coverage of the two-dimensional coordinate space.

In the left panel of Fig. 2, the energy distribution of the training energies data indicates that approximately 90% of the computed points lie within the first dissociation channel of the system (see Table S1 in the SI), demonstrating that the most physically relevant regions of the PES are well represented in the training set.

In turn, in the right panel of Fig. 2 shows a systematic improvement in the RKHS ML-PES accuracy as the size of the

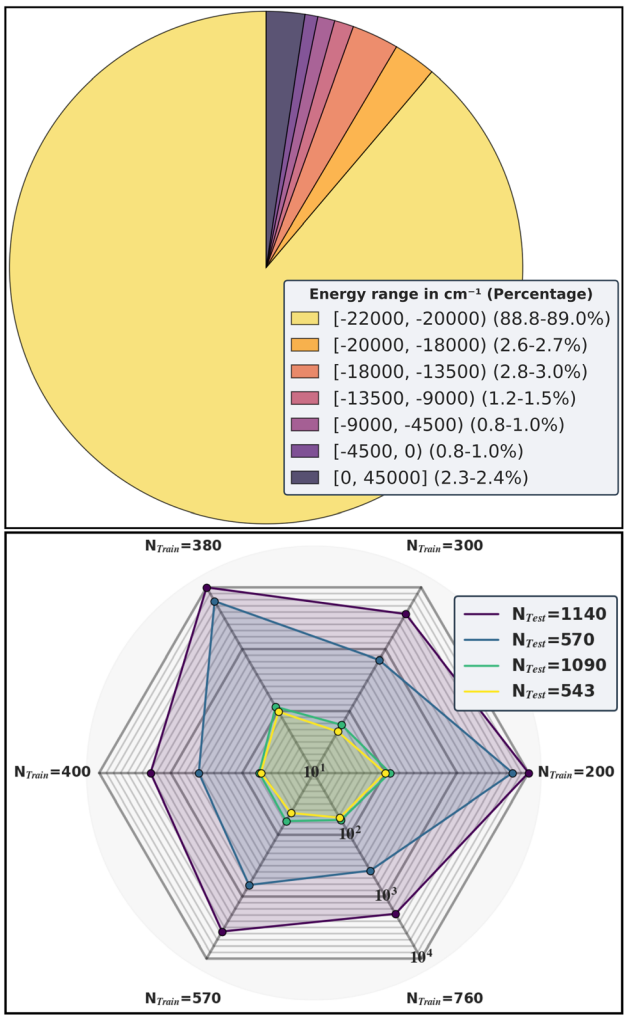

Fig. 2 Chart plots presenting the energy distribution of all configurations included in the training sets, ranging from −22 000 to 45 000 cm^{-1} values for the He_3H^+ (upper panel), and the corresponding RMSE values of the six ML-PES models as a function of training set size (lower panel).

training dataset increases. The root-mean-square-error (RMSE) values were evaluated over several independent testing sets, including the full 1140-point validation grid, the 1090-points below the dissociation limit, and two randomly selected subsets of 570 and 543 points, and all exhibit the same monotonic error reduction. This trend is clearly observed across the 200/380-, 300/570-, and 400/760-point comparisons, highlighting the direct correlation between model performance and the amount of training data.

A closer inspection of the radial sampling reveals that increasing the number of R points from 20 to 40 leads to a substantial and consistent decrease in RMSE, with values ranging from 56.66 to 34.78 cm^{-1} for energies up to the total dissociation. Notably, the errors remain essentially unchanged when comparing sparsely sampled angular grids ($200 = 20 \times 10$, $300 = 30 \times 10$, $400 = 40 \times 10$) with their more densely sampled counterparts ($380 = 20 \times 19$, $570 = 30 \times 19$, $760 = 40 \times 19$). This confirms that the θ coordinate plays a minor role in the overall accuracy, whereas the number of radial points is the dominant factor controlling the PES quality. For these reasons, we have selected the 40×10 training model for the construction of the He_3H^+ PES.

The correlation plot in Fig. 3 illustrates the performance of the selected RKHS ML-PES model against both the training set (400 configurations) and the full testing set (1140 configurations), covering the attractive and repulsive regions of the potential. The corresponding RMSE and mean-absolute-error (MAE) values for different energy intervals are also plotted in the figure. The similar RMSE and MAE values indicates that the RKHS model does not produce sporadic large errors and that its predictive performance remains uniform across configurations of different energies. Importantly, the model preserves a controlled error behavior outside the training region, maintaining deviations of approximately 200 cm^{-1} for the 1090 configurations lying below the dissociation threshold, and errors up to 3% for regions above the dissociation limit.

