CHEMISTRY
A European Journal

www.chemeurj.org

Reprint

A Journal of
ChemPubSoc Europe
ACES
Asian Chemical Editorial Society
WILEY-VCH

ChemPubSoc Europe

DOI: 10.1002/chem.201701031

CHEMISTRY
A European Journal
Communication

Noncovalent Interactions |*Hot Paper*|

The Surprising Importance of Peptide Bond Contacts in Drug–Protein Interactions

Robert M. Parrish,[a] Doree F. Sitkoff,[b] Daniel L. Cheney,*[b] and C. David Sherrill*[a]

Abstract: The study of noncovalent interactions, notably including drug–protein binding, relies heavily on the language of localized functional group contacts: hydrogen bonding, π–π interactions, CH–π contacts, halogen bonding, etc. Applying the state-of-the-art functional group symmetry-adapted perturbation theory (F-SAPT) to an important question of chloro versus methyl aryl substitution in factor Xa inhibitor drugs, we find that a localized contact model provides an incorrect picture for the origin of the enhancement of chloro-containing ligands. Instead, the enhancement is found to originate from many intermediate-range contacts distributed throughout the binding pocket, particularly including the peptide bonds in the protein backbone. The contributions from these contacts are primarily electrostatic in nature, but require ab initio computations involving nearly the full drug–protein pocket system to be accurately quantified.

Computational drug design, despite several notable achievements,[1] remains a frontier area for theoretical chemistry, with numerous unresolved challenges. Formally, the disciplines of electronic structure theory and statistical mechanics can provide arbitrarily accurate predictions of desired observables, e.g., of the ligand–protein binding energy ΔG_{bind}. However, in practice, the computational effort required to accurately encapsulate the physics for these observables is intractable.[2] For example, to accurately compute ΔG_{bind} for a given ligand–protein system, one must implicitly or explicitly capture the contributions from the gas-phase interaction energy ΔE_{int}, solvent effects, including the desolvation penalties of the ligand and binding pocket, and the dynamic vibrational and conformational corrections.[2,3] Moreover, all of these effects should be modeled using an ab initio quality potential surface. Though much progress toward accurate characterization of ΔG_{bind} has been made along the lines of quantum chemistry[4] and free energy perturbation theory,[3,5,6] direct quantification of this metric remains elusive.

An alternative approach involves starting from a known and well-characterized ligand–protein system, and then proposing substitutions to the ligand to enhance the binding affinity. This approach is ostensibly simpler, as one only needs to predict the binding energy difference $\Delta\Delta G_{bind}$ accurately. For simple substitutions involving similar polarities, the differences in ligand desolvation penalties and dynamical contributions may be small relative to $\Delta\Delta G_{bind}$. In such cases, $\Delta\Delta G_{bind}$ will be largely governed by the difference of the interaction energy $\Delta\Delta E_{int}$ between the ligand and the protein, and it suffices to characterize the latter quantity to predict, at least qualitatively, the substituted ligand's affinity. Note that there exist many cases in which the differential desolvation penalties and dynamical contributions are not small, and require direct computation of $\Delta\Delta G_{bind}$ for an accurate result (e.g., in the case of bound ligands which are partially exposed to the solvent)—such systems lie outside the scope of the present study.

With regard to determining $\Delta\Delta E_{int}$, it would be extremely useful if a simple rulebook existed for predicting the sign and magnitude of this quantity without needing to resort to extensive ab initio computations. A system along these lines does exist, and is widely used by chemists to rationalize or even directly predict $\Delta\Delta E_{int}$: the localized contact model. The idea here is that functional group constituency and local geometric topology can identify transferable "contacts" such as hydrogen bonds, π–π interactions, CH–π contacts, halogen bonding, etc., and that $\Delta\Delta E_{int}$ may then be characterized by discussing only a handful of these contacts.

