

MedChemComm

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

A fundamental view of enthalpy–entropy compensation

Ulf Ryde*

Department of Theoretical Chemistry, Lund University, Chemical Centre, P. O. Box 124,
SE-221 00 Lund, Sweden

Correspondence to Ulf Ryde, E-mail: Ulf.Ryde@teokem.lu.se,

Tel: +46 – 46 2224502, Fax: +46 – 46 2228648

2014-05-27

Abstract

In this paper, enthalpy–entropy compensation (EEC) during the association of two molecules is studied by minimising model systems with molecular mechanics (MM) or quantum mechanics (QM), calculating translational, rotational, and vibrational contributions to the enthalpy and entropy with standard statistical-mechanics methods, using the rigid-rotor harmonic-oscillator approach. We start with simple two-atom models, for which dispersion and electrostatics can be studied separately, showing that there is no fundamental difference between dispersion, electrostatics, or even covalent interactions. All three types of interactions give rise to EEC and a saturation of $T\Delta S$ as ΔH becomes strongly negative. Next, homologous series of complexes dominated either by dispersion or hydrogen bonds are studied. We see no qualitative difference between results obtained at the MM or QM level, and for all complexes except two very weak, EEC is observed, owing to the loss of translational and rotational entropy, typically counteracted by the vibrational entropy. Within homologous series, linear relations between $T\Delta S$ and ΔH with slopes of 0.1–1.7 are obtained with no clear difference between dispersive or hydrogen-bonded systems (but ~ 0.01 for ionic and covalent interactions). These relations often reflect the increasing size of the complexes coming from the translational and rotational entropies, but at least for the hydrogen-bonded complexes, it is significantly enhanced also by the vibrational entropy (which depends on the strength of the interaction). Thus, for homologous series of molecules with repeated interactions studied in vacuum, EEC is a rule. However, if water molecules are added, the relation is blurred and it can be predicted that for a real binding reaction in water solution, both enthalpy–entropy compensation and anti-compensation can be observed, depending on the detailed interaction of the two molecules with water before and after binding, further complicated by dynamic effects.

Key words: Ligand binding, enthalpy–entropy compensation, quantum mechanics, molecular mechanics.

Introduction

The enthalpy–entropy compensation (EEC), i.e. the observation that an increase in the enthalpy during the non-covalent association of two molecules (e.g. the binding of a drug candidate to a protein) is often to a large extent cancelled by a concurrent decrease in the entropy, is a much debated phenomenon. The term has been used with several meanings, but recent reviews have clarified the concept [1,2,3]. Even if the EEC in some cases has been attributed to experimental errors and limitations [4,5] or to a natural consequence of thermodynamics [1,6,7], there is now much clear evidence of EEC [8,9,10,11,12]. On the other hand, several examples where enthalpy and entropy enforce each other have also been reported [1,13,14,15]. Owing to the importance of EEC for the understanding of molecular recognition and drug design, it has been thoroughly studied by both experimental and theoretical methods [1,16,17].

In 1981, Jencks pointed out that in the association of two molecules, a favourable binding enthalpy has to overcome an unfavourable loss of translational and rotational entropy when two molecules form a complex, directly leading to EEC [18]. Williams and coworkers suggested that residual motion in the complex will partly compensate for the loss in translational and rotational entropy, and that this residual motion will decrease as the strength of the complex is increased [19]. Therefore, EEC naturally arises in associations and if the entropy is plotted against enthalpy for interactions with increasing strength, it first rapidly becomes more negative and then saturates and asymptotically approaches a limiting value of complete loss of translational and rotational entropy, as the binding enthalpy becomes more negative (cf. Figure 1). In a series of articles, they showed that such a curve is supported by experimental measurements and that it can be used to explain and predict binding affinities and cooperativity [16,20,21,22,23].

Dunitz suggested that the residual motion can be estimated from the vibrational entropy [24]: When a complex is formed, the six translational and rotational degrees of freedom of the ligand are converted to six vibrations between the ligand and the receptor, and as the complex becomes stronger, the vibrations become stiffer. The vibrational entropy can be estimated from the entropy of a harmonic oscillator and for many types of interactions, there is a linear relation between the strength of the interaction and the quadratic force constant of the vibration. Interestingly, he made an error of signs, suggesting that the corresponding vibrational entropy opposes the binding enthalpy, although it actually enforces the binding, partly compensating the loss of translational and rotational entropy. However, when correctly performed, this provides a powerful approach to quantify and understand the EEC with theoretical methods.

Recently, Korth used this approach to study EEC in 88 complexes of small molecules in gas phase [25]. He argued that quantum mechanical (QM) effects are essential for the understanding of the EEC¹ and that interactions dominated by hydrogen bonds or by dispersive interactions show a fundamentally different behaviour, giving rise to distinct types of EEC. A problem with such an approach is that for real molecules studied by QM, there are no pure hydrogen-bonded or electrostatic systems without any dispersion, making the distinction problematic. Moreover, QM investigations are limited to the strength of interactions observed for real chemical systems.

In this paper, we explore the opposite viewpoint, i.e. that the EEC can be understood from a molecular mechanics (MM) perspective. The advantage with MM is that we can design systems that are purely dispersive or electrostatic, and continuously modify the strength of the

1 This argument is quite strange, considering that he primarily discussed results obtained at the dispersion-corrected density-functional theory level, i.e. with dispersion estimated with a molecular-mechanics (MM) approach. The fact that he obtains similar results at the MP2 level shows that at least the dispersion interactions are well modelled by MM.

dispersion or the magnitude of the charges, deriving analytical expressions for the enthalpy and entropy of each system. Thereby, we can fully understand what is expected from the two types of systems at any strength of interaction, thereby deciding whether there is any fundamental difference in EEC between dispersive and electrostatic (hydrogen-bonded) systems. We start from simple two-atom models and continue to more complicated systems, involving many interaction sites with or without a combination of dispersive and electrostatic effects. Throughout the investigation, we compare with results obtained at the QM level, to address whether there is any fundamental difference between QM and MM. In agreement with previous theoretical studies [18-25], we observe EEC for essentially all studied systems. The question then naturally arises why EEC is not observed in all experimental studies of ligand binding. We will discuss possible answers to this fundamental question about molecular interactions.

Theory and Methods

In this paper, we study several variants of the general binding reaction



where M_1 and M_2 are two molecules (or atoms or ions) and M_1M_2 their complex. We study the enthalpy, entropy, and free energy of this reaction, which are denoted ΔH , ΔS , and ΔG .

Throughout the article, all entropies are discussed in energy terms, $T\Delta S$ at $T = 298.15$ K, so that the terms are comparable when the free energy is calculated from

$$\Delta G = \Delta H - T\Delta S \quad (2)$$

We used either QM or MM to calculate the energies of M_1 , M_2 , and M_1M_2 , as well as their vibrational frequencies. From the frequencies, we calculated zero-point vibrational energies, as well as thermal corrections to the enthalpy and entropy of each reactant, based on standard statistical-mechanical expressions [26], using an ideal-gas rigid-rotor harmonic-oscillator (RRHO) approximation. Hence, the enthalpy is given by

$$H = E_{\text{int}} + H_{\text{trans}} + H_{\text{rot}} + H_{\text{vib}} \quad (3),$$

where E_{int} is the internal energy of the reactant (the total QM or MM energy of the molecule). H_{trans} is the translational enthalpy, which is $2.5 RT$ for any molecule (where R is the gas constant and T is the absolute temperature). H_{rot} is the rotational entropy, which is RT for linear molecules, $1.5 RT$ for non-linear molecules, and vanishes for atoms. Finally,

$$H_{\text{vib}} = R \sum_{i=1}^{3N-6(5)} \left(\frac{h\nu_i}{2k_B} + \frac{h\nu_i}{k_B(1 - e^{-h\nu_i/k_B T})} \right) \quad (4)$$

is the vibrational enthalpy, which depends on the $3N - 5$ (linear molecules) or $3N - 6$ (non-linear molecules) vibrational frequencies ν_i (k_B is the Boltzmann constant and N is the number of atoms in the molecule or complex; the first term of the sum is the zero-point vibrational energy).

Similarly, the entropy is given by

$$S = S_{\text{trans}} + S_{\text{rot}} + S_{\text{vib}} \quad (5),$$

where the translational, rotational, and vibrational enthalpies are given by

$$S_{trans} = \frac{5}{2} R \ln \left(\frac{RT}{P N_A} \left(\frac{2\pi M k_B T}{h^2} \right)^{3/2} \right) \quad (6)$$

$$S_{rot} = \frac{R}{2} \left(3 + \ln \left(\frac{\sqrt{\pi}}{\sigma} \left(\frac{8\pi^2 k_B T}{h^2} \right)^{3/2} \right) \sqrt{I_1 I_2 I_3} \right) \quad (7)$$

$$S_{vib} = R \sum_{i=1}^{3N-6(5)} \left(\frac{h\nu_i}{k_B T (e^{h\nu_i/k_B T} - 1)} - \ln(1 - e^{-h\nu_i/k_B T}) \right) \quad (8)$$

From this, it can be seen that the translational entropy depends on the molar mass of the molecule (M ; P is the pressure, assumed to be 1 atm, and h is Planck's constant), the rotation entropy depends on symmetry number ($\sigma = 1$ for most molecules, except for those with rotational symmetry) and the moments of inertia (I_1 , I_2 , and I_3 , which depend on the coordinates and mass of the atoms) of the molecule (the formula is somewhat different for linear molecules [26]), whereas the vibrational entropy depends on the vibrational frequencies.

MM calculations were either performed analytically (diatomic molecules) or with the nmodes program in the AMBER 9 software suite [27], using the GAFF force field [28] and either zeroed or AM1-BCC charges [29,30], except the β sheet models, which used the Amber ff99SB force field [31]. The alkane models with water molecules used a +0.1 e charge on the H atoms and a compensating negative charge on the C atoms, similar to the ff99SB charges on non-polar amino acids. For water, TIP3P parameters were used [32].

QM calculations were performed with the Turbomole 6.5 software [33,34], using the TPSS density-functional theory method [35] with the def2-TZVPPD basis set [36,37] and the empirical DFT-D3 dispersion correction [38]. The calculations were sped up by expanding the Coulomb potentials in auxiliary basis sets, the resolution-of-identity approximation [39,40,41].

To obtain more stable results for the vibrational contribution in the MM and QM (but not the analytical) calculations, low-lying vibrational modes were treated by the free-rotor approximation, using the interpolation model suggested by Grimme with the parameter $\omega_0 = 100 \text{ cm}^{-1}$ [42] (but the term was still assigned to the vibrational contribution). These calculations were performed with the thermo program, kindly provided by Prof. Stefan Grimme. This approximation does not affect the general conclusions of the paper; it only makes the curves smoother.

We will discuss two types of EEC: The first (which we will denote EEC1 in the following) involves a single binding reaction of the type in Eqn. 1. Then, there is an EEC if ΔH and $T\Delta S$ have the same signs. This can be characterised by the quotient

$$q_{SH} = \frac{T\Delta S}{\Delta H} \quad (9),$$

which is positive in case of EEC1.

The second type of EEC (which we will denote EEC2 below) involves several binding reactions, typically in some series of homologous molecules. Strictly, EEC2 implies that any difference in ΔH between two reactions in this series ($\Delta\Delta H$) is counteracted by a difference in

$T\Delta S$ ($T\Delta\Delta S$) with the same sign. This requires that $T\Delta S$ is a monotonous function of ΔH . In a less stringent, but more useful version, EEC2 implies that there is a significant correlation between ΔH and $T\Delta S$ with a positive slope for a series of homologous compounds. EEC2 is probably closer to what is normally meant by EEC, but it depends on how the homologous series is defined.

