# Journal Name

# Intrinsic Bond Strength Index as a halogen bond interaction energy predictor[†]

Ona Šivickytė,[a,b] and Paulo J. Costa,[a,b]*

Halogen bonds (XBs) have become increasingly popular over the past few years with numerous applications in catalysis, material design, anion recognition, and medicinal chemistry. To avoid a *post factum* rationalization of XB trends, descriptors can be tentatively employed to predict the interaction energy of potential halogen bonds. These typically comprise the electrostatic potential maximum at the tip of the halogen, $V_{S,max}$, or properties based on the topological analysis of the electronic density. However, such descriptors either can only be used with confidence for specific families of halogen bonds or require intense computations and, therefore, are not particularly attractive for large datasets with diverse compounds or biochemical systems. Therefore, the development of a simple, widely applicable, and computationally cheap descriptor remains a challenge as it would facilitate the discovery of new XB applications while also improving the existing ones. Recently, the Intrinsic Bond Strength Index (IBSI) has been proposed as a new tool to evaluate any bond strength, however, it has not been extensively explored in the context of halogen bonding. In this work, we show that IBSI values linearly correlate with the interaction energy of diverse sets of closed-shell halogen-bonded in the ground state, and therefore, can be used to quantitatively predict this property. Although the linear fit models that use quantum-mechanics-based electron density provided MAEs typically below 1 kcal mol$^{-1}$, this type of calculation might still be computationally heavy in large sets or systems. Therefore, we also explored the exciting possibility to use a promolecular density approach (IBSI$^{PRO}$), which only requires the geometry of the complex as an input, being computationally cheap. Surprisingly, the performance was comparable to the QM-based methods, thus opening the door for the usage of IBSI$^{PRO}$ as a fast, yet accurate, XB energy descriptor in large datasets but also in biomolecular systems such as protein–ligand complexes. We also show that $\delta g^{pair}$ descriptor emerging from the Independent Gradient Model that leads to IBSI can be seen as a term proportional to the overlapping van der Waals volume of the atoms at a given interaction distance. Overall ISBI can be thought of as a complementary descriptor to $V_{S,max}$ for situations when the geometry of the complex is available and QM calculations are not feasible whereas the latter still remains a hallmark of XB descriptors.

## 1 Introduction

A halogen bond (XB) is a directional non-covalent interaction between a Lewis base (B), *e.g.*, the lone pairs on an N-, O-containing molecule, and a halogen atom (X) in a molecular entity acting as a Lewis acid (R–X···B)[1–3]. Indeed, while typically halogens are perceived as electron-rich electronegative species behaving as nucleophiles, the picture is more complicated when they are covalently bound to another atom (R–X) as the electrons are anisotropically distributed, forming regions of higher and lower electron density (ED). The region of lower ED located at the tip of the halogen opposite to the covalent bond corresponds to the so-called $\sigma$-hole[4]. This site is typically electropositive and can interact with nucleophiles, thus offering an electrostatic explanation for the formation of XBs. A larger polarizability of X corresponds to a larger $\sigma$-hole and consequently to a stronger XB, and therefore, the XB interaction energy typically increases along the halogen series: Cl < B < I[5,6]. A seemingly opposing view describes XBs as charge-transfer (CT) complexes explained by the existence of electron transfer from a filled donor orbital of the Lewis base to the accepting R–X $\sigma^*$ orbital of the halogenated molecule[3,7,8], following the same trends as mentioned above.

$^a$ *BioISI - Instituto de Biosistemas e Ciências Integrativas, Faculdade de Ciências, Universidade de Lisboa, 1749-016, Lisboa, Portugal; E-mail: pjcosta@ciencias.ulisboa.pt*
$^b$ *Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal*

However, while both views might reveal different sides of a dual XB nature[9], it has been argued that both essentially describe the same phenomenon[10,11] or that CT is practically negligible for the overall interaction[12].

XBs are seen as hydrophobic counterparts for hydrogen bonds (HBs), but they are often considered to be more versatile[13] since halogen atoms can act as both a Lewis base (HB acceptor) and a Lewis acid (XB donor). This versatility also arises from their directionality and tunability as the XB length, the R–X···B angle, and the magnitude of the $\sigma$-hole largely depend on the halogen, the existence of other substituents on the XB donor, and the nature of the Lewis base[13–15]. All these factors can easily be chosen or adjusted to meet a set of unique specifications thus, XBs span a wide range of interaction energies[13,16,17]. In principle, such factors could be represented by a combination of descriptors of electronic and/or electrostatic effects which could be used to estimate the interaction energies of of XBs[7,18]. However, it is also admitted that the applicability of XBs is often rationalized *post factum* as it remains a challenge to accurately predict the outcomes and their interaction energy in complex systems (*e.g.* protein-ligand) and thus, we are still far from taking full advantage of XBs in the rational design of new systems[19–21], even though various potential applications are constantly emerging[22–25]. There have been attempts to overcome this challenge and provide a solid basis for designing new halogen-bonded structures[8,21,26–28], but accurate modelling of XBs is still not straightforward. This issue is paramount given the increased attention put on XBs and and their broad application in catalysis[19,29–32], material design[33–35], supramolecular[36–38] and medicinal[39–41] chemistry, among other areas.

Several approaches allow the estimation of the XB interaction energies and this topic is tightly related to the discussion regarding the nature of this interaction and the importance of various bonding components to the overall XB stability[12,42]. The most commonly employed XB interaction energy descriptor is the electrostatic potential maximum at the tip of the halogen ($V_{S,max}$), *i.e.* the magnitude of the $\sigma$-hole. This simple, yet powerful descriptor with a clear physical interpretation encompasses only the electrostatic component of XBs and does not always adequately predict the interaction energy[43–45]. This occurs mainly due to its static nature as it is computed in the absence of the base, thus neglecting the contribution from the XB acceptor. It can be corrected by adding polarization to the static $V_{S,max}$, evaluating its magnitude in the presence of a negative point charge, yielding an extended electrostatic model[46–48]. It has also been proposed that the minima of the local attachment energy, analogous to the average ionization energy but reflecting the susceptibility towards nucleophiles, can be used to complement $V_{S,max}$ or as an independent descriptor to predict XB energies in methyl- and aryl halides[49]. Alternatively, some authors approached the incompleteness of the $V_{S,max}$ by combining it with CT descriptors such as the charge transfer energy[45,50] or the C–X $\sigma^*$ orbital energy[8], often leading to improved XB energy predictions[8]. Other attempts to predict XB interaction energy include the usage of ED properties such as the kinetic, potential, and total energy density[51,52], also its Laplacian and curvature[53], evaluated at the bond critical point.

All the mentioned approaches typically require *ab initio* or DFT calculations to obtain the descriptors and therefore, could be computationally demanding, hindering their application in large datasets or large molecules such as protein-ligand systems[54]. Machine-learning (ML) approaches could offer an alternative, as highlighted by a statistical model trained against high–accuracy *ab initio* calculations, which depends on only two fitted parameters along with the equilibrium distance. This model, whose computational cost is negligible, outperforms some of the best density functionals[55]. However, the physical interpretation of fitted ML parameters is often not straightforward.

The above considerations indicate that the need to develop more straightforward and easily accessible XB energy estimators persists. In this context, the Intrinsic Bond Strength Index (IBSI)[56], emerging from the Independent Gradient Model (IGM)[57,58], can evaluate the strength of the interaction between a given pair of atoms which could prove useful in the context of XBs. It is a score that allows to quantitatively compare interactions and estimate their nature, *i.e.*, distinguish covalent from non-covalent interactions, based on threshold values. Although methods relying on topological analysis of the ED are common in identifying and characterizing chemical bonds, *e.g.*, electron localization function, these are often not able to quantify interactions[59]. In contrast, with IBSI, the quantification is outstandingly easy to interpret and is becoming a common tool to evaluate other types of interactions[60–62]. However, despite a few XB complexes being included in the original study[56], IBSI has not yet been systematically explored in the context of these interactions. Herein we report an exploratory study on how IBSI can be used to fairly predict XB interaction energies. Most strikingly, we will show that IBSI values calculated using a promolecular approach that does not require any QM calculation, also linearly correlate with interaction energies while providing similar accuracy. These exploratory results open the door for the development of fast methods to estimate XB energies in large datasets and/or biomolecular systems, and also for the usage of IBSI as a fast-obtainable XB feature for ML models.