Herein we present the first application of the recently developed functional-group symmetry-adapted peturbation theory (F-SAPT)[7,8] to protein–ligand complexes. F-SAPT is capable of providing the strength of noncovalent contacts between functional groups, including their makeup in terms of electrostatics, polarization, London dispersion, and exchange–repulsion. In particular, we compare chloro- versus methyl-substituted ligands targeting factor Xa (fXa), a key enzyme in the blood coagulation cascade. Ligands of this general class have recently been introduced as effective therapeutic agents for the prevention of deep vein thrombosis and stroke[9,10]. A key to the development of these orally available ligands was the introduction of a neutral ligand moiety targeting the negatively

[a] *Dr. R. M. Parrish, Prof. C. D. Sherrill*
Center for Computational Molecular Science and Technology
School of Chemistry and Biochemistry
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0400 (USA)
E-mail: sherrill@gatech.edu

[b] *Dr. D. F. Sitkoff, Dr. D. L. Cheney*
Molecular Structure and Design,
Bristol-Myers Squibb Company,
311 Pennington-Rocky Hill Road,
Pennington, NJ 08534 (USA)
E-mail: daniell.cheney@bms.com

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/chem.201701031.

charged S1 pocket of fXa. For this consideration, an interesting experimental finding is that ligand moieties with a variety of chloro-substituted (Cl) aromatic P1 groups are often potent fXa binders, while the corresponding methyl-substituted variants (Me) are generally weaker ligands.[11–13] This question is the crux of the present work: why do the Cl-type ligands bind more strongly than their Me-type counterparts? Note that in the present study, we will confine ourselves to the form of this question that pertains only to $\Delta\Delta E_{int}$ (which we find correlates well in this particular case with experimental $\Delta\Delta G_{int}$); the focus of this work is to demonstrate the utility of F-SAPT in understanding the enthalpic component of drug binding.

Figure 1 illustrates the binding of a Cl-type ligand in the S1 pocket of fXa, as seen in the crystal structure PDB ID 3ENS. For this ligand, the salient feature is a 3-chloroindole moiety in the Cl variant and a 3-methyindole moiety in the Me variant. This single substitution changes the in vitro IC_{50} from 2.4 nM for Cl to 118 nM for Me, that is, a 49× enhancement for the Cl-type ligand.[11] Similar, though often less marked, enhancements are observed for a wide variety of other fXa inhibitors: along these lines, we also consider ligand/fXa crystal structure complexes 2CJI, 2PR3, and 2W26 (see Table 1). Potential origins of this enhancement that have previously been explored include stronger favorable interactions between the Ar–Cl group and the π-system of the Tyr228 phenol, differential dipole–charge interactions between the Ar–X group and Asp189, differential desolvation penalties of the ligands, and the formation of a stronger hydrogen bond between the indole N–H moiety and the peptide bond between Gly219 and Cys220 (e.g., refs. [11, 15, and 16]). This last hypothesis is unique to the indole-type aromatic ring in 3ENS, and thus might explain the larger enhancement for 3ENS versus the other ligands, but not the overall enhancement in Cl versus Me-type ligands.

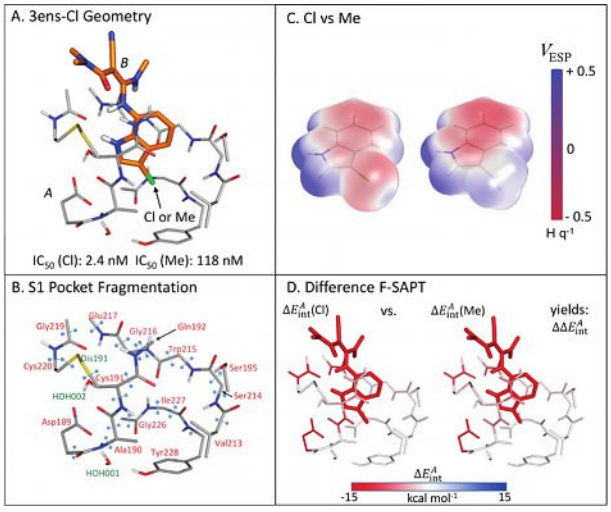
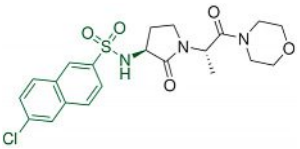
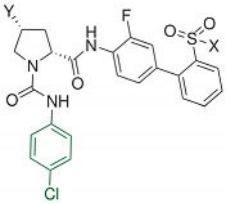
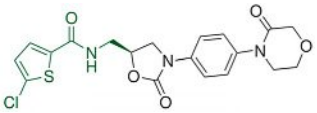
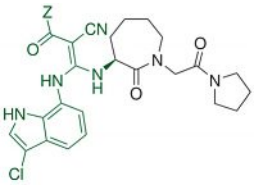


Figure 1. Overview of the difference F-SAPT computational experiment for ligand 3ENS. (A) Geometry of 3ENS-Cl ligand S1 pocket complex. (B) F-SAPT fragmentation Scheme for the S1 pocket–peptide bonds are numbered according to the lower of the two residues which they bridge. (C) 3-chloroindole vs. 3-methylindole electrostatic potential plots (Hartree–Fock/jun-cc-pVDZ, consistent with F- SAPT computations, isosurface at 0.002 electron $bohr^{-3}$). (D) F-SAPT computations (F-SAPT0/jun-cc-pVDZ) for ΔE^A_{int} for Cl and Me variants (the difference, $\Delta\Delta E^A_{int}$, is displayed in Figure 2).