Results

In this paper, we study a number of binding reactions for simple atoms, ions, and molecules at the QM and MM levels, estimating the binding enthalpies and entropies from the standard statistical-mechanics formulae in Eqns. 3–8. Individual reactions will be discussed in separate sections. We aim at answering three questions: Is there a fundamental difference in EEC between systems dominated by dispersion or electrostatics (hydrogen bonds)? Is there a fundamental difference in EEC when studied by MM or QM? and Why is EEC not observed experimentally for all binding reactions?

Pure dispersion

We will start with the simple association of two neutral atoms, i.e. a purely dispersive system without electrostatics. At the MM level, this is obtained by assuming that they interact with a Lennard-Jones potential:

$$E_{\text{int}}(r) = \frac{A}{r^{12}} - \frac{B}{r^6} \quad (10)$$

where r is the distance between the two atoms and A and B are constants describing the strength of the exchange-repulsion and dispersion interactions, respectively. This energy attains a minimum at r_{min} , which easily can be found analytically. By differentiation, the corresponding vibrational frequency can also be found. The system is completely described by four parameters, A , B , and the masses of the two atoms, m_1 and m_2 (besides the temperature and pressure, which we assume are fixed at 298.15 K and 1 atm thorough this article).

Figure 1a shows how the enthalpy, entropy, and free energy for the binding reaction depends on the B Lennard-Jones dispersion constant for the interaction of two argon-like atoms. It can be seen when $B > 200 \text{ \AA}^6\text{kJ/mol}$ (for smaller B , the RRHO approximation breaks down and the entropy tends towards infinity), the systems shows EEC1 (ΔH and $T\Delta S$ are both negative), as is illustrated by a positive q_{SH} (Eqn. 9). For intermediate values of B , ΔH is small, $T\Delta S$ is dominating, ΔG is positive (unfavourable binding), and q_{SH} is rather large ($q_{\text{SH}} \approx 6$ for $B = 11\,160 \text{ \AA}^6\text{kJ/mol}$ which gives the same minimum energy and distance as a QM calculation of Ar_2). However, for $B > 40\,000 \text{ \AA}^6\text{kJ/mol}$, ΔG becomes negative and q_{SH} decreases below 1.

The various enthalpic and entropic contributions are shown in Figure 1b. In this diatomic system, the three translational degrees of freedom of each of the free atoms are converted to one vibrational, two rotational, and three translational degrees of freedom in the complex. ΔH_{trans} is constant, $-2.5 RT$ (-6 kJ/mol ; the two atoms atom and the complex contribute by $2.5 RT$ each), i.e. favouring the complex formation. ΔH_{rot} is also constant, and for this diatomic system, it is positive, $+RT$ (2 kJ/mol), because the free atoms do not have any rotational contributions. ΔH_{vib} is also almost constant, $\sim 3 \text{ kJ/mol}$; it depends on the single vibrational frequency (ν , cf. Eqn. 4), but for the small frequencies typical for non-covalent interactions (for the range shown in Figure 1b, $\nu = 0\text{--}200 \text{ cm}^{-1}$), the term is small and the dependence on

the frequency is weak.

The $T\Delta S_{\text{trans}}$ term is also constant but large and negative, -44 kJ/mol. It depends on the mass of the atoms (Eqn. 6) and reflects the loss of translational entropy when the two atoms form a complex. It is the main cause of EEC1, as suggested by Jencks [18]. $T\Delta S_{\text{rot}}$ is also rather large. For this diatomic system, it is positive (because the free atoms do not have any rotational contributions, so it compensates for two of the lost translational degrees of freedom). As can be seen in Eqn. 7, it depends on the moments of inertia, which are functions of the masses and positions of the atoms, but for this diatomic system, it is simply the product of the mass and the distance between the two atoms (r_{min} , also shown in Figure 1b). Thus, it decreases somewhat as the diatomic complex becomes more compact. Finally, $T\Delta S_{\text{vib}}$ is positive and shows a strongly decreasing trend. As can be seen in Eqn. 8, it depends on the vibrational frequency, but, in variance to the H_{vib} term, it is largest for small frequencies and it becomes insignificant for frequencies above 1000 cm^{-1} . It tends to compensate for the third lost translational degree of freedom, especially for weak complexes. Consequently, the net enthalpy correction ($\Delta H_{\text{trans}} + \Delta H_{\text{rot}} + \Delta H_{\text{vib}}$) is almost constant (-1 kJ/mol), whereas the total entropy correction becomes more negative as B and therefore ΔH increases, owing to the concerted decrease of both the rotational and vibrational contributions. This gives rise to a strict EEC2, illustrated in Figure 1c.

Figure 1c shows that similar results are obtained also for other values of A . In all cases, $T\Delta S$ first sharply becomes more negative and then saturates as the enthalpy is increased in magnitude, exactly as suggested by Williams [19]. In this figure, we have also included QM calculations of a number pairs of noble-gas atoms (it should be noted that the QM results for these weak complexes are very sensitive to the level of theory and details of the calculations and therefore only should be taken as typical, but not particularly accurate results). It can be seen that they follow quite closely the MM results, showing that there is not any major difference between the QM and MM methods for this simple model system. However, it can be seen that the QM results do not give rise to any EEC2 (the MM and QM calculations represent different homologous series – with MM only A is varied, whereas with QM all parameters, including the masses change, because we can only study real systems). In fact, there is essentially no correlation between ΔH and $T\Delta S$ for the QM calculations (the correlation coefficient r^2 is 0.1). If the noble-gas homo- and heterodimers are considered separately, both groups actually show a weak anticorrelation ($r^2 = -0.1$ or -0.2).

Pure electrostatics

Next, we turn to a system without any dispersive interactions but instead with a favourable electrostatic interaction, reminding of a hydrogen bond. With MM, this is obtained by setting $B = 0$ in Eqn.10 and adding an electrostatic term:

$$E_{\text{int}}(r) = \frac{A}{r^{12}} + \frac{q_1 q_2}{4 \pi \epsilon_0 r} \quad (11),$$

where q_1 and q_2 are the charges on the two atoms and ϵ_0 is the vacuum permittivity. Again, we can analytically find the minimum energy and distance of this interaction, as well as the vibrational frequency and thereby calculate the binding enthalpy and entropy in the same manner as for the dispersive system. This time, the results depend on five parameters, A , m_1 , m_2 , q_1 , and q_2 , but we will assume that $q_1 = -q_2 = q$, so that the complex is neutral.

Figure 2a shows how ΔH , $T\Delta S$, and ΔG vary with the charge for some typical parameters.

It can be seen that the curves are very similar to those in Figure 1a: The enthalpy is always negative, $T\Delta S$ is also negative for all $q > 0.004 e$ (again the harmonic-oscillator approximation breaks down for very weak complexes), therefore counteracting the enthalpy and giving rise to an EEC1 ($q_{SH} > 0$). There is also a strict EEC2 (the curves are monotonous). The entropy is dominating at small charges, whereas the enthalpy dominates for large charges. Consequently, ΔG changes sign at $q = 0.18 e$. For a system with unit charges ($q = 1 e$), the entropy dominates completely (giving $\Delta H = -559$ kJ/mol and $T\Delta S = -22$ kJ/mol with the parameters in Figure 2a). This illustrates the strength of Coulombic interactions in vacuum. However, in water solvent, it will be scaled down by a factor of 80, giving $\Delta H = -6$ kJ/mol and $T\Delta S = -14$ kJ/mol at the optimum interaction distance.

The various contributions to the enthalpy and entropy are qualitatively identical to those for the dispersive system (Figure 2b), with only minor quantitative differences: $\Delta H_{trans} = -6$ kJ/mol and $\Delta H_{rot} = 2$ kJ/mol, as before, ΔH_{vib} is nearly constant (3–4 kJ/mol), $T\Delta S_{trans} = -42$ kJ/mol, $T\Delta S_{rot}$ is positive and slowly decreasing with q and E (19–25 kJ/mol, because r_{min} decreases), whereas $T\Delta S_{vib}$ is also positive and more strongly decreasing in magnitude with q and E (1–19 kJ/mol). Again, the EEC1 is mainly caused by $T\Delta S_{trans}$ and the EEC2 is caused by the concerted action of the rotational and vibrational entropy. The only major difference compared to the dispersive system is that the interaction energy varies more and may become much more negative.

In Figure 2c, we show the correlation of the entropy and enthalpies as a function of A with $q = 1.0 e$. Again, it can be seen that the MM curve predicts a strict EEC2. In the plot, we have also included a number of chemical systems, studied by QM and consisting of pairs of an alkali ion and a halogen ion. It can be seen that they follow the MM curves quite well, although the enthalpy is systematically slightly underestimated (because dispersion is omitted in the MM calculations; cf. Eqn. 11). The QM results also show an EEC1 and all QM points, except two pairs (NaCl–RbF and NaCl–CsF) give rise to a strict EEC2. In fact, $r^2 = 0.95$ and the slope of the correlation line is 0.013. The deviation from the perfect correlation comes from the saturation of the entropy, as predicted by the MM theory.

Thus, from a theoretical point of view, there is no fundamental difference between dispersive and electrostatic interactions, although in practice, the latter is often stronger and therefore at a different position on the saturation curve, thereby often giving a smaller slope for the $T\Delta S$ vs. ΔH correlation line.

Covalent bonds

Next, we consider the formation of a covalent bond. It can be described by a Morse potential

$$E_{int}(r) = D \left(e^{-2C(r-r_0)} - 2e^{-C(r-r_0)} \right)^2 \quad (13),$$

which involves the dissociation energy (D), the equilibrium distance (r_0), and a constant (C). It has a minimum at $r = r_0$, with $E_{int}(r_0) = -D$, and $E_{int}(\infty) = 0$. The Hessian at $r = r_0$ is $2C^2D$. Again, we can find analytical expressions for the enthalpy and entropy of the binding, depending on the five parameters, m_1 , m_2 , r_0 , D , and C .

To reduce the number of parameters, we have in a first step set $m_1 = m_2 = 19$ Da (F), $C = 8.87 \text{ \AA}^{-1}$, and assumed an inverse relationship between r and D . Then, we obtain results presented in Figure 3a, which reproduce the QM results of F_2 at $D = 192$ kJ/mol. It should be noted that realistic bond lengths ($r = 0.7\text{--}3.5 \text{ \AA}$) are obtained for $D = 80\text{--}400$ kJ/mol, but we

have included larger ranges to show the possible variation of ΔH , $T\Delta S$, ΔG , and q_{SH} . For example, q_{SH} and ΔG attain maxima ($q_{SH} = 0.9$ and $\Delta G = -1$ kJ/mol) around $D = 7$ kJ/mol. However, for realistic parameters, ΔH and $T\Delta S$ have the same sign, thus exhibiting an EEC1. Moreover, there is a strict EEC2, although $q_{SH} < 0.3$ and $T\Delta S$ shows a quite restricted variation, -22 to -35 kJ/mol. This shows why EEC is normally not discussed for covalent bonds: ΔH is strongly dominating and $T\Delta S$ is essentially constant.