## 2 Methods

### 2.1 IGM and IBSI

Herein, a succinct overview of the IBSI approach is given. Further details can be found in the original publications[56–58]. The concept of IBSI originates from the Independent Gradient Model (IGM)[57,58] which can be viewed as an extension of the NCI analysis method[63]. NCI is based on the analysis of the reduced density gradient $s$ (also denoted RDG). However, the NCI approach has a semiquantitative character since the integration of quantities over NCI regions is not trivial[57]. On the contrary, the IGM approach, by providing a virtual non-interacting reference system[58], allows quantification of the interactions.

For a system with interacting fragments A and B and with electron density (ED) $\rho$, the absolute value of each ED gradient component can be expressed as the sum of the individual terms (Equa-

tion 1).

$$\left|\frac{\partial \rho^{pair}}{\partial \mu}\right| = \left|\sum_{i=A,B}^{N} \frac{\partial \rho_i}{\partial \mu}\right| \quad (\mu = x,y,z) \qquad (1)$$

Then, as mentioned above, a virtual non-interacting reference system can be defined as

$$\left|\frac{\partial \rho^{IGM,pair}}{\partial \mu}\right| = \sum_{i=A,B}^{N} \left|\frac{\partial \rho_i}{\partial \mu}\right| \quad (\mu = x,y,z) \qquad (2)$$

Here, the absolute values of the individual fragment components are summed. Since the individual signs of the individual ED derivatives are ignored in the summation, the values will not suffer the attenuation usually observed from the addition of two atomic ED gradients having opposite signs in the region between the interacting fragments A and B. Therefore, this quantity can be considered the upper limit of the true ED gradient (virtual non-interacting reference). From the resulting norm $|\nabla\rho^{pair}|$, obtained from the derivatives of the true ED (Equation 1), along with the norm $|\nabla\rho^{IGM,pair}|$ (Equation 3), the $\delta g^{pair}$ descriptor emerges

$$\delta g^{pair} = |\nabla\rho^{IGM,pair}| - |\nabla\rho^{pair}| \qquad (3)$$

which is a unique bond signature that precisely quantifies the net ED gradient collapse due to the interaction between any pair of interacting atoms/fragments, *i.e*, $\delta g^{pair} \neq 0$ is exclusive of interaction situations. Additionally, $\delta g$ can be plotted against the ED multiplied by the sign of the second eigenvalue of the ED hessian matrix, $sign(\lambda_2)\rho$, producing plots analogous to those obtained in NCI analyses, allowing to discriminate if $\delta g^{pair}$ occurs in attractive ($\lambda_2 < 0$) or repulsive ($\lambda_2 > 0$) regions.

The above corresponds to the original formulation of IGM[57] which is implemented in IGMplot[64] (see Computational details). A different definition of the non-interacting reference and its norm has been proposed in MultiWFN[65] as follows:

$$|\nabla\rho^{pair}| = |\nabla\rho_A| + |\nabla\rho_B| \qquad (4)$$

Notice that this new definition is based on the sum of the norms of the fragments and not on the ED derivative component ($x$, $y$, $z$), likely leading to situations where $\delta g \neq 0$ in non-interacting states. In spite only very small differences in bond signatures are observed when using both formulations[66], the reader is should be aware that another definition than the original IGM is provided in MultiWFN.

With the $\delta g^{pair}$ descriptor in hand, its integral over the interaction volume divided by the square of the internuclear distance $d$ can be used as a global score for a given bond

$$\Delta g^{pair} = \int_V \frac{\delta g^{pair}}{d^2} dV \qquad (5)$$

$\Delta g^{pair}$ is a bond index by itself, however, in order to obtain a score comparable between bond indices and molecules, it has to be normalized for the $H_2$ molecule, thus yielding the Intrinsic Bond Strength Index (IBSI):

$$IBSI = \frac{\Delta g^{pair}}{\Delta g^{H_2}} \qquad (6)$$

IBSI is a dimensionless value that does not correspond to a bond order but reflects the bond strength[56].

## 2.2 ED and partition schemes

IGM and IBSI are dependent on the ED ($\rho$) and the partition scheme used to assign its gradient to individual atoms/fragments. Originally, IGM was developed specifically for promolecular ED-based calculations[57] and this promolecular density is obtained from a sum of simple exponential atomic functions fitted to averaged *ab initio* atomic electron densities. Even though the obtained gradient is approximate as it lacks relaxation, the accuracy is reasonable as long as it is used in the non-covalent regime[57]. This approach is very attractive since minimal computational resources are required and only the geometry is used as input. Given its simplicity, the partition of the total ED gradient into atomic contributions is straightforward. The calculation of IBSI values from promolecular densities (here denoted as IBSI[PRO]) was not considered in the original implementation[56,64] which used QM-based densities and a Gradient-Based Partitioning scheme (see below). However, such calculation is implemented in the MultiWFN package[65] (see Computational Details) despite the different definition of IGM, as stated above.

Another approach takes advantage of the ED obtained from QM calculations. In this case, the ED is in principle more accurate but the partition of the total gradient is not trivial. The Gradient-Based Partitioning (GBP) scheme was introduced in the context of the original IGM definition[58] and is implemented in IGMplot[64]. This method proposes that each individual gradient element $\partial \rho_i / \partial_x$ can be assigned to an atomic orbital $\varphi_i$ and IBSI values calculated within this approach are henceforth termed IBSI[GBP] and used herein as a *gold standard* for QM-based methods.

Recently, an Hirshfeld partition of the ED featuring an hybrid QM/promolecular approach was proposed[67] and implemented in the MultiWFN package[65]. Hirshfeld partition is a very common method to obtain atomic densities[68] and in principle, allows the calculation of all quantitative indices within IGM, including IBSI (here denoted IBSI[H]). However, the sign of the gradient of the free-state atomic density, calculated by MultiWFN is inverted[66,69]. The authors surprisingly argued that this model performs significantly better than correct one[69], which is highly questionable since it has no physical basis and is based on a calculation error. Nonetheless, herein we provide calculated IBSI[H] values (mostly in ESI† and for comparison purposes) and the reader is advised to take them with care.

## 2.3 Data sets

To test a possible correlation of IBSI with XB interaction energies we used 3 data sets containing various X-bonded systems with available optimized equilibrium geometries and energies obtained from high-level QM calculations.

**Set 1** was taken from the Non-Covalent Interactions Atlas, a library containing accurate benchmark non-covalent interaction energies ($E_{int}$)[70]. It comprises halogen-bonded systems, optimized at the B3LYP-D3(BJ)/def2-QZVP level, featuring small

molecules with Cl, Br, and I as XB donors, and various XB acceptors bearing O, N, P, and S, such as acetonitrile, pyrazine, acetone, thiacetone, and molecular halogens. The X-bonded compounds in this library were chosen to cover a wide range of $\sigma$-hole magnitudes and each fragment contains no more than 13 atoms. The benchmark interaction energies, which do not include the deformation energy of the fragments, were calculated using a composite CCSD(T)/CBS scheme based on MP2 and CCSD(T) calculations with very large basis sets. Herein we directly excluded X$\cdots\pi$ bonds because they cannot be unambiguously described by a single IBSI value, therefore yielding a final set of 99 complexes (see Table S1 in ESI†).

**Set 2** consists of A–X$\cdots$B systems, where A = H, F, Cl, Br, I, and X = F, Cl, Br, I taken from reference 55. The data contained originally 140 high-accuracy *ab initio* benchmark interaction energies (CCSD(T)-F12b/CBS) calculated on CCSD(T)-F12b/VTZ-(PP)-F12 optimized structures whose geometry is available. Again, $E_{int}$ values do not include the deformation energy of the fragments. In this work, only 124 of those systems were used (see Table S2 in ESI†) since 10 complexes containing X$\cdots\pi$ contacts were removed for the same reason mentioned above for **Set 1** along with those containing $F_2$ as a XB donor. Notice that fluorine is typically not considered a XB donor and fluorine interactions are fundamentally different from typical XBs[71].

