

Cite this: *Digital Discovery*, 2025, 4, 824

# Quantitative evaluation of anharmonic bond potentials for molecular simulations†

Paul J. van Maaren and David van der Spoel \*

Most general force fields only implement a harmonic potential to model covalent bonds. In addition, in some force fields, all or a selection of the covalent bonds are constrained in molecular dynamics simulations. Nevertheless, it is possible to implement accurate bond potentials for a relatively small computational cost. Such potentials may be important for spectroscopic applications, free energy perturbation calculations or for studying reactions using empirical valence bond theory. Here, we evaluate different bond potentials for diatomic molecules. Based on quantum-chemical scans around the equilibrium distance of 71 molecules using the MP2/aug-cc-pVTZ level of theory as well as CCSD(T) with the same basis-set, we determine the quality of fit to the data of 28 model potentials. As expected, a large spread in accuracies of the potentials is found and more complex potentials generally provide a better fit. As a second and more challenging test, five spectroscopic parameters ( $\omega_e$ ,  $\omega_e X_e$ ,  $\alpha_e$ ,  $B_e$  and  $D_e$ ) predicted based on quantum chemistry as well as the fitted potentials are compared to experimental data. A handful of the 28 potentials tested are found to be accurate. Of these, we suggest that the potential due to Hua (*Phys. Rev. A*, 42 (1990), 2524) could be a suitable choice for implementation in molecular simulations codes, since it is considerably more accurate than the well-known Morse potential (*Phys. Rev.*, 34 (1929), 57) at a very similar computational cost.

Received 27th October 2024  
Accepted 12th February 2025

DOI: 10.1039/d4dd00344f

rsc.li/digitaldiscovery

## 1 Introduction

Prediction of molecular properties can be done, in principle, through theoretical models based on physics or by models based directly on data.<sup>1</sup> Both ways involve approximations and require thorough validation based on experimental data. Importantly, though there has been great progress in data-based models,<sup>2</sup> the laws of physics remain valid.<sup>3</sup> Data and physics-based models share the need for high-quality reference data, and we have recently presented an overview of available quantum-chemistry databases for this purpose.<sup>4</sup> Here, we aim to design parts of a force field for molecular dynamics (MD) simulation, for which empirical potentials are needed that reproduce data from high-quality quantum chemistry or experiments. This means that the fundamental physics should be followed as much as possible without making the potentials impractically complicated.<sup>5</sup> Systematic design of force fields<sup>6</sup> is needed to determine which functions are suitable for predicting properties.<sup>7</sup> By careful design of the data set it is possible to break down the complex empirical potential for molecules into simpler parts. We have, for instance, studied noble gases to evaluate potential forms for exchange and dispersion

interactions<sup>8</sup> and found that the 14-7 potential due to Halgren<sup>9</sup> as well as the generalized Buckingham due to Werhahn *et al.*<sup>10</sup> were sufficiently flexible to reproduce both gas-phase and condensed phase data. On the other hand, the popular Lennard-Jones 12-6 potential<sup>11</sup> was the poorest contender. We obtained similar results in an earlier study of alkali-halides,<sup>12</sup> where multiple potentials were parameterised in exactly the same manner, allowing apples-to-apples comparisons of their predictive power. There we found that the 12-6 potential had three times higher deviation from experimental observables than a modified Buckingham potential.<sup>13</sup>

In this paper, we address potentials for covalent bonds by studying diatomic molecules. The study of diatomic molecules has a long history, with famous old papers such as those by Heitler and London,<sup>14</sup> Morse,<sup>15</sup> Rydberg<sup>16</sup> and Pöschl and Teller.<sup>17</sup> Systematic comparisons of potentials with experimental data have been done, for instance by Royappa *et al.*,<sup>18</sup> and by others,<sup>19,20</sup> usually focused on reproduction of vibrational modes. It should be noted that many of these potentials are “related” to each other<sup>21,22</sup> but we will not discuss this here. Instead, we refer the reader to a recent review by Araújo *et al.* covering 100 years of history of analytical potentials to fit diatomic energy curves.<sup>23</sup> The development of methods for accurate yet affordable prediction of vibrational spectra has been an active research field for a long time<sup>24–26</sup> and the choice of a potential for covalent bonds is a step towards force field based prediction of vibrational spectra.<sup>27,28</sup>

Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden.  
E-mail: david.vanderspoel@icm.uu.se

† Electronic supplementary information (ESI) available: Additional tables and figures. See DOI: <https://doi.org/10.1039/d4dd00344f>



When addressing the question what bond potential is best suitable for molecular simulation, it is important to realize that “best” by necessity involves a compromise between accuracy (deviation from experimental data or high-level quantum chemistry) and computational efficiency. A further consideration is that a potential should be simple to parameterize, which favors functions with fewer parameters. It should be noted that many classical force fields use a simple harmonic function to model chemical bonds although, for instance, the MM3 force field can employ a Morse function instead.<sup>29</sup> In what follows, we consider the root mean square deviation (RMSD) from the reference data as the “accuracy” of the potential. For comparison with earlier studies, we also provide the least-squares Z-score introduced by Murrell and Sorbie.<sup>30</sup> As we have shown in our study on noble gases, it is often advantageous or even necessary to introduce an energy threshold for fitting potentials. The magnitude of such a threshold is important quantitatively, but often not qualitatively.<sup>8</sup> The size of the RMSD changes with energy threshold used for training, but the ranking of potentials does not change much. We discuss the effect of different thresholds on different potential functions below. Finally, we evaluate the quantum chemistry data and a number of empirical potentials by computing spectroscopic parameters and comparing them to experimental data.