Moreover, the upper panel of Fig. 4 compares the RKHS ML-PES (solid lines) with the reference CCSD(T)/CBS[56] energies (circle symbols) as a function of the R and θ coordinates (see inset plot). The potential energy curves are shown for θ ranging from 0 to 90° in increments of 5°. In addition, the full 2D RKHS ML potential for the He_3H^+ cation is also displayed as a contour plot in the lower panel of Fig. 4, in the (θ, R) plane over the energy range −21 450 to −21 050 cm^{-1}. The presence of well-defined minima on the He_3H^+ PES is clearly evident. The global minimum corresponds to a T-shaped configuration with a well depth of −21 427.60 cm^{-1} for both the RKHS ML-PES and the CCSD(T)/CBS reference data. This minimum is located at R = 2.16 Å, which is only 0.15 Å larger than the equilibrium distance predicted by the CCSD(T)/AV6Z geometry optimization. The excellent agreement between the *ab initio* data and the present RKHS ML-PES demonstrates that the RKHS model provides an accurate and smooth representation of this region of the potential energy surface.

3.2 Computational validation of sum-of-potentials modeling approaches

To assess the accuracy of the three many-body (V_{2B}, V_{3B}, and V_{4B}) expansion schemes employed in this work, we performed additional CCSD(T)/CBS[56] calculations for the penta-atomic

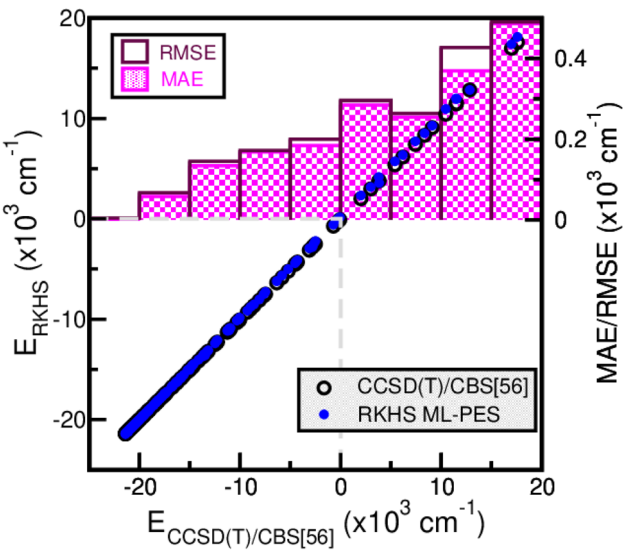

Fig. 3 Correlation plot of the best-performed RKHS ML-PES model against the reference CCSD(T)/CBS[56] energies for testing data, and the corresponding average RMSE and MAE values as a function of energy.

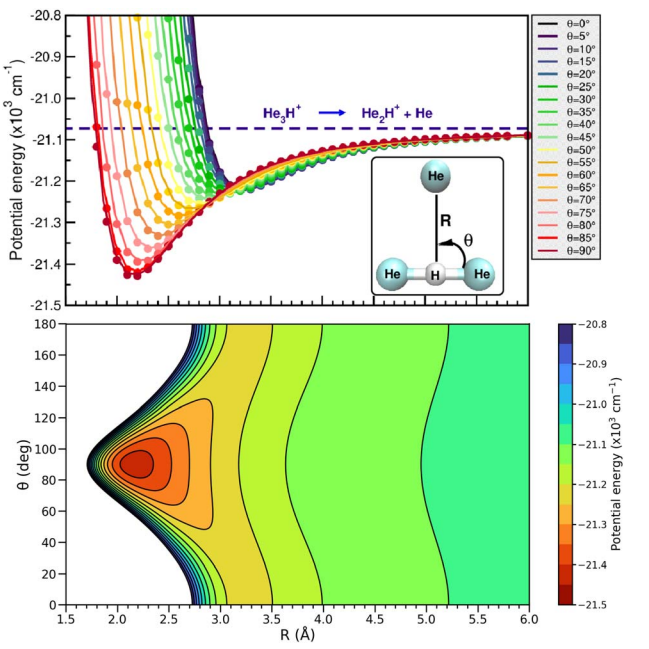

Fig. 4 Potential CCSD(T)/CBS[56] energies (see circle symbols) as a function of R distance at the indicated θ values for the He_3H^+ cation (upper panel). The corresponding RKHS ML-PES curves are also shown as solid lines, while the dissociation energy is plotted by long-dashed line. 2D contour plot of the RKHS ML-PES for He_3H^+ complex in the (θ, R)-plane (lower panel). The equipotential curves are at energies of −21450 to −21050 cm^{-1} in intervals of 50 cm^{-1}.