The various contributions to the enthalpy and entropy are quite different from those of the other two systems (Figure 3b): $\Delta H_{\text{trans}} = -6$ kJ/mol and $\Delta H_{\text{rot}} = 2$ kJ/mol are still constant, as is $T\Delta S_{\text{trans}} = -41$ kJ/mol. However, for reasonable D energies, $T\Delta S_{\text{vib}}$ is insignificant (< 1 kJ/mol), owing to the large vibrational frequency characterising a covalent bond. $T\Delta S_{\text{rot}}$ shows the largest variation, 11 – 19 kJ/mol decreasing when D is increased (owing to changes in the bond length), so this term is the prime cause of the observed EEC2. ΔH_{vib} also shows some variation, owing to the large vibrational frequencies, 4 – 9 kJ/mol (coming almost entirely from the zero-point energy). It provides a positive enthalpy contribution that decreases when D is increased, thereby partly counteracting ΔE , i.e. giving rise to an *energy–enthalpy* compensation. Thus, for typical covalent bond, the entropy is quite constant, although the small variation (from $T\Delta S_{\text{rot}}$) is still correlated with that of ΔH , giving an EEC2.

A similar behaviour is seen for other values of r_0 and B , as is shown in Figure 3c. It can be seen that C determines how rapidly the entropy saturates. An EEC would be apparent only for small values of C and D . QM calculations for the homolytic bond dissociation of hydrogen halogenides, hydrogen chalcogenides (with a $-1 e$ net charge), and first-row homo-nuclear diatomics are also included in Figure 3c. It can be seen that they show a restricted range of $T\Delta S$, -20 to -27 kJ/mol. There is a nearly perfect correlation between ΔH and $T\Delta S$ for the hydrogen halogenides and hydrogen chalcogenides ($r^2 = 0.99$ – 1.00), showing a strict EEC2. This linear correlation comes primarily from the $T\Delta S_{\text{rot}}$ term, reinforced by the change in the mass of the heavy atom (the mass is constant for the MM curves in Figure 3). For the diatomics or all three groups together, the correlation is lower ($r^2 = 0.48$ and 0.36 ; because there is a mixture of single, double, and triple bonds) and the EEC2 is not strict, but still significant. The slopes of the correlation lines are small, 0.006 – 0.018 .

This exercise shows that even covalent bonds are expected to give a qualitatively similar behaviour. However, owing to the strength of typical covalent interactions, $T\Delta S$ is essentially always fully saturated and no EEC is normally discussed for such systems.

Water-like systems

Next, we will study a series of water-like systems, i.e. systems of the form HXH – XH_2 . We use MM methods and assume that the only significant Lennard-Jones interaction is that between the two heavy atoms (a common assumption in many MM water models, e.g. TIP3P [32]). Thereby, the system will only involve three parameters (besides the masses, which we fix at the normal masses of hydrogen and oxygen, and the internal geometry of the monomers, which we fix to experimental values): the X – X Lennard-Jones parameters A and B , and the charge on the H atoms, q_H (we assume that the two H atoms have the same charge and that the monomers are neutral so that $q_X = -2q_H$). In addition, we will keep A constant at the normal value for TIP3P water, $2.435 \times 10^6 \text{ \AA}^{12} \text{ kJ/mol}$ [32]. This system has the advantage that each monomer has three atoms, so that there are rotational contributions to all terms, making it similar to more realistic systems.

We studied four cases: We varied q_H with constant B , either that of TIP3P water ($B =$

2489 Å⁶kJ/mol) or $B = 0$ (no dispersive attraction). Alternatively, we varied B with constant q_H , either that of TIP3P water ($q_H = 0.417 e$) or almost zeroed ($q_H = 0.01 e$; if q_H is zeroed, the H atoms become non-interacting dummy atoms). Thus we can compare purely dispersive and purely electrostatic variants with a realistic water model.

In all four cases, we obtain enthalpy–entropy curves of the same type as the diatomic systems in Figures 1c, 2c, and 3c, i.e. with a sharp bend and saturation at large negative enthalpies, as can be seen in Figure 4. The TIP3P parameters are at the beginning of the saturation plateau and quite close to the QM result. All four curves show strict EEC2, but as the enthalpy (but not the energy) is positive for the weakest interactions, we sometimes do not have EEC1. However, for realistic parameters, $q_{SH} \approx 1.5$. It can be seen that dispersion has a minimal effect on this interaction – the curves for $B = 0$ and $B = 2489 \text{ Å}^6\text{kJ/mol}$ are essentially identical, as is the one for $q = 0.417 e$. It is only the curve for $q = 0.01 e$ that differs and only by a simple translocation of the curve to less negative entropies – the shape is the same.

Looking at the components, $\Delta H_{\text{trans}} = -6 \text{ kJ/mol}$ as always and we now have a non-vanishing rotational contributions of the monomers, leading to a negative $\Delta H_{\text{rot}} = -4 \text{ kJ/mol}$, which will be the same for all large molecules. $T\Delta S_{\text{trans}}$ is constant, -41 kJ/mol , whereas $T\Delta S_{\text{rot}}$ shows a small variation, depending on the length of the hydrogen bond, 3 to -1 kJ/mol , most negative for the strongest complexes (cf. Figure 4b). ΔH_{vib} is positive and quite large, 15–31 kJ/mol, increasing with the binding energy. Therefore, it again gives rise to some energy–enthalpy compensation. $T\Delta S_{\text{vib}}$ shows the largest variation, 7–45 kJ/mol, decreasing when the complex becomes stronger. It is the prime cause of the EEC2, compensating for the loss of translational and rotational freedom. It is also causes the diverging results for the $q = 0.01 e$ curve, by being $\sim 25 \text{ kJ/mol}$ more positive, owing to five very small vibrational frequencies ($< 20 \text{ cm}^{-1}$; the smallest frequency for the water dimer is 142 cm^{-1} at the QM level). This is mainly an artefact of the missing dispersion interactions of the hydrogen atoms.

Consequently, this model show that the same type of EEC curves are obtained also for more realistic multi-atom systems and it gives an indication of the sign of the various components in such systems. Dispersion has a minimal influence on the results for the interaction of two water-like molecules, whereas the magnitude of the charges determines the absolute scale of the entropy, without changing the shape of the curves.

Larger dispersive and hydrogen-bonded systems

So far, we have seen that there is no significant difference between EEC studied at the QM and MM levels and that there are no fundamental difference between dispersive, electrostatic, hydrogen-bonded, or even covalent systems, only quantitative differences caused by the magnitude of ΔH , i.e. the position on the saturation curve. We will now check that these results are not modified when there are several simultaneous interactions between the two binding molecules. We will continue to study (mainly) dispersive and electrostatic (hydrogen-bonded) interactions separately.

As models of dispersive interactions with an increasing strength, we have studied at the MM level dimers of the seven first n -alkanes and the three simplest branched alkanes, as well as eight stacked dimers of simple polyaromatic systems (with 1–6 benzene rings). All had zeroed charges to ensure that the interactions are purely dispersive. The results in Figure 5 show that for the alkanes, there is an EEC1 (except for the methane and ethane dimers, for which the enthalpy is positive) and a reasonable EEC2 with a correlation $r^2 = 0.83$ and a slope of 1.7. In agreement with Figure 1b, there is an indication that the points with the least negative entropy show a larger slope than those with more negative enthalpies.

The aromatic systems show a nearly perfect EEC2 ($r^2 = 0.98$). Compared to the alkanes,

the slope of the correlation line is appreciably lower, 0.3, reflecting that the complexes are appreciably stronger ($\Delta E = -15$ to -94 kJ/mol) and therefore lie further to the left on the saturation curve. However, the larger ΔE comes entirely from the increased number of atoms – with the GAFF force field, benzene gives a less favourable interaction energy and enthalpy (-15 and -10 kJ/mol) than *n*-hexane (which contains the same number of carbon atom; -18 and -12 kJ/mol), but the entropy is smaller ($T\Delta S = -31$ compared to -44 kJ/mol), owing mainly to the vibrational term and the higher symmetry of benzene.

Looking at the enthalpy components, the translational and rotational enthalpies are constant at $\Delta H_{\text{trans}} = -6$ kJ/mol and $\Delta H_{\text{rot}} = -4$ kJ/mol as for all larger systems. The vibrational enthalpy is also essentially constant, increasing by less than 1 kJ/mol for the largest systems ($\Delta H_{\text{vib}} = 15$ – 16 kJ/mol; Figure 5b).

The translational entropy shows a restricted variation, increasing from -40 to -51 kJ/mol when the size of the molecules increases (Figure 5b). $T\Delta S_{\text{rot}}$ shows a similar, but larger variation, increasing from -1 to -35 kJ/mol with the size of the molecule. For the alkanes, the EEC2 comes entirely from these two terms, which give correlations of 0.74 and 0.69 to the enthalpy, respectively. Thus, the EEC2 is a pure size effect, which is expected as dispersion is present everywhere and therefore will increase with the size of the molecules. For the aromatic systems, the vibrational entropy also contributes to the EEC2, with a variation of 8 kJ/mol and a decent correlation to the enthalpy ($r^2 = 0.64$). Thus, we can conclude that realistic dispersive systems also show pronounced EEC2, but different homologous series can give different slopes, depending on the strength of the interactions.

Next, we consider two sets of hydrogen-bonded systems, which are shown in Figure 6a. The first is dimers of linear glycine oligomers, as a model of a protein β sheet. We studied oligopeptides with 1, 2, 4, 6, 8, 10, 12, and 14 units and no capping groups. The larger systems have $n - 2$ hydrogen bonds, where n is the number of glycine units, whereas the three smallest systems all have two hydrogen bonds (they were truncated differently). The second was a set of dimers of polycarboxylic acids with 1–5 propionic acid units (i.e. with three carbon atoms between each pair of acid groups). There were two hydrogen bonds between each pair of acidic group in the dimer.

The results in Figure 6a shows that the β sheets give an almost perfect EEC2 with a correlation between the enthalpy and $T\Delta S$ of 0.99 and a slope of 0.18. For the polypropionic acid dimers, there is also an EEC2, but it is much less linear, $r^2 = 0.89$ and the slope is 0.56. The reason for this is that the hydrogen bonds become increasingly less favourable at the terminals of the larger dimers, owing to internal strain in the carbon backbone. In fact, with six units, the last hydrogen bond breaks. The three smallest dimers give a better correlation with $r^2 = 0.97$ and a slope of 0.40.

All three entropy components show a good correlation to the total entropy, $r^2 = 0.89$ – 0.99 , with $T\Delta S_{\text{trans}}$ and $T\Delta S_{\text{vib}}$ giving the best correlation and $T\Delta S_{\text{rot}}$ giving the largest contribution to the variation in the total entropy (Figure 6b). It is notable that $T\Delta S_{\text{vib}}$ changes sign from positive to negative for the largest systems of both models, so that it no longer counteracts the other two terms, but instead contributes to EEC1. Interestingly, for these strong complexes, the vibrational frequencies become so large that also ΔH_{vib} shows a significant variation, 17–28 kJ/mol, with a good correlation to the total enthalpy ($r^2 = 0.96$ – 0.98). It therefore gives rise to a significant *energy*–enthalpy compensation.

It is interesting that the three smallest β -sheet systems follow the same linear EEC2 although they all have two hydrogen bonds. This is caused by both electrostatic and dispersive interactions: The pair-wise difference in binding energy between the $n = 1$ and 2 complexes (19 kJ/mol) comes mainly (14 kJ/mol) from electrostatics, with a contribution of 6 kJ/mol

from bonded interactions (the smallest complex is strongly distorted by the two hydrogen bonds) and an opposing 1 kJ/mol contribution from dispersion. The difference between the $n = 2$ and 4 complexes (58 kJ/mol) comes mainly (46 kJ/mol) from electrostatics, but the dispersion also contributes by 13 kJ/mol, whereas the bonded contribution is only 1 kJ/mol and counteracting. For all the other complexes, adding two Gly residues increases the interaction energy by 63 kJ/mol, of which 53, 13, and -2 kJ/mol come from electrostatics, dispersion, and bonded interactions, respectively.