Table 1. Crystallographic and experimental data for fXa/ligand complexes.

Ligand[a]	$\Delta\Delta G_{bind}$(Cl–Me) [kcal mol^{-1}][b]	PDB Code/ Resolution[b]	Ref
	−1.7	2CJI 2.1 Å	13
	−1.4 X = NH_2 Y = H Racemic mixtures	2PR3 X = Me Y = OMe 1.5 Å	15
	−1.1	2W26 2.08 Å	12
	−2.3 Z = OMe	3ENS Z = $N(Me)_2$ 2.38 Å	11

[a] 2D structures for the studied ligands, with the portion used in the SAPT computations colored in green. [b] Small differences, where they exist, are noted between the crystal ligands and the ligands for which fXa binding data were reported; computations used the latter versions.

Additionally, given that the experimental hydration free energies of toluene and chlorobenzene are so similar (−0.89 and −1.12 kcal mol^{-1}, respectively),[17] differential desolvation of Cl versus Me-type ligands is unlikely to account for the observed differences in binding affinity. Considering this, the similar size and rigidity of the small Cl and Me substitutents (Panel C of Figure 1), and the buried placement of the S1 pocket, it is plausible that the enhancement is well-described by $\Delta\Delta E_{int}$. In our computational experiment described below, we find support for this hypothesis.

The question now is what drives $\Delta\Delta E_{int}$: Cl–aryl interactions with Tyr228, differential dipole–charge interactions with Asp189, or some other effect? We can answer this directly by applying the recently developed F-SAPT approach,[7,8] which computes ΔE_{int} as a series of effective two-body interactions between pairs of functional groups. By construction, the pairwise F-SAPT contributions naturally sum to the many-body SAPT limit[19,20] of the interaction energy under the chosen level of SAPT, distinguishing F-SAPT from a number of fragmentation approaches. In the present work, we compute the SAPT0/jun-cc-pVDZ[18,21–23] interaction energy between the S1 pocket (monomer A) and the Cl-type P1 analogue (monomer B), using F-SAPT to keep track of the interaction energy contributions from the protein side chains, peptide bonds, disulfide bonds, and structural waters [yielding ΔE^A_{int}(Cl), in which *A* indicates a protein moiety]. We repeat the analysis with the

Me-type P1 analogue [yielding ΔE^A_{int}(Me)]. We then subtract the two partitions, yielding $\Delta\Delta E^A_{int}$, that is, the contributions of the various protein moieties to the difference energy. This is exactly equivalent to the application of the localized contact model, with the exception that we rely on the F-SAPT partition instead of intuition, and we quantify all interactions with quantum mechanics. A schematic of this difference F-SAPT computational experiment is depicted in Panels B and D of Figure 1, and described in more detail in the Supporting Information, in which results for the other ligands are also presented. The full system in Figure 1 is computed with F-SAPT.