## 2 Methods

### 2.1 Energy calculations

71 diatomic molecules were selected, consisting of first and second row atoms plus sulfur, halogens and alkali halides. The dataset therefore consists of both covalently bound molecules and ion-pairs. Ion-pairs are included because there is a lot of experimental data available to compare fitted potentials to, including spectroscopic data, allowing to test our scripts and the generality of the potentials. In MD simulations, one would not use a covalent potential for ion pairs but rather a combination of Coulomb and van der Waals potentials.<sup>12</sup> For each of these molecules respectively ion pairs, a scan of the energy as a function of distance was made at two levels of theory. First, Møller–Plesset 2nd order (MP2) perturbation theory<sup>31</sup> with the correlation-consistent basis-set aug-cc-pVTZ<sup>32</sup> and second, coupled clusters<sup>33</sup> with singlets and doublets and perturbative triples, or CCSD(T), with the same basis-set. Iodine and iodide were modeled using the aug-cc-pwCVTZ-PP basis set<sup>34</sup> while for potassium, calcium, cesium and rubidium the def2-TZVPP basis set was employed,<sup>35</sup> all downloaded from the basis-set exchange.<sup>36</sup> For singlet molecules a restricted Hartree–Fock procedure was used.<sup>37</sup> Initially, unrestricted Hartree–Fock calculations were performed for the radicals, but in particular for the MP2 method this led to significant spin contamination. For this reason, restricted open-shell Hartree–Fock<sup>38</sup> calculations were used instead, and these were well-behaved.

A list of molecules, their charge, multiplicity and the range of distances used for quantum calculations is provided in Table S1.† The range of distances used in the quantum calculations was based on the equilibrium distance  $r_e$  by dividing it by 1.2 and multiplying it by 1.2 to get the lower and upper limits

respectively (Table S1†). The distance between scanning points was 0.5 pm. Calculations were performed with the Psi4 software suite.<sup>39</sup> Experimental data for the subset of 14 molecules used by Royappa *et al.*<sup>18</sup> were used for fitting here as well. References to experimental data for these molecules are given in Fig. S9–S13, S25, S28, S30, S39, S43, S48–S50, S58 and S59, in the ESI.† Experimental data on frequencies were collected from the database of spectroscopic constants of diatomic molecules<sup>40,41</sup> and the National Institute of Standards Webbook.<sup>42</sup> We note that the database contained several errors and omissions that we corrected based on the original data collection due to Huber and Herzberg,<sup>43</sup> which itself also contained some errors. Corrected files are provided on Zenodo.<sup>44</sup>

### 2.2 Data processing

Energies were stored as text files and processed with a curve-fitting script based on Scientific Python.<sup>45</sup> 28 functions were used to fit the energy curves, 21 of which were evaluated by Royappa *et al.* previously.<sup>18</sup> In addition, we use a harmonic potential, potentials due to Lennard-Jones,<sup>11</sup> Buckingham,<sup>46</sup> Cahill,<sup>47</sup> Tang & Toennies,<sup>48</sup> Xie *et al.*<sup>49</sup> and Wang *et al.*<sup>13</sup> Numerical curve-fitting was extremely tedious. First, a careful shifting of the quantum chemical energy curves to have their minimum energy level at zero was needed. Despite tuning of tolerances, the energies produced by geometry optimization were incompatible with single point calculations. Therefore, we located the position and depth of the energy minimum by implementing a bisection algorithm using single point calculations with exactly the same settings as used in the distance scan. Then, manual curating of the starting values was required, using visual inspection of the fitted curves to validate the correctness of the fit. Unfortunately, the Levenberg–Marquardt algorithm<sup>50,51</sup> used in Scientific Python<sup>45</sup> simply is not fool-proof. Unless starting parameters are close to the correct ones, it cannot be guaranteed that the best solution will be found, in particular since we used some highly non-linear potentials with a relatively small number of correlated data points. The total number of fits to be checked in this manner was well over 10 000. This suggests that, as much as we would like it to, the era of digital discovery has not yet fully started. Even though the curve fitting applied here only requires a small number of parameters to be determined, we are not aware of any guaranteed error-free solution. More elaborate algorithms like Monte Carlo search in parameter space coupled with simulated annealing might help, but any such algorithm can get stuck in a local minimum as well. A further possibility would be the algorithm due to Ho and Rabitz for generating a molecular energy surface from quantum chemistry calculations.<sup>52,53</sup> Finally, for fitting the data a number of different energy thresholds were employed: either all data was used, or just the data points with energy of at most 1000 cm<sup>-1</sup> respectively 5000 cm<sup>-1</sup>.

Fitting and evaluation of the goodness of fit was done using the Z-score<sup>30</sup>

$$Z = \frac{1}{N\Delta r} \sum_{i=1}^N (E_{\text{obs}} - E_{\text{fit}})^2 \quad (1)$$



where  $\Delta r$  is the range of distances considered in the fit and  $N$  is the number of data points.  $E_{\text{obs}}$  is the energy from experiment or quantum chemistry and  $E_{\text{fit}}$  the energy according to the analytical potential. We also list the conventional root mean square deviation (RMSD)

$$\text{RMSD} = \left[ \frac{1}{N} \sum_{i=1}^N (E_{\text{obs}} - E_{\text{fit}})^2 \right]^{1/2} \quad (2)$$

or, in other words  $\text{RMSD} = (Z\Delta r)^{1/2}$ . The  $Z$ -scores were averaged linearly over molecules to obtain one number per empirical potential. The RMSD were squared before averaging over molecules, followed by taking a square root. Due to non-linear relation between  $Z$  and RMSD, the order of empirical potentials may differ slightly in the Tables 1, 2, S2 and S3.† For compatibility with earlier papers (e.g. by Royappa *et al.*<sup>18</sup>) we present the  $Z$ -scores in wave numbers, whereas the RMSD are given in  $\text{J mol}^{-1}$  which is customary in the molecular simulation community.