Thus, we can conclude that for both the dispersive and the hydrogen-bonded systems, a pronounced EEC2 is observed. We do not observe any qualitative difference between the dispersive and hydrogen-bond interactions, besides that the latter are typically stronger. This is in accordance with our results in the previous sections of this article.

Competition with water

So far, all studied systems (except the methane and ethane dimers with positive ΔH) have shown EEC1 and most of them also positive correlation between ΔH and $T\Delta S$, indicating EEC2. We have also seen how such EEC arises naturally either from the increase in $T\Delta S_{\text{trans}}$ and $T\Delta S_{\text{rot}}$ with the molecular mass (and therefore with more interactions) or from $T\Delta S_{\text{vib}}$ increasing with the strength of the interactions. Why is then not EEC always observed in all molecular systems? A natural answer to this question is that all the calculations were performed for isolated molecules in vacuum, whereas real interactions take place in water solution. To examine this possibility, we have also performed some tests of how the addition of water molecules may affect the results for some types of interactions. Of course, the present approach cannot properly model the full solvation of the molecules or the dynamics of water, but we believe that it can illustrate some of the expected effects of water.

For the series of nine alkane dimers, we added one water molecule to each monomer, ensuring that the two water molecules in the dimer form a hydrogen bond. The results in Figure 7 shows that the calculations with water give more negative enthalpies than those without water by ~ 24 kJ/mol, i.e. the enthalpy of the new hydrogen bond between the two water molecules. $T\Delta S$ also becomes more negative by a similar amount. For the nine complexes with water, there is only a faint EEC2 ($r^2 = 0.16$). However, this is mainly caused by a single outlier (isopentane); if it is excluded, $r^2 = 0.78$ and the slope is 1.5. In fact, this is quite close to the EEC correlation line of the alkanes without water, with a slope of 1.0 for the joint systems ($r^2 = 0.82$).

We performed also a similar study of the polypropionic-acid systems, for which we added one water molecule to each acidic group, i.e. 1–6 for the various monomers. No attempt was made to add the water molecules at any systematic positions. Instead, they were added more or less random from a pre-equilibrated box of water molecules according to their distance to the acidic molecule. The results of these calculations are also shown in Figure 7. It can be seen that in this case, the solvated calculations fall in a similar range as those without water. This is expected, because hydrogen bonds with water and neutral carboxylic acids have a similar strength. However, there is a tendency that the calculations with water give a less negative enthalpy and a more negative entropy.

At first, the non-systematic manner in which the water molecules were added may seem to be a disadvantage. It is the reason why the correlation between ΔH and $T\Delta S$ is modest ($r^2 = 0.49$). However, in practice, this unsystematic study gives more information of the variability of the hydrogen-bond strength and how the enthalpy and entropy terms vary with this strength. The enthalpy and entropy contributions for the six complexes in Table 1 show that $T\Delta S_{\text{trans}}$ and $T\Delta S_{\text{rot}}$ become somewhat more negative as the size of the complex increases (-47 to -53 and

–28 to –44 kJ/mol, respectively). The former term depends only on the molecular mass of the complex, whereas the latter varies also with the geometry (compactness) of the complex. However, only the two vibrational terms show a variation coupled to the strength of the formed hydrogen bonds. ΔH_{vib} show a rather small variation, 17–25 kJ/mol, slightly increasing with the size of the complexes, but more anticorrelated to the strength of the complex ($r^2 = 0.79$ with respect to ΔE). Thus, it slightly counteracts the binding energy, but the slope of the correlation line is only –0.06, so the effect is quite small. $T\Delta S_{\text{vib}}$ shows a larger variation, 27 to –2 kJ/mol, but it is more correlated to the size of the complex than to ΔH ($r^2 = 0.46$). This is the reason for the quite poor correlation between the net ΔH and $T\Delta S$: the entropy terms show only a weak correlation to the binding energy and the variation of the ΔH_{vib} term is much smaller than the variation in ΔE . This also explains why most points with water lie below the old correlation line between ΔH and $T\Delta S$: When the hydrogen bonds formed in the dimer are not ideal, ΔE and therefore also ΔH becomes less negative. However, both $T\Delta S_{\text{trans}}$ and $T\Delta S_{\text{rot}}$ depend mainly on the size of the complex, so they will still become more negative and therefore actually overcompensating the enthalpy.

The chelate effect

Finally, we will to shortly discuss the chelate effect, which has been discussed in several publications [16,18,19,20,21]. As already mentioned, the formation of a complex from two free molecules leads to major unfavourable translational and rotational entropy terms. However, this term is paid only once when the complex is formed, i.e. when the first interaction between the two molecules is formed. If there are additional interactions between the two molecules, these should have smaller entropy penalties, coming mainly from the vibrational part.

This can be illustrated by N-methylacetamide. In the dimeric complex, it can form either one or two hydrogen bonds, giving $\Delta E = -32$ or -57 kJ/mol at our QM level. As the complexes consists of the same molecules, ΔH_{trans} , ΔH_{rot} , and $T\Delta S_{\text{trans}}$ are all identical for the two complexes. Moreover, ΔH_{vib} agree to within 0.3 kJ/mol (Table 2). On the other hand, the complex with two hydrogen bonds is more compact so that $T\Delta S_{\text{rot}}$ is 2 kJ/mol more negative for that complex. Likewise that complex is slightly stronger (both NH–O hydrogen bonds are 1.81 Å, compared to 1.90 Å for the single hydrogen bond in the other complex), so $T\Delta S_{\text{vib}}$ is also 2 kJ/mol more negative. Consequently, when the first hydrogen bond is formed, the quite favourable energy ($\Delta E = -32$ kJ/mol) is more than compensated by the unfavourable ΔH_{vib} (16 kJ/mol) and entropy terms ($T\Delta S = -46$ kJ/mol) so that the net binding free energy is unfavourable ($\Delta G = 21$ kJ/mol). However, when the second hydrogen bond forms, all these costs have already been paid and the extra energy of the second hydrogen bond ($\Delta E = -26$ kJ/mol) is only counteracted by a slight increase in the ΔH_{vib} , $T\Delta S_{\text{rot}}$, and $T\Delta S_{\text{vib}}$ terms, 4 kJ/mol in total. Therefore, the free energy becomes favourable ($\Delta G = -2$ kJ/mol).

Do we see this effect also in the other systems? In fact, the effect is not qualitatively different from the effects seen in the other homologous series (where we also pay the prime entropic penalty in the first complex and then only add smaller increases in the entropy as the complexes are made larger). Seen in this way, the N-methylacetamide complexes show a 4-kJ/mol decrease in $T\Delta S_{\text{rot}} + T\Delta S_{\text{vib}}$ as ΔE or ΔH become more negative by 26 kJ/mol. This is an EEC2 with a slope of 0.14, i.e. similar to what was found for the β -sheets (0.18). All homologous series will take some advantage of the chelate effect, but they will still show an

EEC2, as $T\Delta S_{\text{vib}}$ often $T\Delta S_{\text{rot}}$ becomes more negative with the strength of the complex. Moreover, the effect is blurred by the fact that for real molecules in homologous series, the mass increases when more interactions are formed, leading to an increase also in $T\Delta S_{\text{trans}}$ and $T\Delta S_{\text{vib}}$.

Discussion

In this section, we give a general discussion of all our results (QM or MM results for 89 real molecular systems and MM results for a large number of simplified analytical models). We will first discuss which systems and energy components give rise to the EEC1, then how EEC2 arises. Finally, we will compare our results with those obtained by Korth [25] with QM methods for the S22 [43] and S66 [44] sets of interacting molecules and discuss whether different types of interactions give rise to different types of EEC.

Among our 89 molecular systems, ΔH is negative for all, except two, viz. the methane and ethane dimers, which have slightly positive ΔH (3 and 1 kJ/mol) owing to the ΔH_{vib} term (ΔE is negative, -2 and -4 kJ/mol). On the other hand, $T\Delta S$ is negative for all systems (-9 to -120 kJ/mol). Thus all systems show an *energy-entropy* compensation and all except two show EEC1. The same is seen in Korth's data (he also obtained a positive enthalpy for the methane dimer). The quotient q_{SH} shows a large variation, ranging from 29 for the propane dimer (owing to the low ΔH) to 0.3 for the largest β sheets and 0.03 for LiF.

Looking at the enthalpy components, ΔH_{trans} is always -6 kJ/mol, favouring the bonding and enhancing the enthalpy. For the diatomic complexes, ΔH_{rot} is positive (+2 kJ/mol), counteracting the binding, but for all more realistic complexes, it is -4 kJ/mol, enhancing the enthalpy. On the other hand, ΔH_{vib} is always positive, thereby counteracting the binding. For Korth's complexes of rather similar sizes, it is almost constant, 13-17 kJ/mol, but in our diatomic complexes it can be as small as 2 kJ/mol and for the largest hydrogen-bonded systems it can increase to 29 kJ/mol. Thus, the net enthalpy correction is in general positive and often ~ 5 kJ/mol, although it is -1 for several of the diatomic complexes, up to 19 kJ/mol for the systems with many hydrogen bonds, and 23 kJ/mol for H_2 .

$T\Delta S_{\text{trans}}$ is always negative, -30 kJ/mol to -55 kJ/mol, although it continues to decrease logarithmically (become more negative) with the size of the complex. It is numerically the largest entropy term and therefore always gives the largest contribution to the EEC1. $T\Delta S_{\text{rot}}$ is positive (4-26 kJ/mol) for the diatomic complexes, counteracting the EEC1. However, for the more typical polyatomic complexes, it is negative and therefore contributing to the EEC1. It increases roughly with the size of the complex, varying between 0 kJ/mol (water dimer) and -49 kJ/mol for the largest model, thus providing a sizeable contribution to the EEC1. Finally, $T\Delta S_{\text{vib}}$ may have a varying sign. For most complexes, it is positive and 10-40 kJ/mol. However, for the periodic systems with an increasing number of hydrogen bonds, it decreases with size and eventually becomes negative for the five largest β sheets and acid dimers. Thus, for most complexes, this term compensates the loss of translational and rotational freedom, reducing the EEC1, but for the largest models it may actually contribute to EEC1.

The EEC2 is instead determined by the differences between two complexes in a homologous series. Of course, this somewhat changes the conclusions; for example, both ΔH_{trans} and ΔH_{rot} are constant and therefore do not contribute to EEC2. However, it also means that the results depend on how you define a homologous series.

If the mass of the complexes do not change, e.g. in the chelate effect or for the toy systems in Figures 1-4, $T\Delta S_{\text{trans}}$ is also constant and does not contribute to the EEC2.

However, in most homologous series (e.g. in all our series of real molecules), the mass increases. Then, both $T\Delta S_{\text{trans}}$ and $T\Delta S_{\text{rot}}$ become more negative, but typically by only a few kJ/mol (in our cases, by 0–8 kJ/mol, often slightly more for $T\Delta S_{\text{rot}}$ than for $T\Delta S_{\text{trans}}$). Still, it is not evident that this will lead to an EEC2, because it depends on how ΔE changes in the series. For the ionic QM systems in Figure 2c and the covalent QM systems in Figure 3c, the strength of the interaction actually decreases with the mass, so that $T\Delta S_{\text{trans}}$ and $T\Delta S_{\text{rot}}$ give rise to a (weak) enthalpy–entropy anti-compensation. For the noble gases and the systems with added water molecules, the trend of ΔE with respect to the mass is varying, so that no conclusion can be reached regarding the EEC2. The same also applies if isomers are included in the alkane and aromatic series (some of the complexes have the same masses, but different ΔE). However, for the pure alkane, polyaromatic, β -sheet, and polypropionic-acid series, the $T\Delta S_{\text{trans}}$ and $T\Delta S_{\text{rot}}$ terms give significant contributions to EEC2. The same actually also applies for $T\Delta S_{\text{rot}}$ in some systems with a constant mass (the model systems in Figures 1–4 and the chelate complex), because there is direct relation between the strength of the complex and the interaction distance, which contributes to the compactness of the complex (the moment of inertia) and therefore to $T\Delta S_{\text{rot}}$.