**Set 3** was taken from reference 72 which revises and corrects some values earlier reported for the XB18 and XB51 benchmarking sets[73]. These benchmarks consist of 69 systems bearing only neutral fragments with Cl-, Br-, and I-containing molecules as XB donors, and N, O, P, and Cl as acceptor atoms. XB18 contains 18 systems with NCH and $OCH_2$ as acceptors. Here, the geometries were optimized at CCSD(T)/aVQZ, and dissociation energies ($E_{diss}$) were calculated at the CCSD(T)/CBS level of theory. Notice that $E_{diss}$ values take into account the deformation of the fragments, thus allowing to infer the potential effect of geometry relaxation in the models. XB18 was intentionally constructed using only small molecules so that highly accurate calculations could be easily performed. The XB51 is an extended version of XB18 and includes a wider range of both donor and acceptor fragments. The geometry optimizations for XB51 were performed at $\omega$B97X/aVTZ level of theory with single-point energies computed using an MP2-based extrapolation of the CCSD(T) energy. Herein, we merged both XB18 and XB51 and in cases where dissociation energies and geometries were available from both, the data were taken from XB51, yielding a total of 64 complexes (see Table S3 in ESI†). In order to maintain the consistency of the sign between Sets, the energy values reported for these data sets correspond to $-E_{diss}$.

### 2.4 Computational details

All QM calculations were performed using the Gaussian 09 program package[74]. Since optimized geometries were available, to obtain the QM-based ED for the IBSI estimation (IBSI$^{GBP}$ and IBSI$^H$), single-point calculations were performed at the DFT M06-2X/def2-TZVP level of theory[75] in the gas phase with the associated effective core potential for iodine. This func-

tional is commonly applied in XB studies[8,31,76] with good performances[70,73,77] and is also recommended by the IBSI method[56]. Additional calculations using def2-SVPD[75,78] and def2-QZVP[79] were also carried out in order to evaluate the significance of the basis set (see below). An ultrafine integration grid was applied in all the calculations. Checkpoint (chk) or wave function files (wfn/wfx) were stored for further analysis and calculation of IBSI. IBSI$^{PRO}$ and IBSI$^H$ were calculated with MultiWFN 3.7[65]. As mentioned above, IBSI$^{PRO}$ only required the optimized geometry while for IBSI$^H$, the M06-2X/def2-TZVP wave function file was provided. In both cases, the reported values are normalized to $H_2$ by the $\Delta g^{H_2}$ value obtained in the same conditions. IBSI$^{GBP}$ values were obtained with IGMplot 2.6.9b[64] using the same wave function files. Herein, the values are internally normalized for $H_2$ at the M06-2X/6-31G** level of theory and no re-normalization was performed for M06-2X/def2-TZVP values. Notice that this does not have any impact on the statistics of the obtained linear-fit models.

### 2.5 Statistical analysis

The data in the three sets were fitted separately to the following equations for **Sets 1-2** and **Sets 3**, respectively

$$E_{int} = m \times IBSI + b \tag{7}$$

$$-E_{diss} = m \times IBSI + b \tag{8}$$

via the $m$ and $b$ parameters using an ordinary least squares (OLS) regression model. The quality of the fit was analyzed by evaluating the coefficient of determination $R^2$, the Pearson correlation coefficient $r$, the Spearman's Rank Correlation Coefficient $\rho$, and the Kendall rank correlation coefficient ($\tau$) using in-house python scripts. The Mean Absolute Error (MAE) was employed as a performance metric of each model. Before the fitting stage, an explanatory data analysis (EDA) was performed to characterize each set. Multivariate outliers, *i.e.* the unusual combination of $E_{int}$ (or $-E_{diss}$) and IBSI values, were discarded using the minimum covariance determinant (MCD) method[80,81] with a significance level threshold of 0.001 using in-house python scripts. This is a highly robust estimator of multivariate location and scatter as the MCD is computed using only a subset of the sample, thus, the outlying points will have a small impact on the MCD location or shape estimate. Further information can be found below.

## 3 Results and discussion

### 3.1 Basis set influence on IBSI

In the original IBSI implementation based on GBP the authors showed that IBSI$^{GBP}$ values are typically independent of the method and basis set, therefore, stable results are expected as long as the same method is used for comparative studies[56]. However, we are not aware how the magnitude of the IBSI$^{GBP}$ values compare with the IBSI$^{PRO}$ approach. Moreover, and despite the possible drawbacks that IBSI$^H$ might have (see above), it would also be important to assess the sensitivity of this method to wave function quality. A low dependence on the basis set was claimed in the original paper though the data was not dis-

closed[67]. We therefore selected 3 complexes from **Set 3** featuring a strong (FI···pyr, -20.34 kcal mol$^{-1}$), a mild (FI···OPH$_3$, -13.36 kcal mol$^{-1}$) and a weak (FI···PCH, -2.74 kcal mol$^{-1}$) XB, and calculated IBSI$^{GBP}$, IBSI$^{PRO}$, and IBSI$^H$ values (Table 1). For both QM-based methods, the values obtained vary little with the basis set. A slight noticeable deviation is found for the strongest XB, however, the values are still acceptable. For the weaker XBs, the values remain almost constant. When comparing the approaches, the magnitude of the values is quite different. Using IBSI$^{GBP}$ as a reference, IBSI$^{PRO}$ values are larger. The difference is understandable owing to the lack of ED relaxation in IBSI$^{PRO}$. Surprisingly, IBSI$^H$ values are very low when compared with the reference IBSI$^{GBP}$ meaning that the indicative threshold of the non-covalent domain (0.15) for IBSI$^{GBP}$ [56] is not applicable to IBSI$^H$ which runs on a completely different scale. This might be due to the questionable nature of the implementation of this partition scheme (see above). We note however that a robustness test of IBSI$^H$ is out of the scope of this manuscript, thus, we still provide the reader with IBSI$^H$ values in ESI† while the following discussion on QM-based ED will be based on the robust IBSI$^{GBP}$ values.

## 3.2 IBSI$^{GBP}$ linearly correlates with interaction and dissociation energies of halogen-bonded complexes

Although XBs were explored in the original IBSI reference[56], a real systematic study for this type of non-covalent bond is yet to be performed. Herein, we explore if a "simple" index such as IBSI$^{GBP}$ linearly correlates with the interaction energy ($E_{int}$, **Set 1–Set 2**) or dissociation energies ($-E_{diss}$, **Set 3**) for large and diverse sets of halogen-bonded systems taken from the literature.

**Set 1**, which besides dihalogens and acetone also include cyclic (pyrazine, pyridine-N-oxide) and sulfur-containing (thioacetone, dimethylthioether) acceptors, span a wide range of $E_{int}$ values, from very weak XBs (Cl$_2$···F$_2$, $-0.66$ kcal mol$^{-1}$) to strong (Br$_2$···N(CH$_3$)$_3$, $-17.14$ kcal mol$^{-1}$) ones (Table S1 in ESI†). However, the distribution of the energies is skewed towards the weak interactions (Figure S1 in ESI†) thus, this Set is more representative of weak to moderate XBs (median = $-2.97$ kcal mol$^{-1}$). The same skewed distribution, is observed for **Set 2** (Figure S2 in ESI†). **Set 2** comprises dihalogen and hydrogen halide XB donors paired up with common XB acceptors, mostly small molecules such as NH$_3$, CH$_2$O, and H$_2$O (Table S2 in ESI†), the energies ranging from $-20.51$ kcal mol$^{-1}$ (FCl···PH$_3$) to $-1.28$ kcal mol$^{-1}$ (HBr···PH$_3$) with a median of $-5.52$ kcal mol$^{-1}$. Finally, **Set 3** possess the weakest and the strongest XBs (FI···FCCH, $-0.29$ kcal mol$^{-1}$ and FI···HLi, $-33.79$ kcal mol$^{-1}$), however, notice that these are $-E_{diss}$ values (thus including a relaxation penalty). The distribution of $-E_{diss}$ is also skewed (Figure S3 in ESI†), with a data gap between the very strong XBs and the remaining values. As for the other sets, **Set 3** is more representative of weak to moderate XBs (median = $-4.17$ kcal mol$^{-1}$).