A subset of the results of the difference F-SAPT analysis are presented in Figure 2. The ab initio SAPT0/jun-cc-pVDZ computations indicate that the enhancement in $\Delta\Delta G_{bind}$ is largely driven by $\Delta\Delta E_{int}$—the latter quantity always favors the Cl variant, and the magnitudes track quite well with the experimental IC_{50} or K_i enhancements across the four ligands tested [e.g., an enhancement of 2.5 (calculated) versus 2.3 (experimental) kcal mol^{-1} for 3ENS].[25] Having established this, we move on to determining the origins of $\Delta\Delta E_{int}$ via F-SAPT. The immediate finding is that the hypotheses of Cl-aryl and differential dipole–charge interactions are not supported—the $\Delta\Delta E^A_{int}$ contributions from Tyr228 and Asp189 are both less than 0.25 kcal mol^{-1} for 3ENS. This negligible contribution holds up across the other three ligands, except that the Asp189 contribution is unfavorable by :1 kcal mol^{-1} in 2CJI. The deeper finding is that the overall difference in Cl versus Me binding is driven by a large number of favorable and unfavorable contributions from all over the S1 pocket which together conspire to provide enhanced binding in Cl. As seen in Panels A and B of Figure 2, the largest contributors are the backbone peptide bonds: for 3ENS, these provide the leading seven contributions by magnitude to $\Delta\Delta E_{int}$. On further reflection, this finding is certainly plausible considering the effect of Cl versus Me on the dipole of the aromatic group, and therefore, the differential dipole-dipole interactions between the ligand and CONH peptide bond moieties (the potential importance of protein amides lining the fXa S1 pocket to binding ligands with aromatic P1 groups is discussed, though not in the context of Cl substitutents, in ref.[25]). Additional analysis in Panel C of Figure 2 shows that the enhancement is driven primarily by the electrostatics term, but that even the optimally sorted partial summations to $\Delta\Delta E_{elst}$ are slowly convergent. While the other four terms are generally converged to 0.5 kcal mol^{-1} within the leading five to ten contacts, $\Delta\Delta E_{elst}$ requires fifteen to twenty contacts to converge. In other words, the Cl versus Me-based ligands have a large number of significant differential electrostatic interactions with the binding pocket fragments. Even the apparent convergence at fifteen to twenty contacts may be an illusion: it is certainly possible that additional quantitative contributors to $\Delta\Delta E_{elst}$ might appear if the next shell of side chains and protein backbone was added to the system (this is currently not computationally feasible—even density-fitted F-SAPT is presently limited to :200–300 atoms), and if the (more distant, but still potentially significant) solvent effects were included. Note that the hypothesis of enhanced hydrogen bonding in 3ENS is borne out: the total enhancement in 3ENS is 1.2 kcal mol^{-1} more favorable than the next leading ligand (2CJI, Supporting Information Table II), and the PEP219 contact (see Figure 1 caption for definition) is the strongest enhancing contributor in 3ENS but not in the other three ligands. However, this effect accounts for, at most, half of the enhancement in 3ENS, and none at all for the other three ligands. In fact, this contact is not even the strongest contributor to $\Delta\Delta E_{elst}$ in 3ENS: an unfavorable contact with PEP215 penalizes the Cl-type ligand by nearly 3 kcal mol^{-1}. The total enhancement is only favorable after summing over a large number of other contacts, in which PEP219 is only one of several important favorable contributors.

A follow-up question is, "what is the minimal system that can capture the true enhancement effect?" On the one hand, it would be computationally beneficial if the enhancement could be elucidated by performing isolated computations between the ligand and various fragments of the protein pocket. We have performed a "cut-and-cap" fragmentation analysis along these lines in the supplemental material, using SAPT0/jun-cc-pVDZ between the ligand-protein fragment pairs. We find that cut-and-cap analysis verifies the qualitative conclusion of many significant contacts throughout the S1 pocket (especially including peptide bonds), but cannot accurately

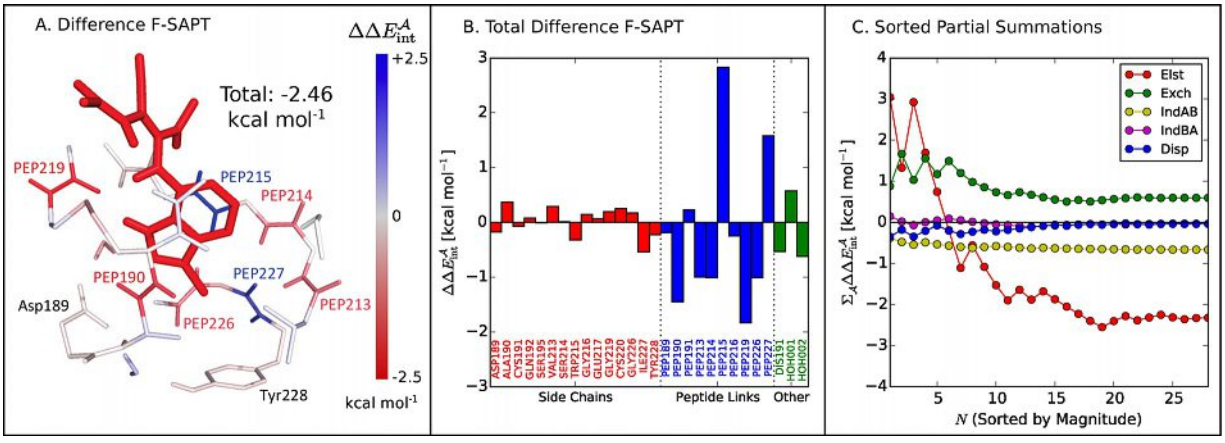


Figure 2. Difference F-SAPT results for ligand 3ENS (F-SAPT0/jun-cc-pVDZ). (A) $\Delta\Delta E^A_{int}$ mapped to the 3ENS-Cl ligand S1 pocket complex geometry. (B) Quantitative $\Delta\Delta E^A_{int}$ data. (C) $\Delta\Delta E^A_{int}$ partial summations sorted by the magnitude of the underlying contributions.