The ΔH_{vib} term gives a varying contribution in the various homologous series. For weak complexes (noble gases, dispersion, alkane, and polyaromatic series), it gives an insignificant contribution. For the ionic and covalent diatomic systems studied by QM, it gives a negative contribution of –1 to –6 kJ/mol, thereby counteracting ΔE (which becomes more positive in these homologous series). For the large hydrogen-bonded systems, it instead gives a positive contribution of 0–4 kJ/mol, but again it means that it counteracts ΔE and therefore gives an energy–enthalpy compensation.

Finally, the $T\Delta S_{\text{vib}}$ term is also varying. For the two-atom toy systems, it is always negative and up to 11 kJ/mol. However, for the real (QM) two-atom systems, it changes very little (less than 1 kJ/mol). For the alkanes, it gives an alternating sign, ± 5 kJ/mol. However, for the other larger systems, it is in general negative and for the hydrogen-bonded systems rather large, –6 to –10 kJ/mol, giving rise to a clear EEC2. Thus, this term changes its effect, depending on how the EEC is defined: In absolute terms, it typically counteracts the other two entropy components and leads to an anti-EEC1, but for most homologous series it enhances EEC2 and often is the dominant entropy term in that respect.

Finally, we consider the data in the same way as Korth [25], i.e. looking for linear correlations in a scatter plot of the absolute ΔH and $T\Delta S$ results of all systems, show in Figure 8. This is essentially some sort of EEC2, including all molecules in the “homologous” series. However, following Korth, we will group them after the dominant type of interaction. Considering all complexes (inset), a major difference in the correlation of the covalent and ionic complexes on one hand (with a slope of 0.01) and all the other complexes (with an average slope of 0.3) is apparent. However, as we saw in Figures 1–3, this does not indicate any fundamental difference between the different types of interactions; instead it only reflects that these interactions are stronger and therefore are on the saturation plateau on the ΔH and $T\Delta S$ curve. The β sheets are also conspicuous owing to the large enthalpies and the nice linear correlation (slope 0.18).

Concentrating on the other complexes with more moderate enthalpies, it is hard to see any clear-cut difference between systems dominated by dispersion or hydrogen bonds. It can be seen that there are a group of hydrogen-bonded systems (with enthalpies around 30 kJ/mol) that fit quite well into the group and slope of the majority of our and Korth's dispersion-dominated systems. Likewise, our data for the polyaromatic complexes (with a slope of 0.3;

dispersion points with ΔH of -40 to -90 kJ/mol in Figure 8) fit quite nicely into both our and Korth's data of hydrogen-bonded complexes. In particular, the slope is lower than for the polypropionic-acid systems (0.56). This confirms our conclusion that there are no clear-cut difference between complexes dominated by dispersive or hydrogen-bond interactions. Instead, individual homologous series have varying slopes and the net slopes obtained from a scatter plot of a large number of complexes mainly depends on the selection of complexes (which series dominate the results).

Conclusions

In this paper we have studied fundamental aspects of enthalpy–entropy compensation based on minimised structures and vibrational frequencies calculated with QM and MM methods and the RRHO approach, in a manner similar to what has been done by Williams, Dunitz, and Korth before [18-25]. Our aim has been to decide whether there is any fundamental difference between dispersive and hydrogen-bond interactions, between QM and MM calculations, and to discuss why EEC is observed in some cases, but not all.

At the MM level, we can study systems that are purely dispersive or electrostatic. Comparing the curves in Figures 1 and 2, it is clear that there is no qualitative difference between the two types of interactions: they both show that $T\Delta S$ first rapidly decreases and then levels out to saturation when ΔH becomes more negative. In fact, also covalent bonds give similar curves (Figure 3). The only difference is the magnitude of ΔE (and therefore ΔH), which is two orders of magnitude larger for the covalent and ionic bonds.

For homologous series of more realistic molecules dominated by dispersive (alkane or polyaromatic dimers) or hydrogen bonds (polyglycine sheets or polypropionic-acid dimers), nearly linear relations between ΔH and $T\Delta S$ were obtained, with slopes of 0.2–1.7. There is a tendency that complexes dominated by hydrogen bonds give a smaller slope, but this is mainly caused by the fact that hydrogen bonds typically provide a larger gain in enthalpy for the same number (mass) of added atoms, i.e. that they are further to the left in the saturation plateau in Figures 1c and 2c. For strong dispersive interactions as in the polyaromatic systems, the slope (0.3) is actually smaller than what is obtained in many hydrogen-bonded systems, as is evident if all studied complexes are included in the same plot (Figure 8).

We have emphasized the importance to define what is meant by EEC and we have used two different definitions. EEC1 simply states that ΔH and $T\Delta S$ have the same sign for a certain reaction. In our simple binding reactions, where two molecules form a complex in vacuum, this is the case for all 89 dimers, except the two smallest alkanes (for which ΔH happens to be positive). This EEC1 is caused by the loss of translational and vibrational freedom (entropy), partly counteracted by the gain in vibrational entropy, as has already been suggested by Jenks, Williams, and Dunitz [18-25]. However, in real binding reactions, it will be blurred, because both the ligand and the receptor are in water solution, rather than in gas phase, which is expected to reduce the translational and rotational freedom. Moreover, it has to be taken into account that water molecules may be associated with the ligand and the binding site before the binding and that some of these may be displaced upon binding.

We have also discussed EEC2, i.e. how ΔH and $T\Delta S$ change in homologous series of compounds. This is probably closer to the common meaning of EEC. Such a definition gives the advantage that several terms cancel ($\Delta\Delta H_{\text{trans}}$, $\Delta\Delta H_{\text{rot}}$, and possibly $T\Delta\Delta S_{\text{trans}}$). On the other hand, it must be defined what is meant by a homologous series and the results will depend on that choice. We have shown that for truly homologous series, where the same structural unit and identical interactions are successively added, nearly perfect linear relations between ΔH

and $T\Delta S$ can be obtained. Such an EEC2 is caused by joint effect of all three entropy terms. Thus, for EEC2, $T\Delta\Delta S_{\text{vib}}$ typically gives a contribution of the same sign as $T\Delta\Delta S_{\text{trans}}$ and $T\Delta\Delta S_{\text{rot}}$ and it gives a dominant contribution at least for the hydrogen-bonded systems. For weaker systems, the EEC2 is dominated by the $\Delta\Delta H_{\text{trans}}$ and $\Delta\Delta H_{\text{rot}}$ terms, i.e. by pure size effect, rather than the strength of the complex (which is correlated with $T\Delta\Delta S_{\text{vib}}$).

However, it is apparent from this study that EEC of any kind is not expected to be strict in real binding studies: Only for perfect homologous series are nice linear relations obtained. As soon as the interactions change (e.g. as in the larger polypropionic-acid dimers), the relation is blurred and if water molecules are added in a non systematic way, the correlation disappears and anticorrelation can easily be obtained between certain pairs. If the complete dynamic effect of water solvation is included any relation between ΔH and $T\Delta S$ can be expected.

In conclusion we do not see any fundamental difference between pure dispersive or electrostatic (hydrogen-bonded) interactions, nor between QM or MM calculations. The RRHO model gives a clear indication when and how EEC1 and EEC2 arise during the binding of two molecules in gas phase. However, the real binding reaction between a ligand and a receptor in water solution is much more complicated, with water molecules giving large effects which easily can blur the expected EEC and even giving anticorrelation. Unfortunately, such interactions are determined by the details of the ligand and the receptor [45] (and the dynamics of the system) making it hard to reach any general conclusions.

Acknowledgements

This investigation has been supported by grants from the Knut and Alice Wallenberg Foundation (KAW 2013.0022) and the Swedish research council (project 2010-5025), as well as by computer resources provided by the Swedish National Infrastructure for Computing (SNIC) at Lunarc, Lund University.

References

- 1 L. Liu, Q.-X. Guo, *Chem. Rev.* 2001, **101**, 673.
- 2 K. Sharp, *Prot. Sci.* 2001, **10**, 661-667.
- 3 J. D. Chodera, D. A. Mobley, *Ann. Rev. Biophys.* 2013, **42**, 121.
- 4 O. Exner, *Chem. Commun.* 2000, 1655.
- 5 A. Cooper, C. M. Johnson, J. H. Lakey, M. Nöllmann, *Biophys. Chem.* 2001, **93**, 215.
- 6 H. Qian, J. J. Hopfield, *J. Chem. Phys.* 1996, **105**, 9292.
- 7 E. B. Starikov, B. Nordén, *Chem. Phys. Lett.*, 2012, **538**, 118.
- 8 C.-E. Chang, M. K. Gilson, *J. Chem. Theory Comput.*, 2012, **8**, 13156.
- 9 S. Moghaddam, Y. Inoue, M. K. Gilson, *J. Am. Chem. Soc.* 2009, **131**, 4012.
- 10 C. H. Reynolds, M. K. Holloway, *ACS Med. Chem. Lett.*, 2011, **2**, 433.
- 11 T. S. G. Olsson, J. E. Ladbury, W. R. Pitt, M. A. Williams, *Protein Science* 2011, **20**, 1607.
- 12 C. Forrey, J. F. Douglas, M. K. Gilson, *Soft Matter*, 2012, **8**, 6385.
- 13 E. Gallicchio, M. M. Kubo, R. M. Levy, *J. Am. Chem. Soc.* 1998, **120**, 4526.
- 14 G. Graziano, *J. Chem. Phys.* 2004, **120**, 4467.
- 15 D. M. Ford, *J. Am. Chem. Soc.* 2005, **127**, 16167.
- 16 D. H. Williams, E. Stephens, D. P. O'Brien, M. Zhou, *Angew. Chem. Int. Ed.* 2004, **43**, 6596.
- 17 S. Moghaddam, Y. Inoue, M. K. Gilson, *J. Am. Chem. Soc.* 2009, **131**, 4012.