Although plotting $E_{int}$ or $-E_{diss}$ as a function of IBSI$^{GBP}$ for the full datasets shows a fair linear correlation between the two properties (Figure S4 in ESI†), outliers are easily identified, even by visual inspection. Those were removed using the minimum covariance determinant (MCD) method as described in Compu-

tational details. Further discussion regarding the outliers can be found below, however, it must be pointed out that all remaining points in the datasets possess IBSI values below the threshold of the non-covalent domain (0.15) as defined in reference 56 except for FI···pyridine in **Set 3** ($-E_{diss}$ = -20.34 kcal mol$^{-1}$ and IBSI$^{GBP}$ = 0.154). Using outlier-free datasets, the plots of $E_{int}$ (**Set 1** and **Set 2**) or $-E_{diss}$ (**Set 3**) as a function of IBSI$^{GBP}$ show a strong linear correlation between the variables (Figure 1) with $R^2 \approx 0.9$ and $|r| > 0.93$. Additionally, $\rho$ and $\tau$ indicate a monotonic association between the variables. The final fitted parameters can be found in Table S4 in ESI†. Noticeably, the intercepts $b$ are $\approx 0$, however, negative for **Set 1-Set 2** and positive for **Set 3**. This could be justified by lack of fragment relaxation in the former sets. Indeed, when X···B IBSI$^{GBP}$ = 0, an attractive residual $E_{int}$ value persists due to ancillary interactions (**Set 1-Set 2**) while when $-E_{diss}$ is used, the inherent positive reorganization energy from the fragments is added, leading to a positive intercept. The performance of the models for each dataset is acceptable with MAE values $< 1.1$ kcal mol$^{-1}$. The worse performance was obtained for **Set 3** that is again connected to the fact that IBSI$^{GBP}$ is an intrinsic bond strength[56] which does not account for geometric and electronic reorganization upon dissociation. The best performance was obtained for **Set 1** taken from the Non-Covalent Interactions Atlas[70], with an MAE of 0.6 kcal mol$^{-1}$. For all sets, larger deviations between the predicted and reference data are observed for stronger XBs (Figure S5 in ESI†), the effect being more visible in **Set 3**. Nonetheless, the deviations are fairly distributed around zero (Figure S6 in ESI†) meaning that no obvious under or overestimation of the predicted values is observed, despite a very slight skewness is observed towards negative deviations in **Set 3**. We also checked the correlation of $E_{int}$ (**Set 1-Set 2**), or $-E_{diss}$ (**Set 3**) with IBSI$^H$. Despite strong correlations are also obtained (Figure S7 in ESI†), the slopes are completely different (Table S4 in ESI†) reflecting the above-mentioned IBSI$^H$ scale.

## 3.3 IBSI$^{PRO}$ as a fast XB interaction energy predictor

In the previous section, we showed that IBSI values obtained using QM-based electron densities (IBSI$^{GBP}$) linearly correlate with XB interaction energies ($E_{int}$) and dissociation energies ($-E_{diss}$) for diverse sets of halogen-bonded complexes, though a larger MAE is observed for the later probably owing to the introduction of the relaxation penalty in the energy. This type of linear relationship can be useful, for instance, to estimate high-level *ab initio* energy values using DFT geometries and EDs. However, such a task still requires the usage of QM-based electron density which could be unpractical not only for large datasets of small molecules but also for biomolecular systems. Therefore, we wondered if IBSI values, calculated using a promolecular approach (IBSI$^{PRO}$) and thus neglecting electron relaxation (among other terms), could also be used similarly. Notice that for the covalent regime, the promolecular ED underestimates the troughs of the ED gradient, hence not describing the bonds correctly[58]. Owing to the disputed varying degree of covalency involved in XBs, it is important to understand if promolecular ED describes them cor-

Table 1 Calculated IBSI$^{GBP}$, IBSI$^{H}$, and IBSI$^{PRO}$ values for 3 halogen bonded systems taken from Set 3. The ED was obtained at the M06-2X/b (b = def2-SVPD, def2-TZVP, and def2-QZVP) level. The reported energies ($-E_{diss}$) are CCSD(T)/CBS values from reference 72

| System | $-E_{diss}$ / kcal mol$^{-1}$ | IBSI$^{GBP}$ | | | IBSI$^{H}$ | | | IBSI$^{PRO}$ |
|---|---|---|---|---|---|---|---|---|
| | | def2-SVPD | def2-TZVP | def2-QZVP | def2-SVPD | def2-TZVP | def2-QZVP | |
| FI···PCH | -2.74 | 0.053 | 0.051 | 0.052 | 0.016 | 0.016 | 0.016 | 0.097 |
| FI···OPH$_3$ | -13.36 | 0.109 | 0.102 | 0.107 | 0.038 | 0.041 | 0.041 | 0.182 |
| FI···pyr | -20.34 | 0.175 | 0.154 | 0.141 | 0.057 | 0.062 | 0.061 | 0.249 |
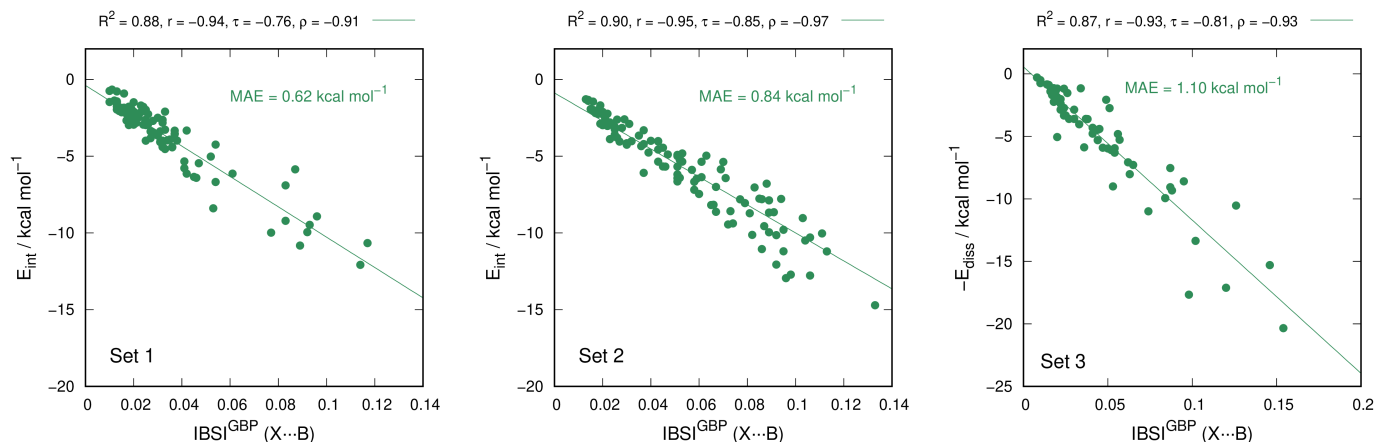


Fig. 1 Interaction energies ($E_{int}$, Set 1-Set 2) or dissociation energies ($-E_{diss}$, Set 3) as a function of IBSI$^{GBP}$. The number of points used for the linear fit was 94, 120, and 62 respectively.

rectly. It is also true that as long the complexes stay in the weak to mild non-covalent regime, the promolecular approach could be enough to capture the correct bond pattern. Indeed, in **Set 1** and after outlier removal (for the full set see Figure S8 in ESI†), the correlation between IBSI$^{PRO}$ values and $E_{int}$ is linear ($R^2 = 0.87$, $|r| = 0.93$) (see Figure 2 left and Table S4 in ESI† for the fitted parameters). Also, the other coefficients, $\rho$ and $\tau$, show a strong monotonic relationship between $E_{int}$ and IBSI$^{PRO}$. In fact, comparing IBSI$^{PRO}$ with IBSI$^{GBP}$, the difference in linearity ($R^2$, $|r|$) is negligible while the MAE is very similar (0.70 kcal mol$^{-1}$ *vs* 0.62 kcal mol$^{-1}$, respectively). Similarly to IBSI$^{GBP}$, IBSI$^{PRO}$ does not strongly over- or underestimate interaction energies, while larger errors are typically associated with stronger XBs (Figure S9 in ESI†). As for QM-based methods, the error is fairly normally distributed around zero (Figure S10 in ESI†). This suggests that using approximate promolecular densities may result in similar accuracy compared to QM methods, even for moderate to strong XBs.

**Set 2** is slightly larger than **Set 1** but the linear correlation between $E_{int}$ and IBSI$^{PRO}$ (Figure 2 center) is again strong ($R^2 = 0.91$, $|r| = 0.95$) and equivalent to that found for IBSI$^{GBP}$. The final fitted parameters can be found in Table S4 in ESI†. The MAE value (0.79 kcal mol$^{-1}$) is lower than that for IBSI$^{GBP}$ (0.84 kcal mol$^{-1}$), thus showing that IBSI$^{PRO}$ can outperform QM-based methods in such simple linear relationships. The difference between estimated and true $E_{int}$ is close to normally distributed around zero (Figure S10 in ESI†), and there is no significant tendency towards consistently over- or underestimating $E_{int}$ values.