He_4H^+ and hepta-atomic He_6H^+ complexes, which serve as structurally significant benchmark configurations. Fig. 5 presents the potential energy curves obtained from these many-body expansions (see eqn (2)–(4)), together with the corresponding *ab initio* interaction energies evaluated at the CCSD(T)/CBS level for the $n = 2$, 3, 4, and 6 clusters. The comparison is shown along the R_1, R_3, R_3/R_4, and $R_3/R_4/R_5/R_6$ coordinates, respectively.

The upper panels of Fig. 5 illustrate the results for the $n = 2$ and $n = 3$ systems. In the $n = 2$ case, the HeH^+ internuclear distance r was fixed at its equilibrium value of 0.92480 Å, and the V_{3B} is represented by the He_2H^+ CCSD(T)/CBS potential, while for $n = 3$, the linear He–H^+–He was fixed at $r_e = 1.8496$ Å, with the V_{4B} described by the present He_3H^+ CCSD(T)/CBS[56] PES. For larger clusters, the V_{2B}, V_{3B} and V_{4B} schemes represent increasingly refined truncations of the many-body expansion. For the $n = 4$ system, we examined selected configurations in which the He–H^+–He core was kept at its equilibrium geometry r_e, and the two additional He atoms were arranged in T-shaped configurations (see the insets in the middle panels). In the middle left panel, the distance to the He atom located outside the core was fixed at 2.10 Å. In the middle right panel, a symmetry constraint $R_3 = R_4$ was imposed. For the $n = 6$ system, the external He atoms were placed in a symmetric square ring around the proton (lower panels of Fig. 5). In the lower left panel, three of the He-H^+ distances were fixed at 2.18 Å, whereas in the lower right panel all four He atoms in the ring were allowed to move symmetrically, satisfying $R_3 = R_4 = R_5 = R_6$.

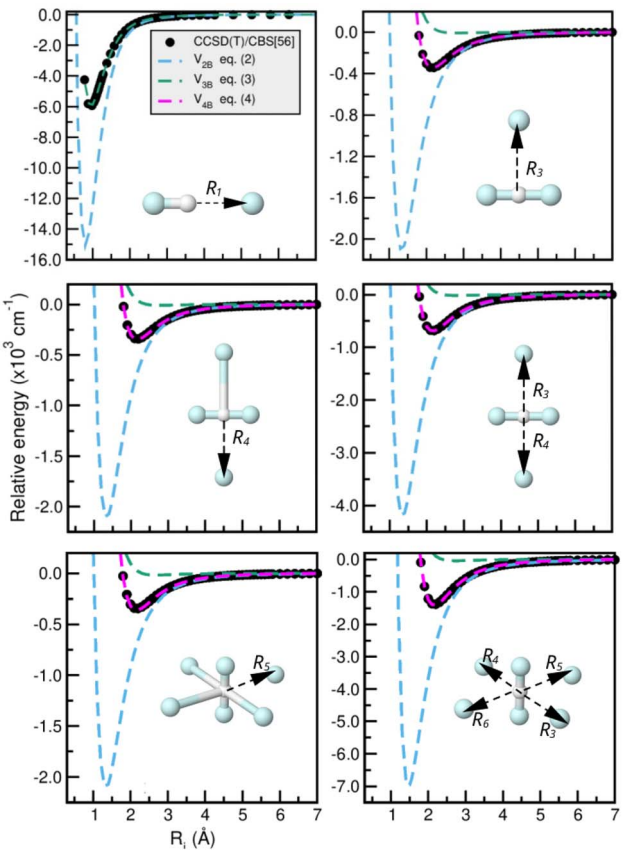

Fig. 5 Potential curves obtained for the He_nH^+ (with $n = 2$, 3, 4 and 6) clusters using the sum-of-potentials approach of eqn (2)–(4) (see color dashed lines), together with the calculated CCSD(T)/CBS[56] interaction energies (black circles) as a function of the indicated R_i coordinates.

Analysis of the results reveals that, for all systems studied, the V_{2B} model exhibits substantial discrepancies relative to the benchmark CCSD(T)/CBS[56] data. In the vicinity of the potential-well minimum, deviations reach approximately 5500 cm^{-1}, clearly demonstrating that a purely pairwise description is insufficient for modeling the energetics of the He_nH^+ clusters. Incorporating higher-order contributions through the V_{3B} model significantly reduces these errors. Upon inclusion of the 3B terms, the discrepancies in the well depth decrease to approximately 400 cm^{-1} for the He_3H^+, He_4H^+, and He_6H^+ complexes, as shown in the top-right, middle-left, and bottom-left panels of Fig. 5, respectively. However, for the second geometrical arrangements of He_4H^+ and He_6H^+ (middle-right and bottom-right panels of Fig. 5), the discrepancies increase to roughly 750 and 1500 cm^{-1}, respectively. These cases indicate a limitation of the V_{3B} model, which still fails to fully capture the interaction energy landscape for certain configurations. A substantial improvement is achieved with the V_{4B} potential model. In this case, the deviations from the CCSD(T)/CBS reference data remain below 20 cm^{-1} across all examined configurations. Based on this validated level of accuracy, the V_{4B} model is adopted for the characterization of interaction energies in all He_nH^+ clusters with $n \geq 3$.