reproduce the quantitative difference energies. The electrostatic contributions to the difference energies agree well with SAPT on the whole system, but the other components generally do not, producing total $\Delta\Delta E_{int}$ which are : 1 kcal mol^{-1} less favorable than are found in the full system (e.g., to the point of predicting Me to be more favorable in 2W26). By contrast, we also consider working with the full S1 pocket but reduced models of the ligand. Removing the "tail" of the ligand and working with only a 3-chloroindole/3-methylindole model for 3ENS is qualitatively acceptable–this yields 2.0 kcal mol^{-1} of the 2.5 kcal mol^{-1} total enhancement. However, truncating more aggressively to a "local substituent" model of HCl or methane is not acceptable—this yields only 0.12 kcal mol^{-1} of the 2.5 kcal mol^{-1} total enhancement. Thus it appears that the minimal physical system for this problem is a large portion of the S1 pocket, treated as a single ab initio monomer, together with the substituted aromatic ring moiety of the ligand.

This study leads to a number of interesting conclusions regarding the prediction of how chemical substitutions enhance drug binding strengths.

1. The localized contact model fails even for the exceedingly simple $\Delta\Delta E_{int}$-dominated Me→Cl substitution considered here. Many contacts may contribute significantly, particularly in the electrostatics term, and especially involving the protein backbone.

2. Reliable modeling of $\Delta\Delta E_{int}$ requires a very large portion of the ligand-protein system, notably including the complete S1 binding pocket and a substantial piece of the substituted ligand.

3. Tools such as SAPT and F-SAPT can provide insight into these enhancements without relying on chemical intuition. SAPT can now be routinely deployed to compute $\Delta\Delta E_{int}$ in systems of this size, and F-SAPT provides a robust partitioning of this many-body $\Delta\Delta E_{int}$ to protein or ligand fragments.

These findings indicate that an intuitive understanding of noncovalent interactions based on a visual identification of various interaction motifs (hydrogen bonds, etc.) may not always be sufficient to understand the differences between two chemically similar drugs. Instead, changes in the interactions with many protein fragments (some of them intermediate-ranged) may be significant. The major recommendation from this study is that one should generally perform a benchmark computation for $\Delta\Delta E_{int}$ using a large-scale model for the ligand and protein binding pocket: tools such as SAPT and F-SAPT now offer the tractability to make this fairly straightforward. Substantial additional efforts will be required to handle cases in which contributions to $\Delta\Delta G_{bind}$ that are outside of $\Delta\Delta E_{int}$ become important.

Acknowledgements

R.M.P. was supported by a US DOE Computational Science Graduate Fellowship (Grant No. DE-FG02-97ER25308). This work is supported by the US National Science Foundation (Grant No. CHE-1566192).

Conflict of interest

The authors declare no conflict of interest.

Keywords: ab initio calculations · drug design · noncovalent interactions · peptides