- 18 W. P. Jencks, Proc. Natl. Acad. Sci. USA, 1981, 78, 4046.
- 19 M. S. Searle, M. S. Westwell, D. H. Williams, J. Chem. Soc. Perkin Trans. 2, 1995, 141.
- 20 M. S. Searle, D. H. Williams, J. Am. Chem. Soc., 1992, 114, 10690.
- 21 M. S. Searle, D. H. Williams, Nucleic Acids Res., 1993, 21, 2051.
- 22 D. H. Williams, M. S. Westwell, Chem. Soc. Rev. 1998, 27, 57.
- 23 M. S. Westwell, M. S. Searle, J. Klein, D. H. Williams, J. Phys. Chem. 1996, 100, 16000.
- 24 J. Dunitz, Chem. Biol. 1995, 2, 70
- 25 M. Korth, Med. Chem. Commun. 2013, 4, 1025.
- 26 F. Jensen, Introduction to Computational Chemistry, 1999, J. Wiley & Sons Ltd, Chichester, p. 298-306.
- 27 D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman (2006), AMBER 9, University of California, San Francisco.
- 28 J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman and D.A. Case. Development and testing of a general Amber force field. J. Comput. Chem. 25, 1157-1174 (2004).
- 29 Jakalian, A, Bush, BL, Jack, DB, Bayly, CI J Comput Chem, 2000, 21, 132-146
- 30 Jakalian, A, Jack, DB, Bayly, CI J Comput Chem , 2002 , 23 , 1623-1641
- 31 Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct. Bioinform.* **2006**, 65, 712-725.
- 32 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
- 33 R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel (1989) Chem. Phys. Lett. 162: 165
- 34 O. Treutler, R. Ahlrichs (1995) J. Chem. Phys. 102: 346
- 35 Tao, J., Perdew, J.P., Staroverov, V.N., Scuseria, G.E, Phys. Rev. Lett. 91 (2003), 146401.
- 36 F. Weigend & R. Ahlrichs, Phys. Chem. Chem. Phys. 7 (2005) 3297-3305.
- 37 D. Rappoport, F. Furche, J. Chem. Phys. 133 (2010) 134105.
- 38 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, J. Chem. Phys, 132 (2010), 154104
- 39 K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs Chem. Phys. Lett. 242: 652 (1995)
- 40 K. Eichkorn, F. Weigend, O. Treutler, R. Ahlrichs, Theor. Chem. Acc. 97: 119 (1997)
- 41 F. Weigend, Phys. Chem. Chem. Phys. 8: 1057 (2006)
- 42 Grimme S (2012) Chem Eur J 18:9955
- 43 P. Jurecka, J. Sponer, J. Cerny, P. Hobza; Phys Chem Chem Phys 2006, 8, 1985-1993
- 44 Řezáč, J., Riley, K.E., Hobza, P., J. Chem. Theory Comput. 7 (2011) 2427-2438.
- 45 S. Genheden, P. Mikulskis, L. Hu, J. Kongsted, P. Söderhjelm, U. Ryde J. Am. Chem. Soc., 2011, 133, 13081-13092.

Table 1. Enthalpy and entropy contributions for the six dimers of polypropionic-acid oligomers (with 1–6 units, n) in Figure 6 with n water molecules. All contributions are defined in Eqns. 3–8, except $H_{\text{trv}} = H_{\text{trans}} + H_{\text{rot}} + H_{\text{vib}}$, and ΔH_{b} is the increase in the number of hydrogen bonds (non-bonded OH–O interactions shorter than 2.5 Å) when the complex is formed.

n	ΔE	ΔH_{trans}	ΔH_{rot}	ΔH_{vib}	$T\Delta S_{\text{trans}}$	$T\Delta S_{\text{rot}}$	$T\Delta S_{\text{vib}}$	ΔH_{trv}	ΔTS	ΔH_{b}
1	-34.9	-6.2	-3.7	17.0	-46.6	-27.5	26.5	7.1	-47.7	0
2	-106.1	-6.2	-3.7	20.7	-49.2	-34.1	16.5	10.8	-66.8	5
3	-144.5	-6.2	-3.7	24.7	-50.7	-37.8	9.8	14.8	-78.7	4
4	-106.7	-6.2	-3.7	22.8	-51.7	-39.9	2.1	12.9	-89.5	4
5	-125.7	-6.2	-3.7	20.9	-52.6	-42.4	-2.0	11.0	-97.0	7
6	-105.2	-6.2	-3.7	22.7	-53.2	-44.1	-1.2	12.7	-98.5	3

Table 2. Hydrogen-bond distance (r_{Hb} in Å), as well as the energy, free energy, enthalpy, and entropy contributions (in kJ/mol) for dimeric complexes of N-methylacetamide forming one or two NH–O hydrogen bonds (Hb), as well as their difference (Diff.). All contributions are defined in Eqns. 3–8, except $H_{\text{trv}} = H_{\text{trans}} + H_{\text{rot}} + H_{\text{vib}}$.

System	r_{Hb}	ΔE	ΔH_{trans}	ΔH_{rot}	ΔH_{vib}	$T\Delta S_{\text{trans}}$	$T\Delta S_{\text{rot}}$	$T\Delta S_{\text{vib}}$	ΔH_{trv}	ΔTS	ΔG
1 Hb	1.90	-31.6	-6.2	-3.7	15.9	-45.8	-25.3	24.6	6.0	-46.5	20.9
2 Hb	1.81	-57.5	-6.2	-3.7	15.6	-45.8	-27.3	23.0	5.7	-50.1	-1.7
Diff.		-25.9	0.0	0.0	-0.3	0.0	-2.0	-1.6	-0.3	-3.6	-22.6

Figure 1. (a) Enthalpy, entropy, free energy (left axis), and the quotient q_{SH} (right axis) as a function of the dispersion parameter B for the purely dispersive interaction of two neutral argon-like atoms with $m_1 = m_2 = 40$ Da and $A = 2.2 \times 10^7 \text{ \AA}^{12} \text{ kJ/mol}$. (b) The enthalpy and entropy contributions (left axis), as well as the distance between the two atoms (right axis) as a function of ΔH for the same values of m_1 , m_2 , and A . (c) Correlation between the enthalpy and entropy as a function of B for different values of A (still with $m_1 = m_2 = 40$ Da). QM results for noble-gas dimers (He_2 , Ne_2 , Ar_2 , Kr_2 , Xe_2 , and Rn_2) and pairs (HeNe , NeAr , NeKr , NeXe , and NeRn) are also included with symbols.

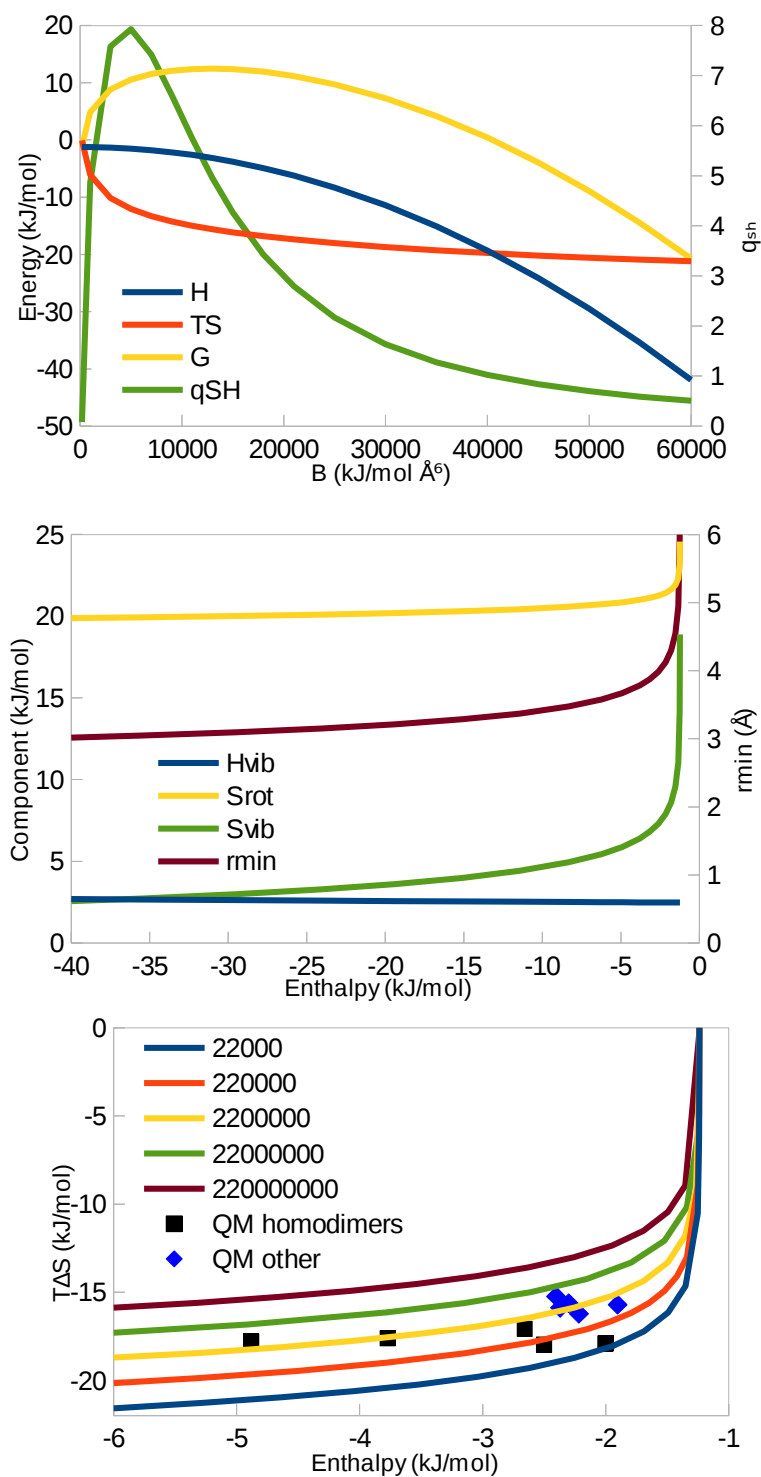


Figure 2. (a) Enthalpy, entropy, free energy (left axis), and q_{SH} (right axis) as a function of the charge q for the interaction of two oppositely charged ions with $A = 1.0 \times 10^6 \text{ \AA}^2 \text{kJ/mol}$. (b) The enthalpy, entropy terms, and the minimum distance (r_{min}) as a function of ΔH for the same value of A . (c) Enthalpy and entropy as a function of A with $q = 1.0 e$. QM results for LiF, NaCl, KBr, RbI, CsAt, NaF, KF, RbF, and CsF are also included as black squares. The inset shows a magnification of the QM results. In all figures, $m_1 = 23 \text{ Da}$ (Na) and $m_2 = 35 \text{ Da}$ (Cl).