3.3 Identification of potential energy surface minima and microsolvated geometries for He_nH^+ clusters

The final step in our analysis of the He_nH^+ clusters addresses the structural arrangements adopted by the He atoms around the proton. To this end, we applied the Basin–Hopping (BH) algorithm[49] to the V_{4B} potential, complemented by independent geometry optimizations performed at the MP2/AVQZ level and single-point CCSD(T)/AVQZ calculations for systems with $n \leq 13$.

The BH algorithm, originally introduced by Li and Scheraga under the "Monte Carlo plus energy minimization" framework,[50] enables the efficient identification of low-energy structures for clusters of up to $n = 34$ He atoms. This method is particularly well suited for exploring stable configurations in atomic and molecular clusters, both neutral and charged, as well as for more general structural optimization problems.[51] By transforming the potential energy landscape into a collection of basins and allowing stochastic hops between local minima, BH provides a robust strategy for sampling the complex configurational space associated with these microsolvated ionic clusters.

All results were obtained using an optimization temperature of $k_BT = 0.8$ meV, and four independent trajectories for each cluster size. In all cases, the global minimum was located within fewer than 10^4 BH steps. Analysis of the optimized geometries of the He_nH^+ complexes (see Table S2 in the SI and Fig. 6)

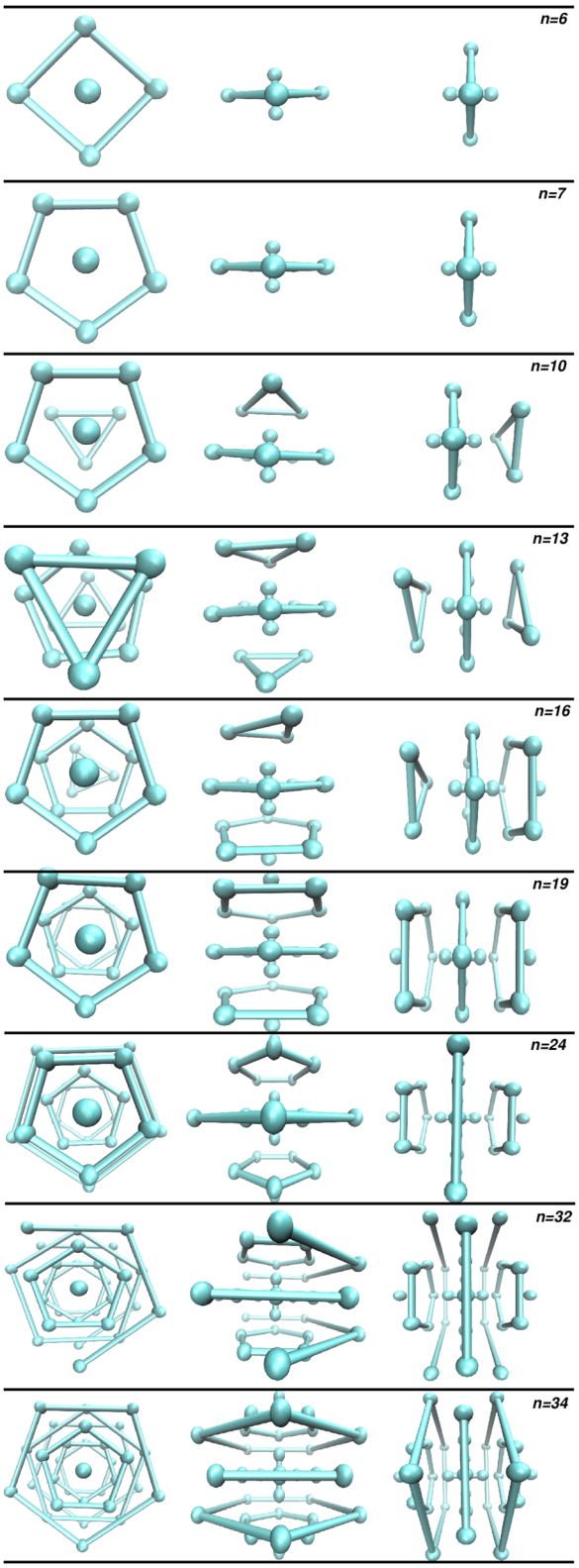

Fig. 6 Optimized structures (top and side views) for selected He_nH^+ clusters.

reveals a consistent structural motif across all cluster sizes: a central linear core that remains essentially unchanged as additional He atoms are added. This core corresponds to two strongly bound He atoms surrounding the proton, with a binding energy of approximately 4560 cm^{-1}, in sharp contrast to the significantly weaker interactions (50–330 cm^{-1}) associated with the remaining He atoms. This energetic distinction is also reflected in the bond lengths, which are markedly shorter in the core (~0.925 Å) compared to the more distant peripheral He-H^+ connections (2.15–2.25 Å), as shown in Fig. 6.