## 3 Results & discussion

### 3.1 Experimental reference data

To validate our Python code, we attempted to reproduce the data included in the paper by Royappa *et al.*<sup>18</sup> The curve-fitting script reproduces Fig. 2–5 from that paper. We found lower average  $Z$ -scores for all of the potentials (Table 1), likely in part because of the manual curation of the data. It is also possible, however, that in some cases a better fit was obtained because

the fitting algorithm switched places between the attractive and repulsive part of the potentials. For example, in the potential due to Linnett<sup>65</sup>

$$U(r) = \frac{a}{r^m} - be^{-nr} \quad (3)$$

the first term is supposed to model the repulsion and the second part the attraction. In our training we find that without exception  $a$  and  $b$  become negative. In other words, we did not enforce parameters that historically have been identified with a certain physical interpretation, such as  $r_e$  and  $D_e$ , to be within experimental range. For the experimental data set, no energy threshold was applied, that is all data points were taken into account. This is more demanding in terms of the functional form and like Royappa<sup>18</sup> and co-workers we find that the more complex functions are better able to reproduce the experimental data (Table 1).

### 3.2 Quantum chemical reference data

Since quantum chemistry for diatomics is relatively cheap it was possible to include 71 diatomic molecules. The quantum chemistry curves are plotted in Fig. S1–S71† and, where available, experimental data are plotted as well. Statistics of the potentials when fitted using an energy threshold of  $1000 \text{ cm}^{-1}$  are given in Table 2. As expected, simple functions like Lennard-Jones<sup>11</sup> and the harmonic function represent the quantum chemistry data poorly. The *ab initio* model due to Xie *et al.*<sup>49</sup> was not accurate for the molecules studied here either, likely due to the fact that most molecules studied here are not just bonded by  $s$ -valence electrons, like in the original paper.<sup>49</sup> The well-known Tang–Toennies potential,<sup>48</sup> that reproduces high-level quantum chemistry interaction functions for noble gases extremely well,<sup>8</sup> is not among the top contenders either. Table S2† gives the corresponding data for a fit with an energy threshold of  $5000 \text{ cm}^{-1}$  and Table S3† without any energy threshold. No significant difference with Table 2 is found, the potentials with the ten or so best fits to the quantum chemistry data are just shuffled in a somewhat different order. For the best performing functions, it seems somewhat easier to reproduce the CCSD( $T$ ) data than the MP2 data (Tables 2 and S2†). It can also be noted that some potentials, like the ones due to Rafi<sup>57</sup> or Levine,<sup>58</sup> are more accurate for non-covalent interactions than covalent bonds. For the purpose of this paper we are mainly interested in potentials that model covalent bonds well, so we will not investigate this further.

### 3.3 Spectroscopic parameters

Based on the quantum-chemical data, the vibrational harmonic frequency, the first anharmonic correction and other vibrational parameters<sup>24</sup> were computed by second order vibrational perturbation theory using the Psi4 software<sup>39</sup> (Table S4†). It should be noted that slightly more accurate vibrational constants may be computed by directly solving the 1D Schrödinger equation<sup>26,70,71</sup> but vibrational perturbation theory is accurate enough to distinguish the accuracy of the potentials under evaluation here. Fig. 1 displays the residual (quantum

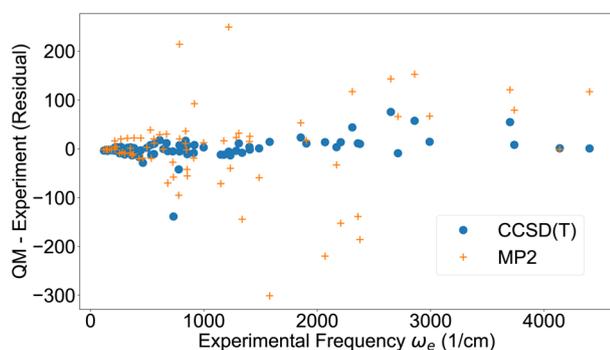
**Table 1** Statistics per function for fits to experimental data for the 14 compounds used by Royappa *et al.*<sup>18</sup>  $M$  is the number of parameters,  $Z$  is the average  $Z$ -score ( $\text{cm}^{-2} \text{ \AA}^{-1}$ ),  $\Delta Z$  indicates the difference between the  $Z$  calculated here and that by Royappa, and RMSD ( $\text{J mol}^{-1}$ ) is the root mean signed error from experimental data without any energy cut-off. Table is sorted after  $Z$ -score