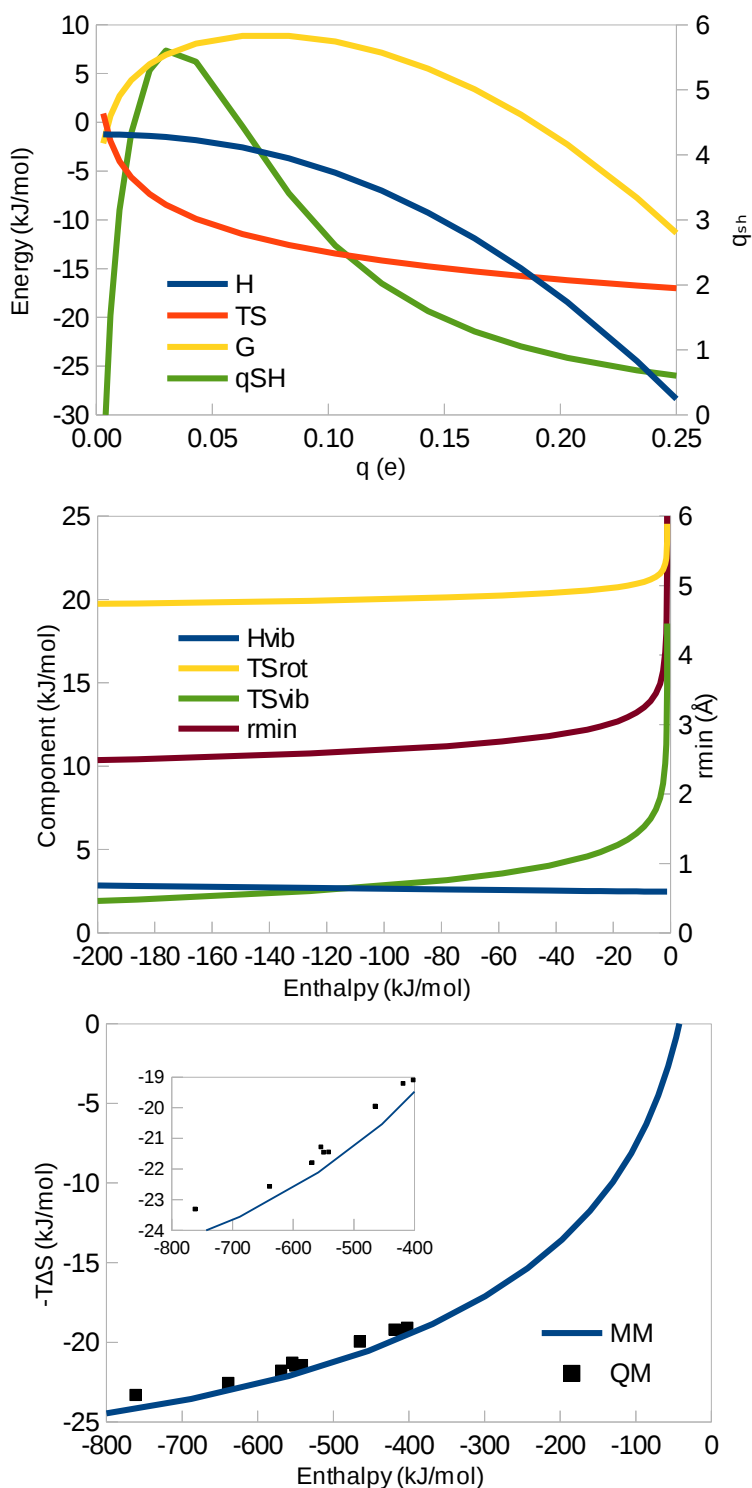


Figure 3. (a) Enthalpy, entropy, free energy, and q_{SH} as a function of D for the formation of a covalent bond with $m_1 = m_2 = 19$ Da (F), $C = 8.87 \text{ \AA}^{-1}$, and $r_0 = 1.42 \text{ \AA} \times 181.6 \text{ kJ/mol} / D$ (which reproduces the QM results for F_2 , $D = 181.6 \text{ kJ/mol}$, $r = 1.42 \text{ \AA}$, and $\nu = 1004 \text{ cm}^{-1}$). (b) The enthalpy, entropy terms, and the minimum distance (r_{min}) as a function of ΔH for the same parameters. (c) Correlation between the enthalpy and entropy as a function of D for different values of r_0 and C (indicated in units of \AA and \AA^{-1} in the legend; still with $m_1 = m_2 = 19$ Da). QM results for HF, HCl, HBr, HI, HAt, HO^- , HS^- , HSe^- , HTe^- , HPO^- , H_2 , Li_2 , C_2 , N_2 , O_2 , and F_2 are also included with symbols.

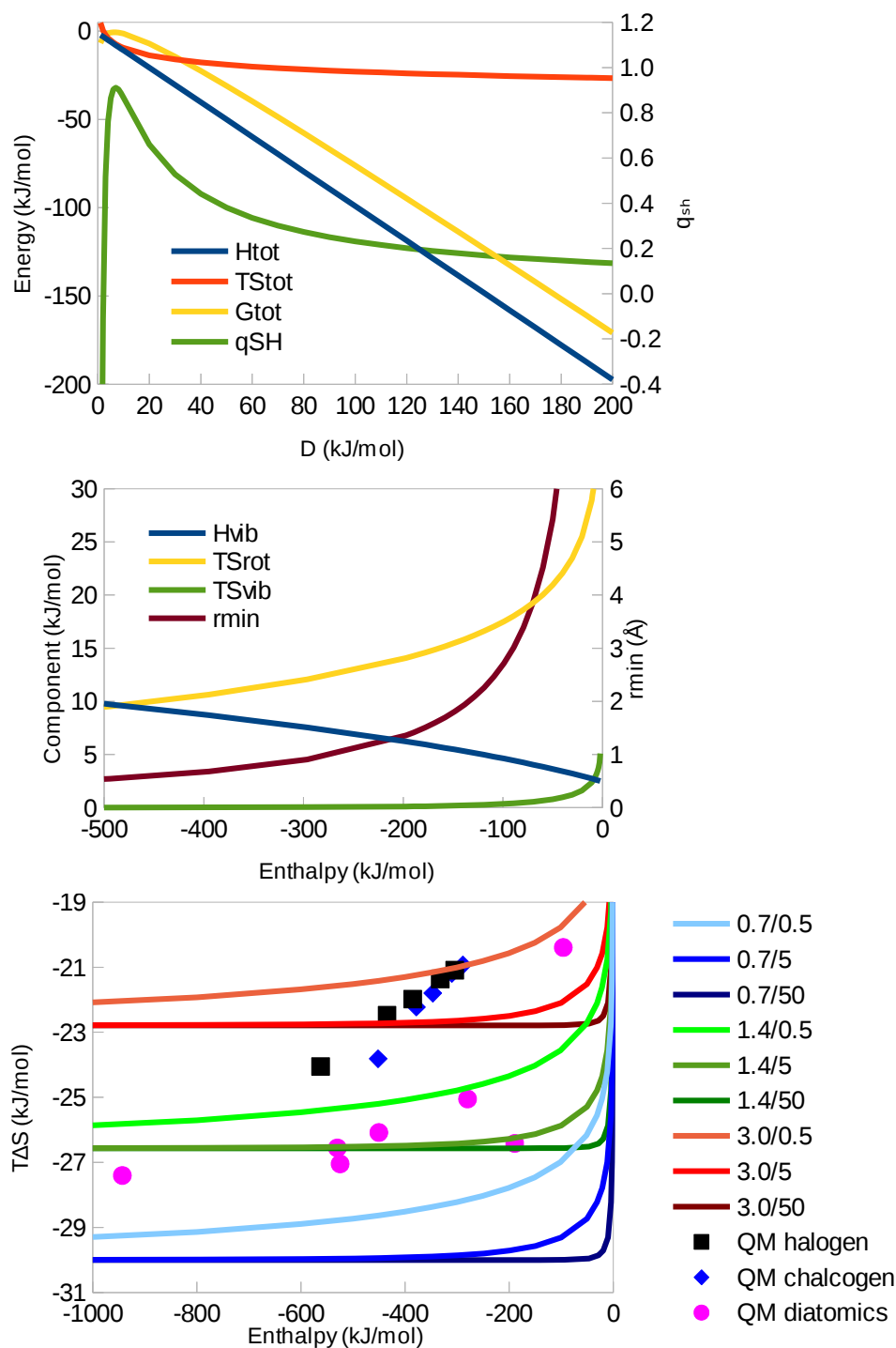


Figure 4. (a) Correlation between the enthalpy and entropy as a function of either q_H or B and with the other parameter fixed at the values given in the legend (and with $2.435 \times 10^6 \text{ \AA}^{12} \text{ kJ/mol}$, $m_H = 1.008 \text{ Da}$, and $m_X = 15.9994 \text{ Da}$). The QM and MM results for $(\text{H}_2\text{O})_2$ are also included with symbols. (b) The enthalpy and entropy components, as well as the minimum distance (r_{\min}) as a function of ΔH for the same parameters, i.e. for each property, the same four lines as in (a) are given. The four lines almost coincide for $T\Delta S_{\text{rot}}$ and r_{\min} , whereas the one for $q_H = 0.01 e$ differs for ΔH_{vib} and $T\Delta S_{\text{vib}}$.

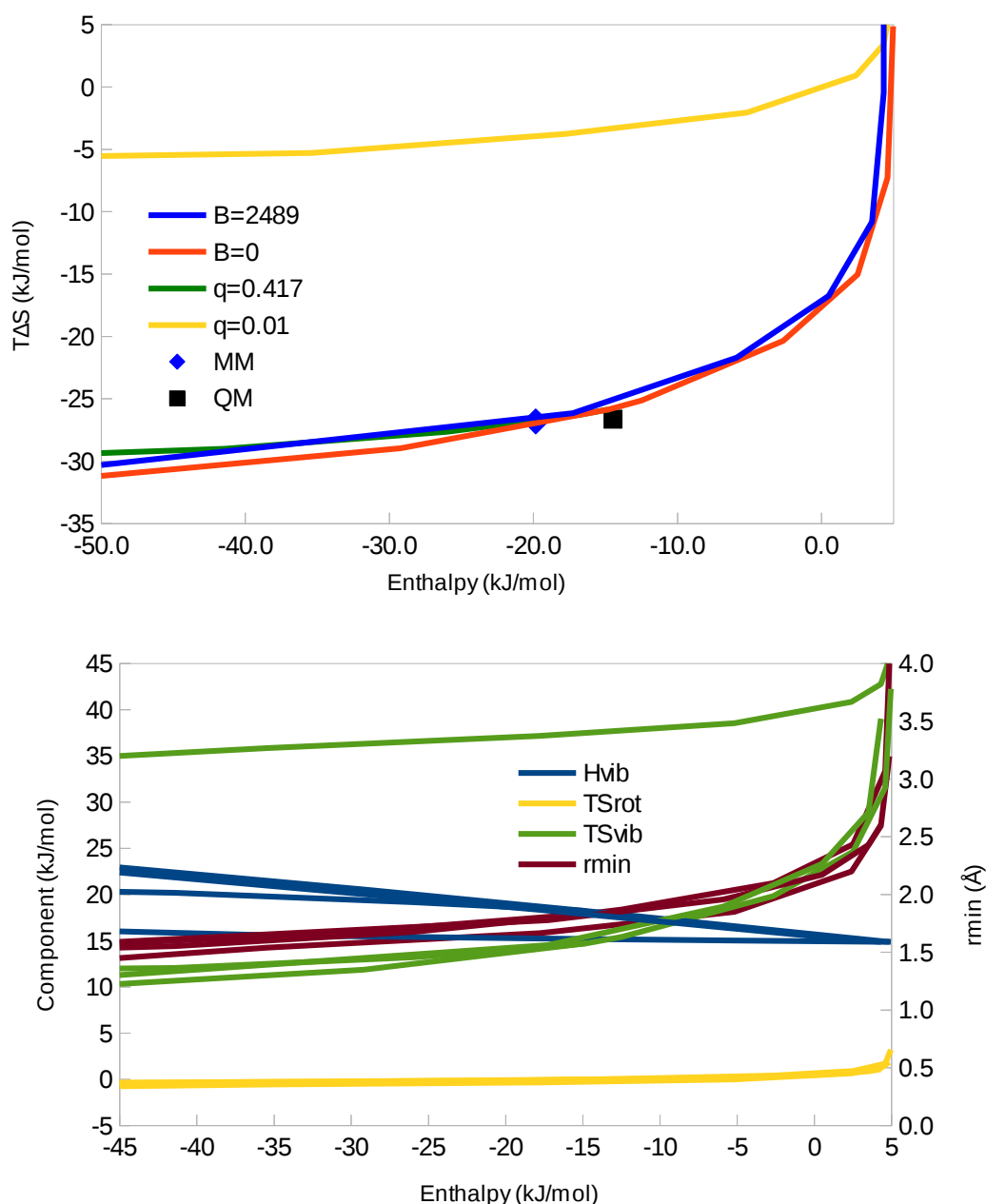


Figure 5. (a) Correlation between the enthalpy and entropy, and (b) the enthalpy and entropy components as a function of the enthalpy for homodimers of the seven smallest *n*-alkanes, and three simplest branched alkanes (isobutane, isopentane, and neopentane), as well as for eight simple polyaromatic homodimers (benzene, naphthalene, anthracene, phenanthrene, tetracene, pyrene, pentacene, and hexacene).