For **Set 3**, the final linear correlation between IBSI$^{PRO}$ and dissociation energies ($-E_{diss}$) is again strong ($R^2 = 0.88$, $|r| = 0.94$, see Figure 2, right), with an MAE value of $\approx 1$ kcal mol$^{-1}$, slightly outperforming the QM-based method IBSI$^{GBP}$. Noticeably, the performance of the linear model for **Set 3**, as evaluated by the MAE value, is again the worse, therefore indicating that including the relaxation of fragments can lead to deterioration of the model, which is expected given the nature of the IBSI score. The deviations are very close to normally distributed (Figure S10 in ESI†), the error increasing with increasing XB stability (Figure S9 in ESI†). The final fitted parameters are listed in Table S4 in ESI†.

The above results suggest that, overall, this quite simple model which uses a promolecular density was able to adequately predict interaction energies (or dissociation energies) in these fairly large and diverse datasets. It is also remarkable that the compounds that were poorly described by IBSI$^{GBP}$ (outliers) are also recurrently observed as outliers with IBSI$^{PRO}$ (see also the following Section and ESI†). It could be expected that larger deviations would be typically observed with increasing XB stability when using IBSI$^{PRO}$ owing to the lack of relaxation of the ED which becomes important in the covalent regime. However, such a tendency (Figure S9 in ESI†) is also observed in IBSI$^{GBP}$ (Figure S5 in ESI†). Therefore, it becomes evident that the accuracy of a simple linear regression estimator based on promolecular properties is, at least, very similar to that obtained by methods that require rigorous QM calculations for these types of halogen bonded complexes (neutral, ground state, and closed-shell). The fact that, with a few exceptions, all complexes are bellow non-covalent domain threshold (IBSI < 0.15) and thus the bond signatures from
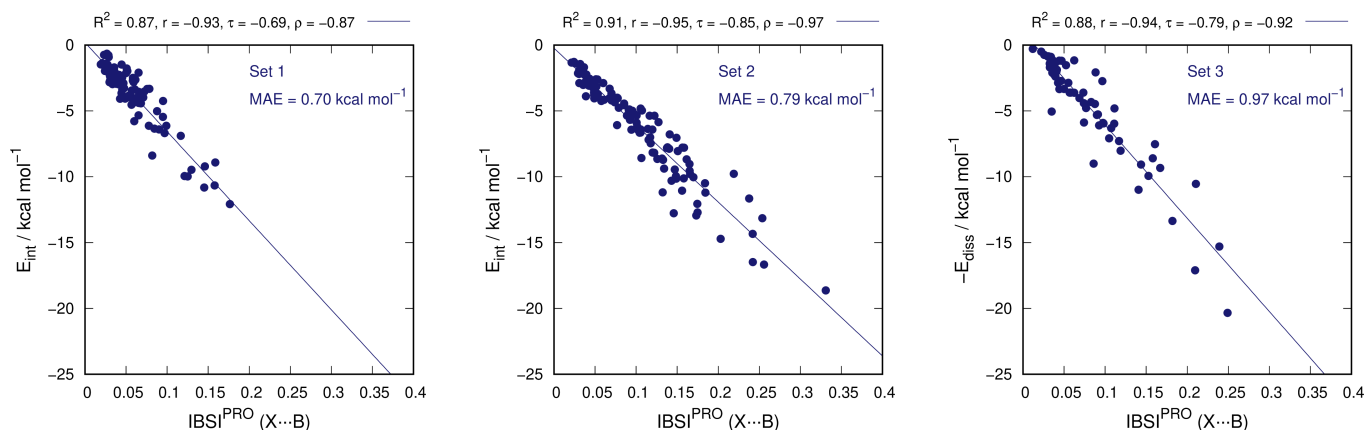
Fig. 2 Interaction energies ($E_{int}$, Set 1-Set 2) or dissociation energies ($-E_{diss}$, Set 3) as a function of IBSI$^{PRO}$. The number of points used for the linear fit was 93, 120, and 61 respectively.

promolecular and GBP approaches should be similar, might also explain the almost identical accuracy. This means that for large systems such as protein-ligand complexes where high-level QM calculations are often not feasible, IBSI$^{PRO}$ can be a fast and reliable solution, provided that proper calibration curves are available.

### 3.4 Why are some complexes outliers? The limits of IBSI as a halogen bond $E_{int}$ predictor

As referred above, the application of the the minimum covariance determinant (MCD) method [80,81] to Set1-Set3 with a significance level threshold of 0.001 identified several unusual combinations of $E_{int}/ -E_{diss}$, IBSI$^{GBP}$ (Table S5 in ESI†) and $E_{int}/ -E_{diss}$, IBSI$^{PRO}$ (Table S6 in ESI†). These points were considered outliers and were removed from the linear fits (See also Figures S4-S8 in ESI†). Nonetheless, it would be important to understand, if possible, why these unusual combinations occur thus helping to establish a "domain of applicability" for the linear models developed herein. Remarkably, the large majority of outliers are shared between the IBSI$^{GBP}$ and IBSI$^{PRO}$ approaches and are represented in Figure 3. Regarding **Set 1**, the data points comprise three complexes featuring molecular chlorine as XB donor along with three featuring trimethylamine as an acceptor. As will be observed for the other sets, the strongest XBs (Br$_2$···trimethylamine, I$_2$···trimethylamine, and Cl$_2$···trimethylamine) are not well described by a linear fit since IBSI is an intrinsic bond strength index which does not explicitly account for electronic (or geometric) relaxation/reorganization occurring when bringing the molecules together to form the complex. Such rearrangement is more extensive as we go from weak non-covalent interactions to the covalent regime. Therefore, it is not surprising that, as mentioned above, the large majority of the complexes used in the final fit possess IBSI values below non-covalent threshold ($< 0.15$). On the other hand, the presence of the other molecular chlorine complexes (Cl$_2$···dimethylthioether and Cl$_2$···thioacetone) is not so straightforwardly explained. We notice however that an overestimation of the interaction energy was previously observed for

complexes of molecular chlorine by several DFT functionals [70] and DFT as used to obtain the ED. That however does not explain why these complexes are also not well described using IBSI$^{PRO}$.

In **Set 2** two complexes with FI as XB donor along with one complex with PH$_3$ as an acceptor are featured as outliers in both IBSI$^{GBP}$ and IBSI$^{PRO}$ approaches. All these are strong XBs and possess IBSI values above or very close to the non-covalent threshold ($< 0.15$). Notice also that in the reference 55, complexes involving FCl, FBr, and FI interacting with phosphine and thiirane were excluded from the set as they behaved substantially different (unusually strong) owing to the formation of Mulliken inner complexes [82,83].

Regarding **Set 3**, it is immediately evident that both Br$_2$···HLi FI···HLi complexes have an extremely high $-E_{diss}$ value corresponding to X···B IBSI values $> 0.25$. Indeed the calculated X···B IBSI values are larger or very close to the covalent Br–Br or I–F bond, indicating a very large degree of bond weakening of these covalent bonds upon binding. Additionally, in **Set 3**, $-E_{diss}$ values take account for the deformation of the fragments and as mentioned, IBSI does not consider the large electronic and geometric relaxation/reorganization upon binding and larger deviations from linearity are thus expected.