Function	$M$	$Z$	$\Delta Z$	RMSD
Sun <sup>54</sup>	8	534	−1264	29
Murrell–Sorbie <sup>30</sup>	5	3300	−1158	66
Hulburt–Hirschfelder <sup>55</sup>	5	4050	−1154	76
Tietz II <sup>56</sup>	5	9521	−1696	144
Rafi <sup>57</sup>	5	9985	−45762	137
Levine <sup>58</sup>	4	10 517	−18734	150
Noorizadeh <sup>59</sup>	5	13 447	−4086	129
Wei Hua <sup>60</sup>	4	13 707	−1127	174
Pöschl–Teller <sup>17</sup>	4	22 749	−61372	175
Frost–Musulin <sup>61</sup>	4	30 581	−1434	223
Morse <sup>15</sup>	3	47 826	−1641	282
Varshni <sup>62</sup>	3	57 698	−1431	315
Rosen–Morse <sup>63</sup>	4	60 451	−94393	349
Rydberg <sup>16</sup>	3	69 734	−1406	357
Pseudo-Gaussian <sup>64</sup>	3	86 717	−1148	352
Linnett <sup>65</sup>	4	107 285	−28842	489
Deng–Fan <sup>66</sup>	3	154 247	−2646	577
Tietz I <sup>56</sup>	5	185 811	−60007	596
Valence-state <sup>67</sup>	4	205 072	−3700	625
Kratzer <sup>68</sup>	2	4 424 629	−6825	2283
Lippincott <sup>69</sup>	3	10 132 095	−18092	3330



**Table 2** Statistics per function for quantum chemistry results.  $M_f$  is the number of parameters used for fitting,  $M_{sim}$  the number of parameters if the minimum is not fixed at zero and when redundancies are removed (see text).  $N$  is the number of compounds,  $Z$  is the average  $Z$ -score ( $\text{cm}^{-2} \text{Å}^{-1}$ ), and RMSD ( $\text{J mol}^{-1}$ ) is the root mean signed error from quantum chemical results. An energy cut-off of  $1000 \text{ cm}^{-1}$  was applied. Table is sorted after  $Z$ -score for covalent compounds computed at the CCSD( $T$ ) level of theory

Function	$M_f$	$M_{sim}$	CCSD( $T$ )				MP2			
			Non-covalent (26)		Covalent (45)		Non-covalent (26)		Covalent (45)	
			$Z$	RMSD	$Z$	RMSD	$Z$	RMSD	$Z$	RMSD
Sun <sup>54</sup>	8	8	0.0	0.5	0.0	0.1	0.0	0.1	0.0	0.1
Hulburt–Hirschfelder <sup>55</sup>	5	5	0.0	0.9	0.0	0.1	0.0	0.4	0.0	0.2
Tietz II <sup>56</sup>	5	4	0.1	2.5	0.0	0.5	0.1	2.0	0.1	1.5
Wei Hua <sup>60</sup>	4	4	0.1	2.7	0.0	0.5	0.1	2.2	0.2	2.2
Cahill <sup>47</sup>	6	6	0.0	0.4	0.0	0.3	3.8	9.4	0.9	4.9
Rafi <sup>57</sup>	5	4	0.1	1.8	0.1	2.1	0.1	1.7	0.5	4.8
Murrell–Sorbie <sup>30</sup>	5	5	0.1	2.7	0.2	2.7	0.2	3.0	0.4	3.6
Frost–Musulin <sup>61</sup>	4	3	0.6	5.6	0.2	3.7	0.6	5.6	1.0	5.0
Pöschl–Teller <sup>17</sup>	4	3	0.6	5.8	0.2	3.7	0.7	5.9	1.1	5.1
Valence-state <sup>67</sup>	4	4	0.6	5.0	0.3	4.3	0.5	4.6	1.2	5.4
Rosen–Morse <sup>63</sup>	4	3	0.9	7.0	0.4	3.8	0.9	7.0	1.2	5.4
Morse <sup>15</sup>	3	3	0.9	7.3	0.4	4.2	1.1	7.3	1.6	6.1
Rydberg <sup>16</sup>	3	3	1.2	8.3	0.4	4.3	1.3	8.4	1.7	6.3
Levine <sup>58</sup>	4	4	0.2	3.5	0.4	5.6	4.2	10	5.8	17
Linnett <sup>65</sup>	4	3	0.6	5.7	0.5	4.6	0.6	5.6	1.5	6.4
Pseudo-Gaussian <sup>64</sup>	3	3	1.9	11	0.6	5.0	2.0	11	1.7	6.5
Tietz I <sup>56</sup>	5	4	1.0	6.9	0.8	6.5	1.0	6.6	4.9	9.4
Varshni <sup>62</sup>	3	3	1.5	9.5	2.1	8.0	3.7	12	4.2	10
Deng–Fan <sup>66</sup>	3	3	0.7	6.0	5.1	13	4.8	11	10	18
Wang–Buckingham <sup>13</sup>	3	3	8.3	22	9.9	20	9.1	22	13	21
Tang (2003) <sup>48</sup>	6	6	17	26	44	29	1.0	6.7	98	43
Buckingham <sup>46</sup>	3	3	4.7	16	44	48	5.1	16	56	53
Noorizadeh <sup>59</sup>	5	4	23	29	273	78	22	29	217	69
Kratzer <sup>68</sup>	2	2	578	176	380	101	483	166	365	93
Lippincott <sup>69</sup>	3	3	1753	307	1471	196	1554	294	1409	186
Xie (2005) <sup>49</sup>	4	3	1747	299	1818	213	1512	284	1767	203
Harmonic	3	2	3848	452	3648	314	3506	438	3531	302
Lennard-Jones <sup>11</sup>	2	2	5487	532	9044	523	5690	534	9287	529

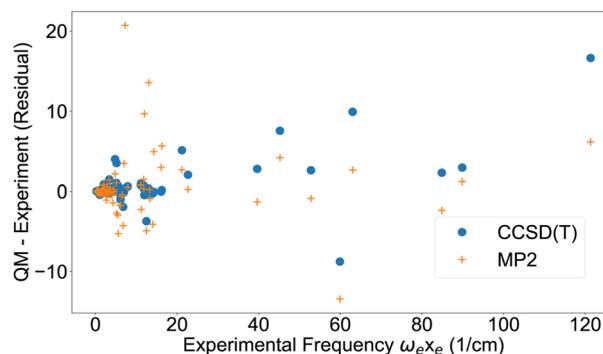


**Fig. 1** Vibrational harmonic frequency  $\omega_e$  from experimental data<sup>42,43</sup> and residual from quantum chemistry.