Furthermore, the optimized geometries reveal that the He atoms outside the central core preferentially occupy positions nearly perpendicular to the principal axis defined by the He–H^+–He trimer. This trend persists as the cluster grows, and the first solvation shell is completed at $n = 6$ with a square bipyramidal arrangement. At $n = 7$, the addition of one helium atom leads to the formation of a pentagonal bipyramid. For $n = 10$, the three additional He atoms form a triangular arrangement on one side of the He–H^+–He axis at distances of approximately 3.05 Å. When the cluster reaches $n = 13$, the corresponding triangular motif on the opposite side of the core is completed. Beyond this size, the growth of the He_nH^+ clusters proceeds through the sequential formation of triangular and pentagonal units on both sides of the central axis, reflecting the layered and increasingly diffuse character of the solvation shells. It is worth noting that the $He_{16}H^+$ cluster also exhibits an asymmetric structure, with one He atom occupying a linear position adjacent to the side pentagon at distance of 3.54 Å from the proton, while the opposite side contains a triangular unit. In contrast, the optimized geometry of $He_{19}H^+$ is fully symmetric, featuring a central pentagon flanked by two complete pentagonal rings, with the remaining two He atoms located along the linear axis at opposite ends of the He–H^+–He core. By $n = 24$, a second central pentagonal ring at distance of 4.2 Å is completed, signaling the onset of a more extended third solvation layer, together with two complete side pentagons at $n = 34$. This structural analysis indicates that the cationic core forms a remarkably strong bond, whereas the additional He atoms are stabilized only through much weaker interactions. Consequently, the overall properties of the clusters are largely governed by the interplay between the He–H^+–He trimer and each subsequently attached He atom.

The optimized potential energies, together with the zero point energy (ZPE) corrected values are presented in Fig. 7 as a function of cluster size up to $n = 34$. The corresponding MP2/AVQZ and CCSD(T)/AVQZ data are also shown for clusters up to $n = 13$. As the number of He atoms increases, the potential energy decreases monotonically, reflecting the progressive stabilization of the microsolvated ion. Notably, the energy curve does not follow a single linear trend; instead, four distinct slopes are observed. The changes in slope observed at $n = 6$, 13, and 19 correspond to the completion of successive solvation layers and are fully consistent with the structural growth patterns identified for the He_nH^+ clusters. These breakpoints mark the formation of the first and second solvation shells, which appear as pentagonal rings arranged around the He–H^+–He axis, and they clearly illustrate the shell-by-shell assembly of the He environment surrounding the central core. This classical layering behavior persists even after the inclusion of ZPE corrections, although the slope transitions become noticeably

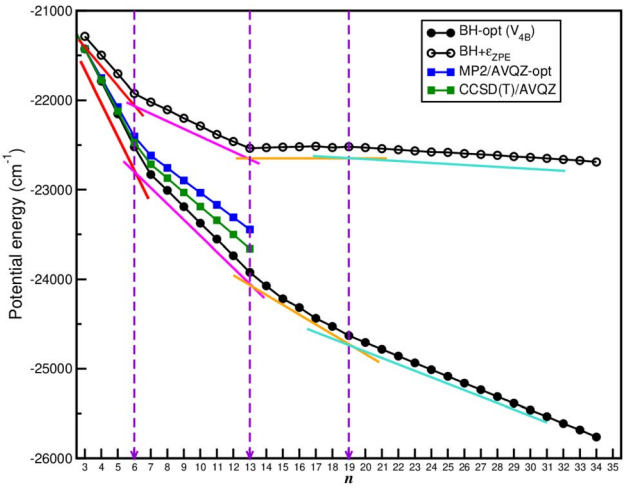

Fig. 7 Optimized BH potential energies and $\mathscr{E}_{ZPE}$ corrections for the He_nH^+ clusters using the V_{4B} SOP expansion.

less pronounced. The ZPE-corrected energies additionally reveal some atypical behavior in the range $n = 14$–19, which we discuss in detail below.