### 3.5 Distance dependence

Our results show that IBSI$^{GBP}$, and most strikingly IBSI$^{PRO}$, present strong linear correlations with interaction energies ($E_{int}$, **Set 1-Set 2**) or dissociation energies ($-E_{diss}$, **Set 3**) for closed-shell complexes in the ground state, as long as strong XBs are removed. Since the interaction energy generally increases with decreasing distance and IBSI requires the $1/d^2$ term (Equation 5) so that $\Delta g^{pair}$ can be compared with properties such as force constants [56], we also sought to explore the influence of the this term by plotting the interaction energies ($E_{int}$, **Set 1-Set 2**) or dissociation energies ($-E_{diss}$, **Set 3**) as a function of $1/d_{int}^2$ (Figure 4, top). Notice that most of the previously mentioned outliers appear in the short X···B distance regime (Figure S11 in ESI†), some of them much lower than sum of their van der Waals radii, indi-
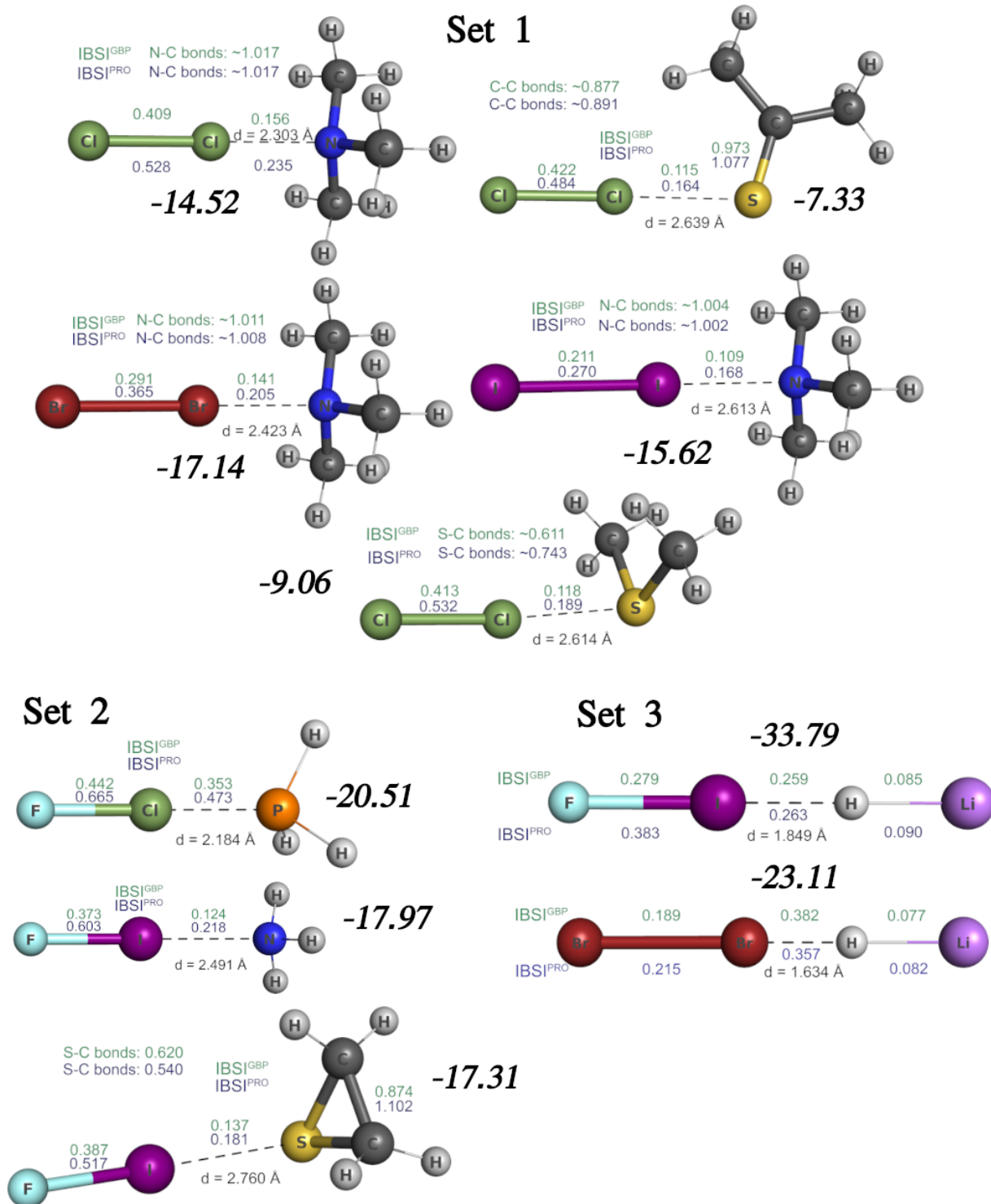
Fig. 3 3D representation of the complexes identified as outliers for Sets 1-3 along with the calculated IBSI$^{GBP}$ and IBSI$^{PRO}$ values. The respective $E_{int}/ -E_{diss}$ values in kcal mol$^{-1}$ are also shown
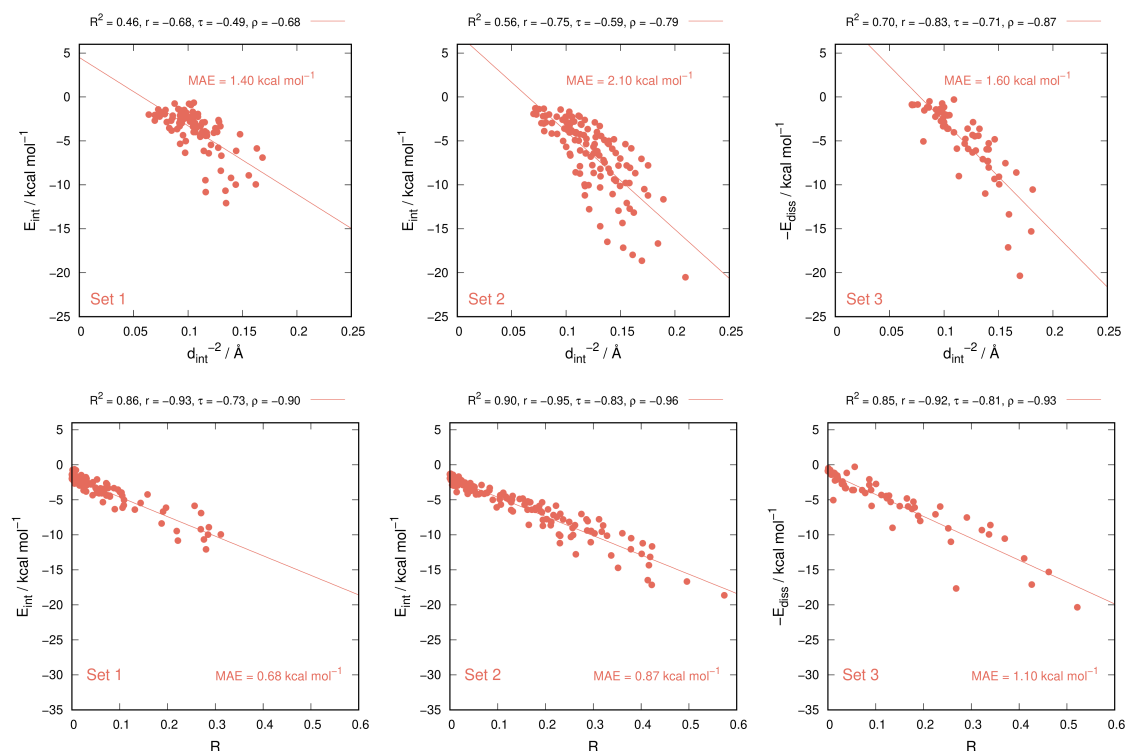
Fig. 4 Top: Interaction energies ($E_{int}$, Set 1-Set 2) or dissociation energies ($-E_{diss}$, Set 3) as a function of $1/d_{int}^2$ (top) or $R$ as defined in Equations 9-10 (bottom).

cating a large covalent character concomitant with a large IBSI value. In such cases where a large electronic rearrangement is observed, along with a large degree of penetration of the electron clouds, IBSI is not a suitable descriptor. The linear correlation between the energies and $1/d_{int}^2$ is poor, as expected, since $\delta g^{pair}$ is not constant, depending on the pair. It is therefore possible that the integral of $\delta g^{pair}$ provides a size-dependent scaling factor to the interaction distance. We plotted $E_{int}$ (**Set 1-Set 2**) or $-E_{diss}$ (**Set 3**) as a function of $R$ defined as

$$R = \frac{V_{overlap}(X \cdots B)}{d_{int}^2} \times f \tag{9}$$

where $f = 1/1\text{Å}$ thus guaranteeing that R is dimensionless, and $V_{overlap}(X \cdots B)$ is the volume of the overlap between X and B, represented as van der Waals spheres of radii $R_{vdw}(X)$ and $R_{vdw}(B)$, and calculated using

$$V_{overlap} = \frac{4}{3}\pi \left( \frac{R_{vdw}(X) + R_{vdw}(B) - d_{int}}{2} \right)^2 \times \tag{10}$$

$$(3\max(R_{vdw}(X), R_{vdw}(B)) - (R_{vdw}(X) + R_{vdw}(B) - d_{int}))/2$$

The radii used for the calculations are given in Table S7 in ESI†. Notice that the numerator emulates a non-relaxed size-dependent scaling factor, as we could assume occurs in IBSI, at least in the IBSI$^{PRO}$ approach. The plot is depicted in Figure 4, bottom whereas Figure S12 in ESI† contains the full set. Strikingly, and despite the ill-defined nature of van der Waals radii[84], the linear fits are very good and the MAE values are comparable to both

IBSI$^{PRO}$ and IBSI$^{GBP}$ thus indicating that, indeed, $\delta g^{pair}$ seems to be proportional to the overlapping volume of the atoms at a given interaction distance and that this distance can play a dominant role in determining both IBSI and $R$.