chemistry minus experiment) of the vibrational harmonic frequency  $\omega_e$  for the 71 molecules studied here. Despite some outliers, the overall trend is that the frequencies are reproduced well when using CCSD( $T$ ), but with considerably more noise for MP2. These results can be compared to results from a machine learning study by Ibrahim and co-workers, who built models to predict spectroscopic constants of diatomic molecules based on

atomic and molecular properties.<sup>72</sup> Their best model produced vibrational harmonic frequencies  $\omega_e$  of similar accuracy to the ones from CCSD( $T$ ) (Fig. 1).

The first anharmonic correction  $\omega_e x_e$ , is overestimated by both the MP2 and CCSD( $T$ ) methods (Fig. 2). These results are corroborated by a summary of statistics in Table 3, showing the deviation from either experiment or CCSD( $T$ ). MP2 has almost



**Fig. 2** First anharmonic correction  $\omega_e x_e$  from experimental data<sup>42,43</sup> and residual from quantum chemistry.



**Table 3** Percent deviation for vibrational harmonic frequency  $\omega_e$ , first anharmonic correction  $\omega_e x_e$ , equilibrium rotational constant  $B_e$ , first correction of the rotational constant  $\alpha_e$ , centrifugal distortion constant  $D_e$ , for different methods and analytical potentials fitted on CCSD(*T*) with an energy threshold of 1000 cm<sup>-1</sup>, from the reference indicated below. Potentials sorted according to deviation of  $\omega_e$  from CCSD(*T*)

Reference	Experiment					CCSD( <i>T</i> )					
	Method	$\omega_e$	$\omega_e x_e$	$B_e$	$\alpha_e$	$D_e$	$\omega_e$	$\omega_e x_e$	$B_e$	$\alpha_e$	$D_e$
CCSD( <i>T</i> )		2.92	18.64	2.77	14.12	26.92					
MP2		6.65	45.47	2.47	20.34	28.21	6.33	41.80	2.13	18.15	9.37
Sun <sup>54</sup>		2.92	18.24	2.77	14.16	26.92	0.01	0.58	0.00	0.09	0.03
Hulburt–Hirschfelder <sup>55</sup>		2.92	18.07	2.77	14.07	26.92	0.01	0.67	0.00	0.13	0.03
Cahill <sup>47</sup>		2.92	18.85	2.77	14.13	26.92	0.02	2.28	0.00	0.17	0.03
Wei Hua <sup>60</sup>		2.92	17.61	2.77	13.89	26.92	0.04	1.68	0.00	0.77	0.08
Tietz II <sup>56</sup>		2.93	18.10	2.77	13.85	26.92	0.05	2.55	0.00	0.93	0.11
Rafi <sup>57</sup>		2.94	17.46	2.77	13.85	26.93	0.08	2.63	0.00	1.19	0.16
Murrell–Sorbie <sup>30</sup>		2.92	20.33	2.77	14.13	26.92	0.08	7.78	0.00	1.00	0.17
Levine <sup>58</sup>		2.93	18.01	2.77	13.76	26.93	0.14	6.34	0.01	2.15	0.30
Deng–Fan <sup>66</sup>		2.93	19.47	2.77	16.68	26.94	0.15	9.49	0.04	8.80	0.36
Morse <sup>15</sup>		2.92	19.74	2.77	14.05	26.94	0.15	9.02	0.01	2.09	0.32
Rydberg <sup>16</sup>		2.91	21.33	2.77	14.00	26.95	0.17	10.71	0.01	2.24	0.36
Tietz-I <sup>56</sup>		2.97	18.86	2.77	14.53	26.93	0.18	10.23	0.01	2.74	0.39
Frost–Musulin <sup>61</sup>		2.93	18.50	2.77	14.04	26.95	0.19	7.70	0.01	2.14	0.39
Pöschl–Teller <sup>17</sup>		2.93	18.87	2.77	14.05	26.95	0.19	8.20	0.01	2.14	0.40
Valence-state <sup>67</sup>		2.98	16.04	2.77	14.34	26.93	0.19	8.96	0.01	2.15	0.40
Varshni <sup>62</sup>		2.90	22.51	2.77	14.74	26.96	0.20	11.60	0.02	4.99	0.42
Tang (2003) <sup>48</sup>		2.92	21.49	2.76	17.32	26.94	0.20	15.19	0.06	14.69	0.40
Pseudo-Gaussian <sup>64</sup>		2.89	23.89	2.77	13.89	26.96	0.21	13.43	0.01	2.74	0.44
Rosen–Morse <sup>63</sup>		2.91	22.51	2.77	14.04	26.96	0.22	12.53	0.01	2.20	0.45
Linnert <sup>65</sup>		2.93	22.89	2.77	14.20	26.94	0.22	14.12	0.01	2.04	0.44
Noorizadeh <sup>59</sup>		2.94	42.83	2.75	36.21	26.94	0.37	40.52	0.14	34.52	0.45
Buckingham <sup>46</sup>		2.98	69.91	2.78	22.85	26.97	0.59	59.92	0.16	21.71	1.24
Wang–Buckingham <sup>13</sup>		2.90	60.73	2.77	15.03	27.01	0.60	52.84	0.02	6.29	1.22
Kratzer <sup>68</sup>		3.32	63.09	2.96	50.86	26.92	0.81	63.64	0.32	50.07	0.76
Lippincott <sup>69</sup>		3.06	100	3.09	100	27.01	0.97	90.31	0.61	100	2.40
Xie (2005) <sup>49</sup>		3.69	87.13	3.09	100	26.92	1.47	87.41	0.58	100	1.38
Harmonic		3.80	—	—	—	—	1.66	—	—	—	—
Lennard–Jones <sup>11</sup>		5.02	100	2.89	100	27.41	4.69	100	1.65	100	4.15

three times higher deviation from experiment than CCSD(*T*) for  $\omega_e$  and  $\omega_e x_e$ , however for the other parameters the difference is smaller.