[1] W. L. Jorgensen, *Acc. Chem. Res.* **2009**, *42*, 724 – 733.
[2] K. M. Merz, *J. Chem. Theory Comput.* **2010**, *6*, 1769 – 1776.
[3] M. K. Gilson, H. Zhou, *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21 – 42.
[4] J. Antony, S. Grimme, *J. Comput. Chem.* **2012**, *33*, 1730 – 1739.
[5] J. L. Knight, C. L. Brooks, *J. Comput. Chem.* **2009**, *30*, 1692 – 1700.
[6] Y. Deng, B. Roux, *J. Phys. Chem. B* **2009**, *113*, 2234 – 2246.
[7] R. M. Parrish, C. D. Sherrill, *J. Chem. Phys.* **2014**, *141*, 044115.
[8] R. M. Parrish, T. M. Parker, C. D. Sherrill, *J. Chem. Theory Comput.* **2014**, *10*, 4417 – 4431.
[9] D. J. P. Pinto, J. M. Smallheer, D. L. Cheney, R. M. Knabb, and R. R. Wexler, *J. Med. Chem.* **2010**, *53*, 6243 – 6274.
[10] N. R. Patel, D. V. Patel, P. R. Murumkar, and M. R. Yadav, *Eur. J. Med. Chem.* **2016**, *121*, 671 – 698.
[11] Y. Shi, D. Sitkoff, J. Zhang, H. E. Klei, K. Kish, E. C-K. Liu, K. S. Hartl, S. M. Seiler, M. Chang, C. Huang, S. Youssef, T. E. Steinbacher, W. A. Schumacher, N. Grazier, A. Pudzianowski, A. Apedo, L. Discenza, J. Yanchunas, P. D. Stein, K. S. Atwal, *J. Med. Chem.* **2008**, *51*, 7541 – 7551.
[12] S. Roehrig, A. Straub, J. Pohlmann, T. Lampe, J. Pernerstorfer, K. H. Schlemmer, P. Reinemer, E. Perzborn, *J. Med. Chem.* **2005**, *48*, 5900 – 5908.
[13] C. Chan, A. D. Borthwick, D. Brown, C. L. Burns-Kurtis, M. Campbell, L. Chaudry, C. Chung, M. A. Convery, J. N. Hamblin, L. Johnstone, H. A. Kelly, S. Kleanthous, A. Patikis, C. Patel, A. J. Pateman, S. Senger, G. P. Shah, J. R. Toomey, N. S. Watson, H. E. Weston, C. Whitworth, R. J. Young, P. Zhou, *J. Med. Chem.* **2007**, *50*, 1546 – 1557.
[14] C. A. Van Huis, C. F. Bigge, A. Casimiro-Garcia, W. L. Cody, D. A. Dudley, K. J. Filipski, R. J. Heemstra, J. T. Kohrt, L. S. Narasimhan, R. P. Schaum, E. Zhang, J. W. Bryant, S. Haarer, N. Janiczek, R. J. Leadley, T. McClanahan, J. T. Peterson, K. M. Welch, J. J. Edmunds, *Chem. Biol. Drug Des.* **2007**, *69*, 444 – 450.
[15] H. Matter, M. Nazare, S. Guessregen, D. W. Will, H. Schreuder, A. Bauer, M. Urmann, K. Ritter, M. Wagner, V. Wehner, *Angew. Chem. Int. Ed.* **2009**, *48*, 2911 – 2916; *Angew. Chem.* **2009**, *121*, 2955 – 2960.
[16] H. G. Wallnoefer, T. Fox, K. R. Liedl, C. S. Tautermann, *Phys. Chem. Chem. Phys.* **2010**, *12*, 14941 – 14949.
[17] D. J. Giesen, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem.* **1995**, *99*, 7137 – 7146.
[18] B. Jeziorski, R. Moszynski, A. Ratkiewicz, S. Rybak, K. Szalewicz, H. L. Williams, In *Methods and Techniques in Computational Chemistry: METECC94*; Clementi, E., Ed.; STEF: Cagliari, **1993**, Vol. B (Medium-Size Systems); p. 79.
[19] B. Jeziorski, R. Moszynski, K. Szalewicz, *Chem. Rev.* **1994**, *94*, 1887 – 1930.
[20] K. Szalewicz, *WIREs Comput. Mol. Sci.* **2012**, *2*, 254 – 272.
[21] E. G. Hohenstein, C. D. Sherrill, *J. Chem. Phys.* **2010**, *132*, 184111.
[22] E. G. Hohenstein, R. M. Parrish, C. D. Sherrill, J. M. Turney, H. F. Schaefer, *J. Chem. Phys.* **2011**, *135*, 174107.
[23] T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno, C. D. Sherrill, *J. Chem. Phys.* **2014**, *140*, 094106.
[24] Experimental binding energy difference computed via $\Delta\Delta G^{3ENS}_{Exp} = k_B T \ln\left(IC^{Cl}_{50}/IC^{Me}_{50}\right) = 0.593 \ln\left(\frac{2.4}{118}\right) = -2.3$ kcal mol^{-1}.
[25] L. M. Salonen, M. C. Holland, P. S. J. Kaib, W. Haap, J. Benz, J. Mary, O. Kuster, W. B. Schweizer, D. W. Banner, F. Diederich, *Chem. Eur. J.* **2012**, *18*, 213 – 222.

Manuscript received: March 6, 2017
Accepted manuscript online: April 4, 2017
Version of record online: May 23, 2017