## 4 Conclusions

Predicting the trends and interaction energies of halogen bonding interactions using simple and computationally cheap molecular descriptors has been recursively addressed in the literature. In this scope, the usage of $V_{S,max}$ of the halogen atom has been an example of such an approach and continues to be the most intuitive descriptor with a clear physical interpretation. However, this descriptor requires the usage of QM calculations which could be hard for very large libraries of compounds. In this exploratory work, we tested the possibility of using the Intrinsic Bond Strength Index (IBSI) as halogen bond energy descriptor for three different datasets containing highly accurate QM-based interaction energies. Notice that XBs were mentioned in the original IBSI reference[56], however, this is the first systematic study regarding the usage of IBSI in halogen bonding. We first addressed an ED partition scheme (GBP) that rely on QM calculations (IBSI$^{GBP}$) providing also some results using the alternative IBSI$^H$ approach whose implementation requires a proper validation (see above). IBSI values that were insensitive to the basis set size for 3 complexes featuring strong, mild, and weak XBs and when applied to **Set 1**–**Set 3**, IBSI$^{GBP}$ linearly correlated with the interaction energy (or dissociatin energies) with the linear models providing MAEs typically below 1 kcal mol$^{-1}$, reaching 0.62 kcal mol$^{-1}$ for IBSI$^{GBP}$ in **Set 1**. Thus, IBSI$^{GBP}$ can in prin-

ciple be used not only from a qualitative perspective to compare halogen bond stability, but also also can be used to provide quantitative estimates of the interaction energy. Despite these exciting results, the usage of QM-based electron density could still hinder applications in large datasets or biomolecular systems. Therefore, we also explored the possibility of obtaining a quantitative model that predicts the interactions energies based in IBSI$^{PRO}$ which relies on the so-called promolecular approach, and hence, it only requires the geometry as an input. In spite of its simplicity, the performance of IBSI$^{PRO}$ was comparable to the QM-based method, actually outperforming IBSI$^{GBP}$ for **Set 2**, suggesting that computationally demanding calculations are not necessary in order to achieve reasonable accuracy. These linear fits should only be applied to closed-shell complexes in the ground state and whose energies are week to mild. Strong halogen bonds undergo a large degree of electronic rearrangement and therefore, IBSI fails to properly describe the target energy and such compounds tend to appear as ouliers. Nonetheless, as long as one stays in the non-covalent regime, which is often the case in halogen-bonded protein-ligand systems systems[54], IBSI should provide a good energy estimator. Given the $1/d^2$ term in the IBSI formulation, we sough to explore its influence on the IBSI values. We showed that in fact, $\delta g^{pair}$ can be seen as a size-dependent scaling factor to the interaction distance since the plots of the overlap volume of X and B, represented as van der Waals spheres, over $d^2$ (which we named $R$), also provides a suitable description of the interaction/dissociation energies. Overall, our exploratory work can open the door to the usage of IBSI$^{PRO}$as a fast and reliable XB interaction energy descriptor in large systems, *e.g.* proteins, provided that proper calibration curves and geometries are available. IBSI, along with empirical approaches such as the descriptor $R$ proposed herein, can be useful and complementary to the universal $V_{S,max}$ descriptor. Indeed, IBSI$^{PRO}$ is an excellent XB interaction energy estimator when the geometries of the complexes are available and QM calculations are hindered by the size of the system or the size of the library. On the other hand $V_{S,max}$, despite typically requiring a QM calculation, can be universally applied to any halogen bond donor, allowing us to estimate the interaction energy even when the geometry of the XB complex is not known.

## Author Contributions

Conceptualization: P.J.C.; Methodology: O.S. and P.J.C.; Validation: O.S. and P.J.C.; Investigation: O.S.; Writing – original draft: O.S.; Writing – review & editing: P.J.C.; Funding Acquisition: P.J.C.; Supervision: P.J.C.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

## Notes and references

1 G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati and G. Terraneo, *Chem. Rev.*, 2016, **116**, 2478–2601.

2 G. R. Desiraju, P. S. Ho, L. Kloo, A. C. Legon, R. Marquardt, P. Metrangolo, P. Politzer, G. Resnati and K. Rissanen, *Pure Appl. Chem.*, 2013, **85**, 1711–1713.

3 P. J. Costa, *Phys. Sci. Rev.*, 2017, **2**, 20170136.

4 T. Clark, M. Hennemann, J. S. Murray and P. Politzer, *J. Mol. Model.*, 2007, **13**, 291–296.

5 P. Politzer, J. S. Murray and T. Clark, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7748–7757.

6 K. E. Riley and K.-A. Tran, *Faraday Discuss.*, 2017, **203**, 47–60.

7 L. P. Wolters, P. Schyman, M. J. Pavan, W. L. Jorgensen, F. M. Bickelhaupt and S. Kozuch, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2014, **4**, 523–540.

8 E. Engelage, D. Reinhard and S. M. Huber, *Chem. – Eur. J*, 2020, **26**, 3843–3861.

9 J. M. Holthoff, R. Weiss, S. V. Rosokha and S. M. Huber, *Chem. – Eur. J*, 2021, **27**, 16530–16542.

10 J. S. Murray and P. Politzer, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2017, **7**, e1326.

11 J. S. Murray and P. Politzer, *ChemPhysChem*, 2021, **22**, 1201–1207.

12 J. Řezáč and A. De La Lande, *Phys. Chem. Chem. Phys.*, 2017, **19**, 791–803.

13 L. Mendez, G. Henriquez, S. Sirimulla and M. Narayan, *Molecules*, 2017, **22**, 1397.

14 Y. Li, B. Guo, Z. Xu, B. Li, T. Cai, X. Zhang, Y. Yu, H. Wang, J. Shi and W. Zhu, *Sci. Rep.*, 2016, **6**, 1–10.

15 J. Lapp and S. Scheiner, *J. Phys. Chem. A*, 2021, **125**, 5069–5077.

16 A. Priimagi, G. Cavallo, P. Metrangolo and G. Resnati, *Acc. Chem. Res.*, 2013, **46**, 2686–2695.

17 V. Oliveira, E. Kraka and D. Cremer, *Phys. Chem. Chem. Phys.*, 2016, **18**, 33031–33046.

18 M. H. Kolar and P. Hobza, *Chem. Rev.*, 2016, **116**, 5155–5187.

19 R. L. Sutar, E. Engelage, R. Stoll and S. M. Huber, *Angew. Chem. Int. Ed.*, 2020, **59**, 6806–6810.

20 M. Kaasik and T. Kanger, *Front. Chem.*, 2020, **8**, 958.

21 E. Margiotta, S. C. C. van der Lubbe, L. de Azevedo Santos, G. Paragi, S. Moro, F. M. Bickelhaupt and C. Fonseca Guerra, *J. Chem. Inf. Model*, 2020, **60**, 1317–1328.

22 L. Schifferer, M. Stinglhamer, K. Kaur and O. G. Mancheño, *Beilstein J. Org. Chem.*, 2021, **17**, 2270–2286.

23 C. Curiac, L. A. Hunt, M. A. Sabuj, Q. Li, A. Baumann, H. Cheema, Y. Zhang, N. Rai, N. I. Hammer and J. H. Del-

camp, *J. Phys. Chem C*, 2021, **125**, 17647–17659.