The potentials, fitted to CCSD(*T*) with a threshold of 1000 cm<sup>-1</sup>, were used to compute vibrational parameters as well (Table 3). The five best potentials from Table 2 are the best here as well. All of these seem to reproduce the CCSD(*T*) energy curves faithfully as they sport the same deviation from experiment as CCSD(*T*) and low deviation from the CCSD(*T*) frequencies as well. The deviation for frequencies  $\omega_e$  from experiment for the harmonic potential are comparable to other potentials, however the deviation from CCSD(*T*) is high and the other spectroscopic parameters are zero by definition.

Table S5† shows that without energy threshold for fitting the analytical potentials to CCSD(*T*), the top ranking ones are the same as fitted with an energy threshold (Table 3). It could be suspected that using a larger energy threshold of 5000 cm<sup>-1</sup> when fitting the potentials, would be advantageous, since the ground state vibration of molecular hydrogen, at 4401.21 cm<sup>-1</sup> (ref. 43) would fall well inside this energy range. However, Table S6† shows that the deviation of frequencies from experiment in fact is larger than that fitted with a lower threshold (Table 3). Finally, it is possible to fit potentials directly on experimental

data. Table S7† shows the RMSD from experimental frequencies for the 14 molecules from Table 1. The deviation for both  $\omega_e$  and  $\omega_e x_e$  is somewhat larger than for the fits to the CCSD(*T*) potential, likely since the ground state vibrations are determined mainly by the shape of the potential close to the minimum. Therefore, fitting a potential to an accurate quantum chemistry calculation may be sufficient if the purpose is to reproduce the vibrational properties listed in Table 3.

## 4 Conclusions

An evaluation of 28 analytical potentials to reproduce quantum chemistry data for 71 diatomic molecules is presented. Several of these potentials fit the quantum chemical reference data excellently (Table 2) and also produce vibrational parameters on par with the CCSD(*T*) ones (Table 3). The potential due to Hua is both accurate and relatively simple and therefore it may be a good choice for implementation in molecular simulation codes. It is given by

$$U(r) = D_e \left[ \frac{1 - e^{-b(r-r_e)}}{1 - c e^{-b(r-r_e)}} \right]^2 \quad (4)$$



where  $D_e$  is the well-depth,  $r_e$  the equilibrium bond length, and  $b$  and  $c$  are constants with  $\|c\| < 1$ . At  $r = r_e$ , the energy is zero, but for MD simulations to reproduce the quantum-chemical energy,  $D_e$  should be subtracted from eqn (4). It has been shown that the Tietz II potential<sup>56</sup> is mathematically identical to the one due to Hua, under the condition that the same  $D_e$  is used,<sup>22</sup> however the Tietz II potential lacks the feature that the energy minimum is zero by definition. For use in MD simulations, the formulation in eqn (4) therefore has the advantage of straightforward interpretation of the meaning of the parameters. The potentials due to Hulburt–Hirschfelder<sup>55</sup> and Sun<sup>54</sup> are more complex and therefore more computationally expensive. In addition, more parameters will make those potentials more cumbersome to parameterize.

Anharmonicity in the frequencies is not reproduced very well by the quantum chemical methods employed here (Fig. 2 and Table 3) and this is reflected in the analytical potentials. Simply stated, these potentials are not better than the reference data, in this case CCSD(T)/aug-cc-pVTZ. Although this is perhaps a trivial conclusion, it applies to both science-driven<sup>1</sup> (like in this work) and data-driven model building, and a careful evaluation of training data is therefore crucial in any machine learning endeavor.<sup>4</sup> A data-science problem that needs to be addressed in both schools of modeling is the range of energies incorporated. We have shown previously<sup>8,73</sup> as well as here that the choice of energy threshold can affect predictive power of models trained on the data. A high threshold, or including a large range of energies, requires complex models to get good results, and simpler models may not be able to compete. A smaller threshold could lead to more complex potentials being underdetermined, but we find rather the opposite, that frequencies are reproduced better when a limited threshold is used (compare Tables 3 and S6†). Finally, it should be noted that the frequencies produced from the MP2 calculations are much less accurate than those from CCSD(T). However, before disregarding MP2 as a basis for systematic design of force fields,<sup>6</sup> studies with larger compounds are needed.

## Data availability

All code and data for this article can be found at: <https://github.com/AlexandriaChemistry/BondPotentials/>. With permanent DOI here: <https://doi.org/10.5281/zenodo.13997971>.

## Author contributions

PvM designed the study, implemented the code to evaluate potentials on quantum chemistry and experimental data and, contributed to analysis and writing. DvdS wrote the first version of the manuscript, contributed to implementations and analysis of the results.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by AI4Research at Uppsala University, Sweden, by the Swedish Research Council (grant 2020-05059) and by eSENCE – The e-Science Collaboration (Uppsala-Lund-Umeå, Sweden). Computer resources were provided by the National Academic Infrastructure for Supercomputing Sweden at the national supercomputing center, partially funded by the Swedish Research Council through (grant 2022-06725). Experimental data reproduced with kind permission of Dr Jorge Luque.