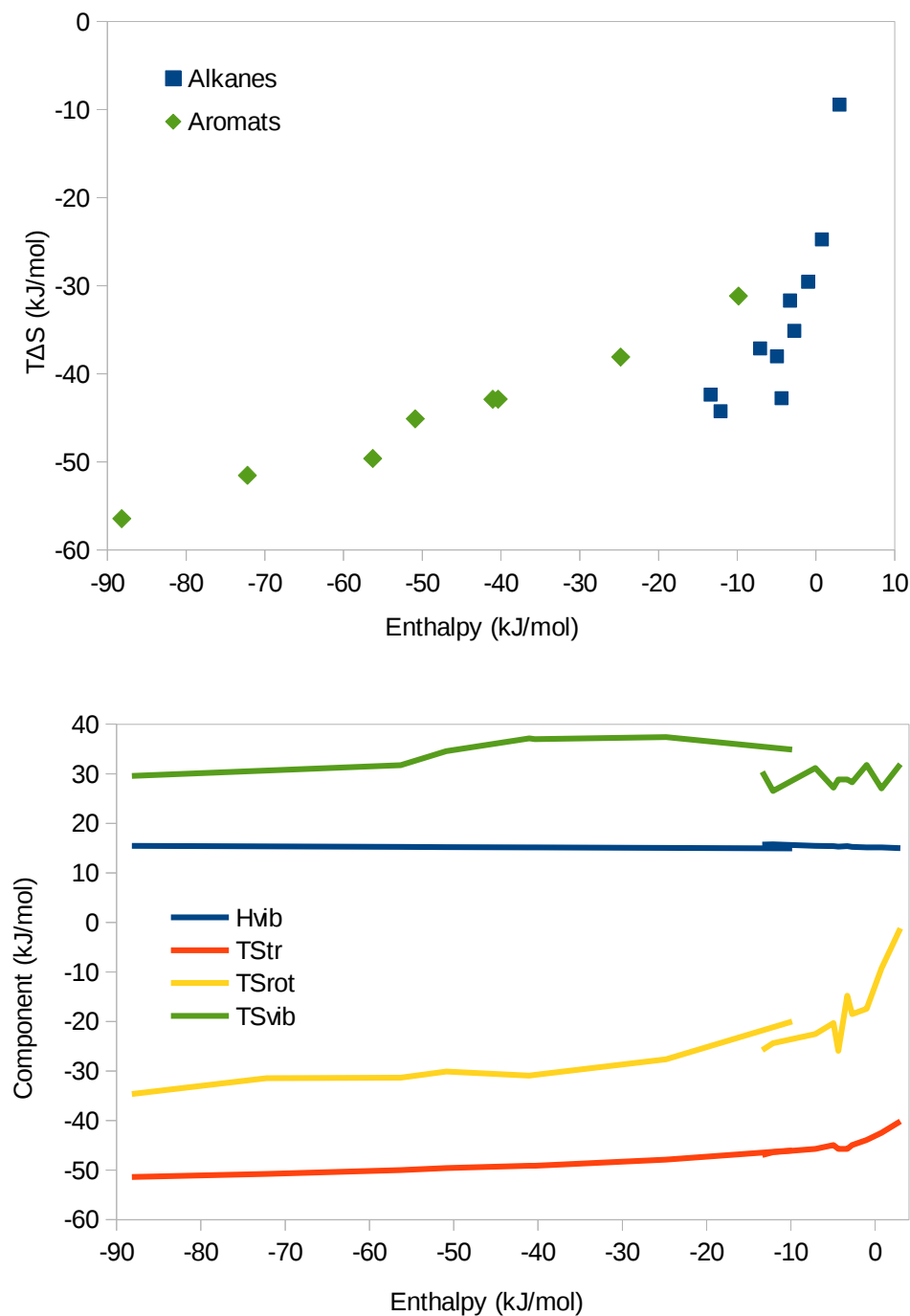


Figure 6. (a) Correlation between the enthalpy and entropy and (b) ΔH_{vib} as well as the entropy components as a function of the enthalpy for the eight dimers of glycine oligomers (with 1, 2, 4, 6, 8, 10, 12, and 14 glycine units), as well as five dimers of polypropionic acid oligomers (with 1–5 units). The insets in (a) show two typical examples.

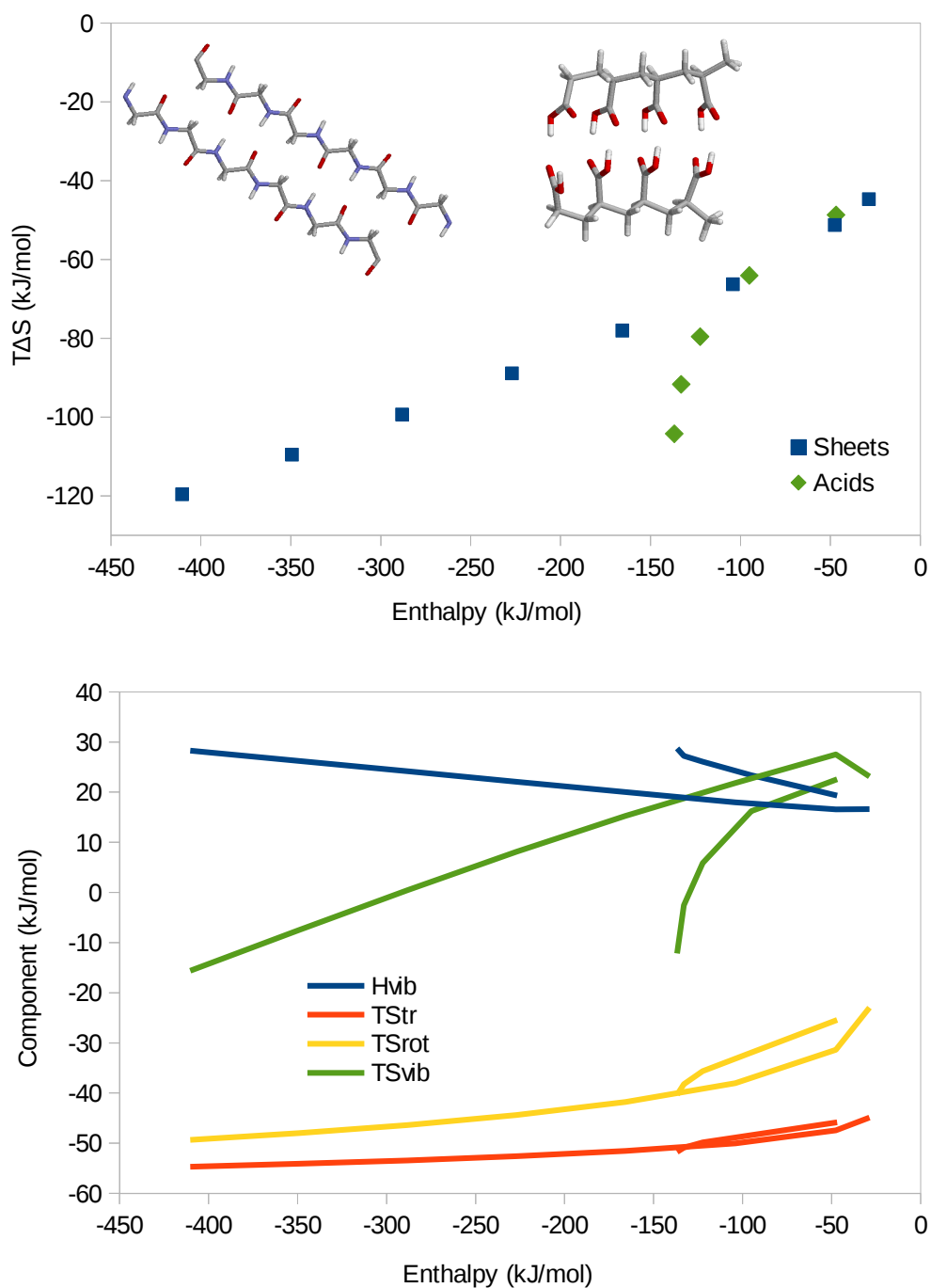


Figure 7. Correlation between the enthalpy and entropy of the ten small alkanes in Figure 5 and the six dimers of the polypropionic acid oligomers in Figure 6 without and with water.

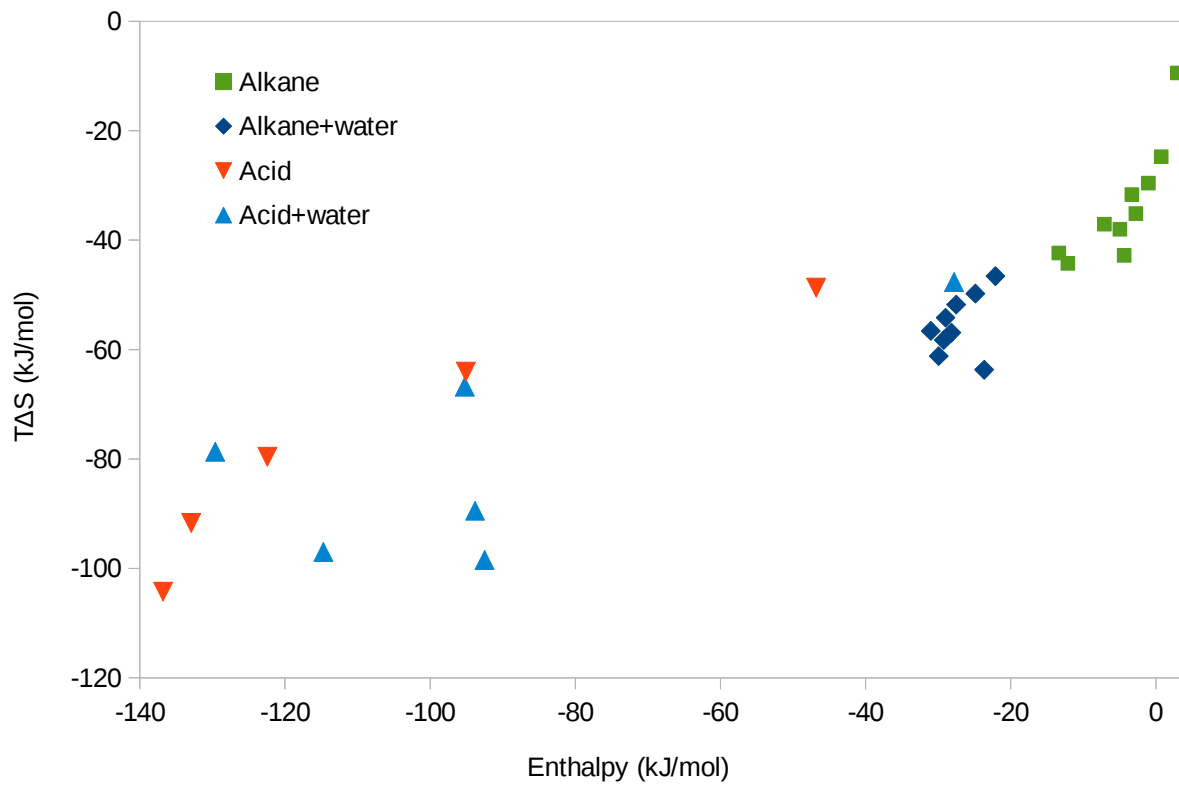


Figure 8. Scatter plots of ΔH vs. $T\Delta S$ for our 89 molecular complexes and Korth's results for the S22 and S60 sets. The complexes are grouped after the major interactions as dispersive, hydrogen bond, ionic, covalent, or mixed. The inset shows the full range of ΔH values.