24 S. An, A. Hao and P. Xing, *ACS Nano*, 2021, **15**, 15306–15315.

25 M. H. H. Voelkel, E. Engelage, M. Kondratiuk and S. M. Huber, *Eur. J. Org. Chem.*, **2022**, e202200211.

26 J.-W. Zou, M. Huang, G.-X. Hu and Y.-J. Jiang, *RSC Advances*, 2017, **7**, 10295–10305.

27 M. R. Scholfield, M. C. Ford, A.-C. C. Carlsson, H. Butta, R. A. Mehl and P. S. Ho, *Biochemistry*, 2017, **56**, 2794–2802.

28 J. Heidrich, L. E. Sperl and F. M. Boeckler, *Front. Chem.*, 2019, **7**, 9.

29 S. H. Jungbauer and S. M. Huber, *J. Am. Chem. Soc.*, 2015, **137**, 12110–12120.

30 P. M. J. Szell, S. Zablotny and D. L. Bryce, *Nat. Commun.*, 2011, **10**, 916.

31 H. Yang and M. W. Wong, *Molecules*, 2020, **25**, 1045.

32 C. Xu, V. U. B. Rao, J. Weigen and C. C. J. Loh, *Nat. Commun.*, 2020, **11**, 4911.

33 H. Wang, H. K. Bisoyi, A. M. Urbas, T. J. Bunning and Q. Li, *Chem. – Eur. J*, 2019, **25**, 1369–1378.

34 X. Miao, Z. Cai, H. Zou, J. Li, S. Zhang, L. Ying and W. Deng, *J. Mater. Chem. C*, 2022, **10**, 8390–8399.

35 R. Shi, D. Yu, F. Zhou, J. Yu and T. Mu, *Chem. Commun.*, 2022, **58**, 4607–4610.

36 A. Vanderkooy and M. S. Taylor, *Faraday Discuss.*, 2017, **203**, 285–299.

37 M. S. Alvarez, C. Houzé, S. Groni, B. Schöllhorn and C. Fave, *Org. Biomol. Chem.*, 2021, **19**, 7587–7593.

38 A. Singh, A. Torres-Huerta, T. Vanderlinden, N. Renier, L. Martínez-Crespo, N. Tumanov, J. Wouters, K. Bartik, I. Jabin and H. Valkenier, *Chem. Commun.*, 2022, **58**, 6255–6258.

39 M. Kokot, M. Weiss, I. Zdovc, M. Hrast, M. Anderluh and N. Minovski, *ACS Med. Chem. Lett.*, 2021, **12**, 1478–1485.

40 S. Jena, J. Dutta, K. D. Tulsiyan, A. K. Sahu, S. S. Choudhury and H. S. Biswal, *Chem. Soc. Rev.*, 2022, **51**, 6255–6258.

41 R. S. Nunes, D. Vila-Viçosa and P. J. Costa, *J. Am. Chem. Soc.*, 2021, **143**, 4253–4267.

42 S. M. Huber, E. Jimenez-Izal, J. M. Ugalde and I. Infante, *Chem. Commun.*, 2012, **48**, 7708–7710.

43 I. Nicolas, F. Barriere, O. Jeannin and M. Fourmigue, *Cryst. Growth Des.*, 2016, **16**, 2963–2971.

44 J. Thirman, E. Engelage, S. M. Huber and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2018, **20**, 905–915.

45 B. Inscoe, H. Rathnayake and Y. Mo, *J. Phys. Chem. A*, 2021, **125**, 2944–2953.

46 T. Clark and A. Heßelmann, *Phys. Chem. Chem. Phys.*, 2018, **20**, 22849–22855.

47 T. Clark, J. S. Murray and P. Politzer, *ChemPhysChem*, 2018, **19**, 3044–3049.

48 T. Brinck and A. N. Borrfors, *J. Mol. Model.*, 2019, **25**, null.

49 T. Brinck, P. Carlqvist and J. H. Stenlid, *J. Phys. Chem. A*, 2016, **120**, 10023–10032.

50 R. Nunes and P. J. Costa, *Chem. – Asian J.*, 2017, **12**, 586–594.

51 M. L. Kuznetsov, *Int. J. Quantum Chem.*, 2019, **119**, e25869.

52 M. L. Kuznetsov, *Molecules*, 2019, **24**, 2733.

53 M. L. Kuznetsov, *Molecules*, 2021, **26**, 2083.

54 P. J. Costa, R. Nunes and D. Vila-Viçosa, *Expert. Opin. Drug. Discov.*, 2019, **14**, 805–820.

55 R. A. Shaw and J. G. Hill, *Inorganics*, 2019, **7**, 19.

56 J. Klein, H. Khartabil, J.-C. Boisson, J. Contreras-García, J.-P. Piquemal and E. Hénon, *J. Phys. Chem. A*, 2020, **124**, 1850–1860.

57 C. Lefebvre, G. Rubez, H. Khartabil, J.-C. Boisson, J. Contreras-García and E. Hénon, *Phys. Chem. Chem. Phys.*, 2017, **19**, 17928–17936.

58 C. Lefebvre, H. Khartabil, J.-C. Boisson, J. Contreras-García, J.-P. Piquemal and E. Hénon, *ChemPhysChem*, 2018, **19**, 724–735.

59 J. Contreras-García and Y. Weitao, *Acta Physico-Chimica Sinica*, 2018, **34**, 567.

60 T. O. Unimuke, H. Louis, E. A. Eno, E. C. Agwamba and A. S. Adeyinka, *ACS Omega*, 2022, **7**, 13704–13720.

61 E. A. Katlenok, A. V. Rozhkov, O. V. Levin, M. Haukka, M. L. Kuznetsov and V. Y. Kukushkin, *Cryst. Growth Des.*, 2020, **21**, 1159–1177.

62 M. Ponce-Vargas, J. Klein and E. Hénon, *Dalton Trans.*, 2020, **49**, 12632–12642.

63 E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498–6506.

64 J. Klein, E. Pluot, G. Rubez, J. C. Boisson and E. Hénon, *IGM-Plot (Revision 2.6.9b)*, `http://igmplot.univ-reims.fr/`.

65 T. Lu and F. Chen, *J. Comput. Chem.*, 2011, **33**, 580–592.

66 C. Lefebvre, J. Klein, E. Pluot, G. Rubez, H. Khartabil, J.-C. Boisson and E. Henon, *IGMPlot Documentation (Revision 3.0)*.

67 T. Lu and Q. Chen, *J. Comput. Chem.*, 2022, **43**, 539–555.

68 F. L. Hirshfeld, *Theor. Chimi. Acta*, 1977, **44**, 129–138.

69 T. Lu and Q. Chen, 2022, chemrxiv:10.26434/chemrxiv-2022-g1m34.

70 K. Kriz and J. Řezáč, *Phys. Chem. Chem. Phys.*, 2022, **24**, 14794–14804.

71 K. Eskandari and M. Lesani, *Chem. – Eur. J*, 2015, **21**, 4739–4746.

72 L. N. Anderson, F. W. Aquino, A. E. Raeber, X. Chen and B. M. Wong, *J. Chem. Theory Comput.*, 2018, **14**, 180–190.

73 S. Kozuch and J. M. L. Martin, *J. Chem. Theory Comput.*, 2013, **9**, 1918–1931.

74 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi,

N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Revision A.2*, 2009.

75 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

76 D. Bulfield, E. Engelage, L. Mancheski, J. Stoesser and S. M. Huber, *Chem. – Eur. J*, 2020, **26**, 1567–1575.

77 A. Otero-De-La-Roza, E. R. Johnson and G. A. DiLabio, *J.*

*Chem. Theory Comput.*, 2014, **10**, 5436–5447.

78 D. Rappoport and F. Furche, *J. Chem. Phys.*, 2010, **133**, 134105.

79 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.

80 M. Hubert, M. Debruyne and P. J. Rousseeuw, *Wiley Interdiscip. Rev. Comput. Stat.*, 2018, **10**, e1421.

81 J. Hardin and D. M. Rocke, *J. Comput. Graph. Stat.*, 2005, **14**, 928–946.

82 J. G. Hill, *Phys. Chemi. Chem.l Phys.*, 2014, **16**, 19137–19140.

83 R. A. Shaw, J. G. Hill and A. C. Legon, *J. Phys. Chem. A*, 2016, **120**, 8461–8468.

84 P. Politzer and J. S. Murray, *Struct. Chem.*, 2021, **32**, 623–629.