## Notes and references

- 1 J. T. Margraf, *Angew. Chem., Int. Ed. Engl.*, 2023, **62**, e202219170.
- 2 A. Aspuru-Guzik, R. Lindh and M. Reiher, *ACS Cent. Sci.*, 2018, **4**, 144–152.
- 3 O. Willcox, K. E. Ghattas and P. Heimbach, *Nat. Comput. Sci.*, 2021, **1**, 166–168.
- 4 K. Kříž, L. Schmidt, A. T. Andersson, M.-M. Walz and D. van der Spoel, *J. Chem. Inf. Model.*, 2023, **63**, 412–431.
- 5 A. Hagler, *J. Comput.-Aided Mol. Des.*, 2019, **33**, 205–264.
- 6 D. van der Spoel, *Curr. Opin. Struct. Biol.*, 2021, **67**, 18–24.
- 7 L. Wang, P. K. Behara, M. W. Thompson, T. Gokey, Y. Wang, J. R. Wagner, D. J. Cole, M. K. Gilson, M. R. Shirts and D. L. Mobley, *J. Phys. Chem. B*, 2024, **128**, 7043–7067.
- 8 K. Kříž, P. J. van Maaren and D. van der Spoel, *J. Chem. Theory Comput.*, 2024, **20**, 2362–2376.
- 9 T. A. Halgren, *J. Am. Chem. Soc.*, 1992, **114**, 7827–7843.
- 10 J. C. Werhahn, E. Miliordos and S. S. Xantheas, *Chem. Phys. Lett.*, 2015, **619**, 133–138.
- 11 J. E. Jones, *Proc. R. Soc. London, Ser. A*, 1924, **106**, 463–477.
- 12 M. M. Walz, M. M. Ghahremanpour, P. J. van Maaren and D. van der Spoel, *J. Chem. Theory Comput.*, 2018, **14**, 5933–5948.
- 13 L.-P. Wang, J. Chen and T. V. Voorhis, *J. Chem. Theory Comput.*, 2013, **9**, 452–460.
- 14 W. Heitler and F. London, *Z. Phys.*, 1927, **44**, 455–472.
- 15 P. M. Morse, *Phys. Rev.*, 1929, **34**, 57–64.
- 16 R. Rydberg, *Z. Phys.*, 1932, **73**, 376–385.
- 17 G. Pöeschl and E. Teller, *Z. Phys.*, 1933, **83**, 143–151.
- 18 A. T. Royappa, V. Suri and J. R. McDonough, *J. Mol. Struct.*, 2006, **787**, 209–215.
- 19 G.-D. Zhang, J.-Y. Liu, L.-H. Zhang, W. Zhou and C.-S. Jia, *Phys. Rev. A*, 2012, **86**, 062510.
- 20 R. K. Pingak, A. Z. Johannes, Z. S. Ngara, M. Bukit, F. Nitti, D. Tambaru and M. Z. Ndi, *Results Chem.*, 2021, **3**, 100204.
- 21 G. A. Natanson, *Phys. Rev. A*, 1991, **44**, 3377–3378.
- 22 C.-S. Jia, Y.-F. Diao, X.-J. Liu, P.-Q. Wang, J.-Y. Liu and G.-D. Zhang, *J. Chem. Phys.*, 2012, **137**, 014101.
- 23 J. P. Araújo and M. Y. Ballester, *Int. J. Quantum Chem.*, 2021, **121**, e26808.
- 24 J. L. Dunham, *Phys. Rev.*, 1932, **41**, 721–731.
- 25 M. Miotto and L. Monacelli, *npj Comput. Mater.*, 2024, **10**, 240.