Fig. 8 displays the single-atom evaporation energies ($E_{n-1} - E_n$) as a function of cluster size, obtained from the BH and BH + ZPE simulations and compared with the experimental measurements available for the He_nH^+ clusters. The experimental data reveal several notable stability features. Ref. 24 reports measurements up to $n = 30$, identifying peaks at $n = 6$ and $n = 13$, along with smaller features at $n = 9$ for He_nH^+. In contrast, ref. 26, which extends the measurements to $n = 50$, finds enhanced stability for the He_nH^+ only at $n = 6$, 11, and 13. Despite these differences, all available experimental datasets consistently support enhanced stability at $n = 6$ and $n = 13$, and very similar pattern for $n > 15$ He_nH^+ clusters. The features at $n = 9$ and $n = 11$, however, appear to vary between experiments, suggesting a dependence on experimental conditions or detection sensitivity.

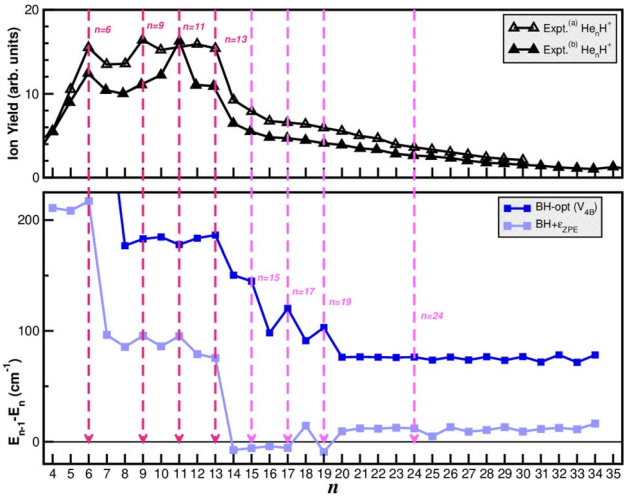

Fig. 8 Single-atom evaporative energies (solid lines) for He_nH^+ clusters, obtained from BH optimizations using the V_{4B} potential and including harmonic ZPE corrections. Experimentally measured ion-yield distributions from (a) ref. 24 and (b) ref. 26 are also shown.

Our BH + ZPE calculations reproduce the experimental peaks for the He_nH^+ at $n = 6$, 9, 11 and 13 with remarkable fidelity. Likewise, this approach predicts new peaks at $n = 18$, and 24, which currently lack experimental confirmation but appear as potential candidates for future verification. The observed changes in the experimental ion-yield curves for n between 15 to 23 and 24 to 32 may therefore be rationalized as signatures of such structural reorganizations within the growing patterns. However, the region between $n = 14$ and 19 requires particular attention. In this range, the harmonic ZPE approach appears to break down, as both the potential and evaporation energy curves exhibit an unusually flat dependence on cluster size. This behaviour is consistent with the presence of energetically close local minima and indicates substantial anharmonicity and He delocalization, suggesting that more accurate quantum treatments should be employed. Such features mirror the mobility-enhanced regions observed experimentally/theoretically,[25,27] which are considered as classical signatures of microscopic superfluidity. For comparison, the BH potential calculations also show peaks at $n = 6$ and 13, but produce a different overall profile, with additional peaks shifted to $n = 10$, 15, 17, and 19. Again the discrepancy with experimental curves highlights the critical importance of quantum ZPE corrections, especially in larger clusters where the high delocalization of He atoms significantly influences the relative stability of protonated He clusters. In this context, comparison with Ar_nH^+ clusters[14,26] shows that, beyond differences in the strengths of the underlying interactions, quantum effects are expected to be significantly more pronounced in He_nH^+. Likewise, in contrast to alkali-ion, *e.g.* Li^+, microsolvation in He, where ion-induced dipole interactions dominate, while the presence of a covalently stabilized proton-bound core in He_nH^+ leads to sharper structural reorganizations and enhanced quantum contributions.

4 Summary and conclusions

In this work, we have developed a data-driven description of the interactions governing the He_nH^+ clusters, combining CCSD(T)/CBS reference data with RKHS machine-learning techniques and a systematic many-body expansion approach. The resulting potential energy surfaces were benchmarked with additional CCSD(T)/CBS calculations, and demonstrate that inclusion of up to four-body contributions is essential to achieve quantitative fidelity for larger aggregates.

Using the V_{4B} validated PESs, global optimization searches reveal a strongly bound linear He–H^+–He structural core across all cluster sizes, onto which additional, weakly bound He atoms attach forming outer layers. The first solvation shell closes at $n = 6$ forming a square bipyramidal geometry; further growth proceeds through sequential formation of triangular, tetragonal and pentagonal units on both sides of the central pentagon ring around the He–H^+–He axis, as seen for $n = 19$, 24 and 34.

Zero-point-corrected BH + ZPE energetics reproduce with high fidelity the experimentally established stability enhancements at $n = 6$, 9, 11, and 13, and predict additional stability features at $n = 18$ and 24, which remain to be experimentally verified. Importantly, the energetic behaviour between $n = 14$ to 19 deserves particular attention, as both the BH potential and evaporation energy curves, including harmonic ZPE corrections, exhibit an anomalously flat dependence on cluster size. This flattening correlates with the presence of multiple energetically near degenerate minima, signalling significant anharmonicity and pronounced delocalization of the He atoms, allowing permutation-like motion typical of superfluid response. Such effects lie beyond the harmonic ZPE approximation and indicate that more advanced quantum treatments (*e.g.*, anharmonic corrections, path-integral or diffusion Monte Carlo methods) are required to accurately capture the size-dependent stability and structural reorganization occurring in such protonated He clusters.

Overall, this study provides a detailed mechanistic understanding of proton microsolvation in He atoms, characterizing the structural, energetic, and experimental stability trends across a wide cluster sizes. Beyond offering new predictions for unexplored cluster sizes, the methodology established here is broadly transferable and can be extended to other weakly bound protonated noble-gas clusters, enabling systematic and size-consistent modeling of complex quantum solvation phenomena.

Conflicts of interest

There are no conflicts to declare.

Data availability

The authors declare that the data supporting the findings of this study are available within the article and its supplementary information (SI). Additional data are available from the corresponding author upon reasonable request. Supplementary information is available. See DOI: https://doi.org/10.1039/d6ra00552g.

Acknowledgements

The authors thank to the Centro de Calculo del IFF/SGAI-CSIC and CESGA-Supercomputing centre for allocation of computer time. This work has been supported by the Comunidad de Madrid grants ref: IND2018/TIC-9467 and Investigo grant ref: INV23-IFF-M2-09, the MCIU grant no. PID2024-155666NB-I00, the CSIC-PEICT ref. 2024AEP119 and the COST Actions CA21126(NanoSpace) and CA21101(COSY).

Notes and references

1 J. S. Rigden, *Hydrogen: the Essential Element*, Harvard University Press, 2003.
2 J. M. McMahon, M. A. Morales, C. Pierleoni and D. M. Ceperley, *Rev. Mod. Phys.*, 2012, **84**, 1607–1653.
3 M. Barlow, B. Swinyard, P. Owen, J. Cernicharo, H. Gomez, R. Ivison, O. Krause, T. Lim, M. Matsuura, S. Miller, G. Olofsson and E. Polehampton, *Science*, 2013, **342**, 1343–1345.
4 H. Müller, S. Muller, P. Schilke, E. Bergin, J. Black, M. Gerin, D. Lis, D. Neufeld and S. Suri, *Astron. Astrophys.*, 2015, **582**, L4.
5 R. Güsten, H. Wiesemeyer, D. Neufeld, K. Menten, U. Graf, K. Jacobs, B. Klein, O. Ricken, C. Risacher and J. Stutzki, *Nature*, 2019, **568**, 357–359.
6 F. Grandinetti, *Int. J. Mass Spectrom.*, 2004, **237**, 243–267.
7 F. Grandinetti, *Front. Chem.*, 2020, **8**, 462.
8 S. Borocci, N. Bronzolino, M. Giordani and F. Grandinetti, *J. Chem. Phys. A.*, 2010, **114**, 7382–7390.
9 S. Borocci, F. Grandinetti and N. Sanna, *Molecules*, 2021, **26**, 1305.
10 M. J. Montes de Oca-Estévez and R. Prosmiti, *Front. Chem.*, 2021, **9**, 187.
11 M. J. Montes de Oca-Estévez and R. Prosmiti, *J. Mol. Graph. Model.*, 2023, **124**, 108562.
12 M. J. Montes de Oca-Estévez, B. Darna, B. García-Ruiz, R. Prosmiti, T. González-Lezana and D. Koner, *ChemPhysChem*, 2023, **24**, e202300450.
13 M. J. Montes de Oca-Estévez and R. Prosmiti, *AI Chem.*, 2024, **2**, 100059.
14 M. J. Montes de Oca-Estévez and R. Prosmiti, *Molecules*, 2024, **29**, 4084.
15 M. J. Montes de Oca-Estévez, Á. Valdés and R. Prosmiti, *Phys. Chem. Chem. Phys.*, 2024, **26**, 7060–7071.
16 M. J. Montes de Oca-Estévez, Á. Valdés, D. Koner, T. González-Lezana and R. Prosmiti, *Chem. Phys. Lett.*, 2024, **856**, 141641.
17 M. J. Montes de Oca-Estévez, Á. Valdés and R. Prosmiti, *Molecules*, 2025, **30**, 2440.
18 M. J. Montes de Oca-Estévez, P. Villarreal, J. Hernández-Rojas, J. Bretón, A. Sarsa and R. Prosmiti, *Chem. Phys. Lett.*, 2025, **877**, 142242.
19 S. Werbowy and B. Pranszke, *Astron. Astrophys.*, 2011, **535**, A51.
20 S. Bag, M. R. McCoustra and T. Pradeep, *J. Phys. Chem. C*, 2011, **115**, 13813–13819.
21 E. Herbst, *Philos. Trans. R. Soc., A*, 2000, **358**, 2523–2534.
22 T. Oka, *Chem. Rev.*, 2013, **113**, 8738–8761.
23 T. M. Kojima, N. Kobayashi, Y. Kaneko and Z. P. D - Atoms Molec, *Clusters*, 1992, **23**, 181–185.
24 P. Bartl, C. Leidlmair, S. Denifl, P. Scheier and O. Echt, *Chem. Phys. Chem*, 2013, **14**, 227–232.
25 A. G. Császár, T. Szidarovszky, O. Asvany and S. Schlemmer, *Mol. Phys.*, 2019, **117**, 1559–1583.
26 L. Lundberg, P. Bartl, C. Leidlmair, P. Scheier and M. Gatchell, *Molecules*, 2020, **25**, 1066.
27 O. Asvany, S. Schlemmer, T. Szidarovszky and A. G. Császár, *Chem. Phys. Lett.*, 2019, **10**, 5325–5330.
28 D. C. McDonald, D. T. Mauney, D. Leicht, J. H. Marks, J. A. Tan, J.-L. Kuo and M. A. Duncan, *J. Chem. Phys.*, 2016, **145**, 231101.