- 26 L. K. McKemmish, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **11**, e1520.
- 27 H. Henschel, A. T. Andersson, W. Jaspers, M. Mehdi Ghahremanpour and D. van der Spoel, *J. Chem. Theory Comput.*, 2020, **16**, 3307–3315.
- 28 H. Henschel and D. van der Spoel, *J. Phys. Chem. Lett.*, 2020, **11**, 5471–5475.
- 29 R. J. Shannon, B. Hornung, D. P. Tew and D. R. Glowacki, *J. Phys. Chem. A*, 2019, **123**, 2991–2999.
- 30 J. Murrell and K. Sorbie, *J. Chem. Soc., Faraday Trans.*, 1974, **70**, 1552–1556.
- 31 C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618–622.
- 32 T. H. Dunning Jr and K. A. Peterson, *J. Chem. Phys.*, 2000, **1113**, 7799–7808.
- 33 J. Čížek, *J. Chem. Phys.*, 1966, **45**, 4256–4266.
- 34 K. A. Peterson and K. E. Yousaf, *J. Chem. Phys.*, 2010, **133**, 174116.
- 35 F. Weigend, F. Furche and R. Ahlrichs, *J. Chem. Phys.*, 2003, **119**, 12753–12762.
- 36 B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson and T. L. Windus, *J. Chem. Inf. Model.*, 2019, **59**, 4814–4820.
- 37 J. A. Pople and R. K. Nesbet, *J. Chem. Phys.*, 1954, **22**, 571–572.
- 38 C. C. J. Roothaan, *Rev. Mod. Phys.*, 1960, **32**, 179–185.
- 39 J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, N. J. Russ, M. L. Leininger, C. L. Janssen, E. T. Seidl, W. D. Allen, H. F. Schaefer, R. A. King, E. F. Valeev, C. D. Sherrill and T. D. Crawford, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **2**, 556–565.
- 40 X. Liu, S. Truppe, G. Meijer and J. Pérez-Ríos, *J. Cheminf.*, 2020, **12**, 31.
- 41 Y. Wang, D. Julian, M. A. Ibrahim, C. Chin, S. Bhattiprolu, E. Franco and J. Pérez-Ríos, *J. Mol. Struct.*, 2023, **398**, 111848.
- 42 P. J. Linstrom and W. G. Mallard, *NIST Chemistry WebBook*, NIST, Standard Reference Database Number 69, 2011, <http://webbook.nist.gov>.
- 43 K. P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure*, Springer-Verlag, Berlin, Germany, 1979.
- 44 P. J. van Maaren and D. van der Spoel, Bond potentials software package and data repository, 2025, DOI: [10.5281/zenodo.14842997](https://doi.org/10.5281/zenodo.14842997).
- 45 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt and SciPy 1.0 Contributors, *Nat. Methods*, 2020, **17**, 261–272.
- 46 R. A. Buckingham, *Proc. R. Soc. London, Ser. A*, 1938, **168**, 264–283.
- 47 K. Cahill and V. A. Parsegian, *J. Chem. Phys.*, 2004, **121**, 10839–10842.
- 48 K. T. Tang and J. P. Toennies, *J. Chem. Phys.*, 2003, **118**, 4976–4983.
- 49 R.-H. Xie and J. Gong, *Phys. Rev. Lett.*, 2005, **95**, 263202.
- 50 K. Levenberg, *Q. Appl. Math.*, 1944, **2**, 164–168.
- 51 D. W. Marquardt, *J. Soc. Ind. Appl. Math.*, 1963, **11**, 431–441.
- 52 T. Ho and H. Rabitz, *J. Chem. Phys.*, 1996, **104**, 2584–2597.
- 53 W. Hwang, S. L. Austin, A. Blondel, E. D. Boittier, S. Boresch, M. Buck, J. Buckner, A. Caflisch, H.-T. Chang, X. Cheng, Y. K. Choi, J.-W. Chu, M. F. Crowley, Q. Cui, A. Damjanovic, Y. Deng, M. Devereux, X. Ding, M. F. Feig, J. Gao, D. R. Glowacki, J. E. Gonzales II, M. B. Hamaneh, E. D. Harder, R. L. Hayes, J. Huang, Y. Huang, P. S. Hudson, W. Im, S. M. Islam, W. Jiang, M. R. Jones, S. Käser, F. L. Kearns, N. R. Kern, J. B. Klauda, T. Lazaridis, J. Lee, J. A. Lemkul, X. Liu, Y. Luo, A. D. MacKerell Jr, D. T. Major, M. Meuwly, K. Nam, L. Nilsson, V. Ovchinnikov, E. Paci, S. Park, R. W. Pastor, A. R. Pittman, C. B. Post, S. Prasad, J. Pu, Y. Qi, T. Rathinavelan, D. R. Roe, B. Roux, C. N. Rowley, J. Shen, A. C. Simmonett, A. J. Sodt, K. Töpfer, M. Upadhyay, A. van der Vaart, L. I. Vazquez-Salazar, R. M. Venable, L. C. Warrensford, H. L. Woodcock, Y. Wu, C. L. Brooks III, B. R. Brooks and M. Karplus, *J. Phys. Chem. B*, 2024, **128**, 9976–10042.
- 54 W. Sun, *Mol. Phys.*, 1997, **92**, 105–108.
- 55 H. Hulburt and J. Hirschfelder, *J. Chem. Phys.*, 1941, **9**, 61–69.
- 56 T. Tietz, *Can. J. Phys.*, 1971, **49**, 1315.
- 57 F. M. Rafi, *Phys. Lett. A*, 1995a, **205**, 383–387.
- 58 I. Levine, *J. Chem. Phys.*, 1966, **45**, 827–828.
- 59 S. Noorizadeh and G. Pourshams, *J. Mol. Struct.*, 2004, **678**, 207–210.
- 60 W. Hua, *Phys. Rev. A*, 1990, **42**, 2524–2529.
- 61 A. Frost and B. Musulin, *J. Chem. Phys.*, 1954, **22**, 1017–1020.
- 62 Y. Varshni, *Can. J. Chem.*, 1988, **66**, 763–766.
- 63 N. Rosen and P. Morse, *Phys. Rev.*, 1932, **42**, 210–217.
- 64 M. Sage, *Chem. Phys.*, 1984, **87**, 431–439.
- 65 J. Linnett, *Trans. Faraday Soc.*, 1940, **36**, 1123–1134.
- 66 Z. H. Deng and Y. P. Fan, *Shandong Univ. J.*, 1957, **7**, 162.
- 67 D. Gardner and L. von Szentpály, *J. Phys. Chem. A*, 1999, **103**, 9313–9322.
- 68 A. Kratzer, *Z. Phys.*, 1920, **3**, 289–307.
- 69 E. Lippincott, *J. Chem. Phys.*, 1953, **21**, 2070–2071.
- 70 R. J. Le Roy, *J. Quant. Spectrosc. Radiat. Transfer*, 2017, **186**, 167–178.
- 71 S. N. Yurchenko, L. Lodi, J. Tennyson and A. V. Stolyarov, *Comput. Phys. Commun.*, 2016, **202**, 262–275.
- 72 M. A. E. Ibrahim, X. Liu and J. Pérez-Ríos, *Digital Discovery*, 2024, **3**, 34–50.
- 73 K. Kříž and D. van der Spoel, *J. Phys. Chem. Lett.*, 2024, **15**, 9974–9978.