29 F. Filippone and F. A. Gianturco, *Europhys. Lett.*, 1998, **44**, 585.
30 F. A. Gianturco and F. Filippone, *Chem. Phys.*, 1999, **241**, 203–212.
31 I. Baccarelli, F. A. Gianturco and F. Schneider, *Int. J. Quantum Chem.*, 1999, **74**, 193–212.
32 B. Balta and F. Gianturco, *Chem. Phys.*, 2000, **254**, 203–213.
33 M. Beyer, A. Lammers, E. V. Savchenko, G. Niedner-Schatteburg and V. E. Bondybey, *Phys. Chem. Chem. Phys.*, 1999, **1**, 2213–2221.
34 B. Balta, F. Gianturco and F. Paesani, *Chem. Phys.*, 2000, **254**, 215–229.
35 R. C. Fortenberry and L. Wiesenfeld, *Molecules*, 2020, **25**, 2183.
36 C. E. Dykstra, *J. Mol. Struct. Theochem.*, 1983, **103**, 131–138.
37 C. J. Stephan and R. C. Fortenberry, *Mon. Not. R. Astron. Soc.*, 2017, **469**, 339–346.
38 L. González-Sánchez, E. Yurtsever, R. Wester and F. A. Gianturco, *J. Phys. Chem. A*, 2021, **125**, 3748–3759.
39 O. Denis-Alpizar, L. D. Cabrera-González, D. Páez-Hernández and R. Pino-Rios, *ACS Earth Space Chem.*, 2022, **6**, 1924–1929.
40 J. N. Murrell, *Molecular Potential Energy Functions*, Wiley, 1984.
41 R. A. Aziz and M. J. Slaman, *J. Chem. Phys.*, 1991, **94**, 8047–8053.
42 H.-J. Werner, P. Knowles, G. Knizia, F. Manby, M. Schütz and *et al.*, *MOLPRO, Version 2022.2, a Package of ab initio Programs*, 2022, see http://www.molpro.net.
43 *DENEB 1.30 beta: The Nanotechnology Software by Atelgraphics.*, See http://www.atelgraphics.com, 2020.
44 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.
45 A. K. Wilson, T. van Mourik and T. H. Dunning, *J. Mol. Struct.*, 1996, **388**, 339–349.
46 C. Schwartz, *Phys. Rev.*, 1962, **126**, 1015–1019.
47 S. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553–566.
48 T. Ho and H. Rabitz, *J. Chem. Phys.*, 1996, **104**, 2584–2597.
49 D. J. Wales and J. P. Doye, *J. Phys. Chem. A*, 1997, **101**, 5111–5116.
50 Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 6611–6615.
51 J. Hernández-Rojas, F. Calvo, J. Bretón and J. Gomez Llorente, *J. Phys. Chem. C*, 2012, **116**, 17019–17028.