Unraveling individual host–guest interactions in molecular recognition from first principles quantum mechanics: Insights into the nature of nicotinic acetylcholine receptor agonist binding

Michael Edmund Beck1 | Christoph Riplinger2 | Frank Neese3 | Giovanni Bistoni3

1Department Computational Life Science, Bayer AG, Division Cropscience, Monheim am Rhein, Germany
2FAccTs GmbH, Köln, Germany
3Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr, Germany

Correspondence
Michael Edmund Beck, Bayer AG, Division Cropscience, Department of Computational Life Science, Alfred-Nobel-Str. 50, Monheim am Rhein 40789, Germany.
Email: michael.beck@bayer.com

Christoph Riplinger, FAccTs GmbH, Köln, Germany.
Email: riplinger@faccts.de

Frank Neese and Giovanni Bistoni, Max-Planck-Institut für Kohlenforschung, Kaiser Wilhelm-Platz 1, Mülheim an der Ruhr 45470, Germany.
Email: frank.neese@kofo.mpg.de (F. N.) and Email: giovanni.bistoni@kofo.mpg.de (G. B.)

Funding information
Deutsche Forschungsgemeinschaft; Max-Planck-Gesellschaft

Abstract
Drug binding to a protein target is governed by a complex pattern of noncovalent interactions between the ligand and the residues in the protein's binding pocket. Here we introduce a generally applicable, parameter-free, computational method that allows for the identification, quantification, and analysis of the key ligand–residue interactions responsible for molecular recognition. Our strategy relies on Local Energy Decomposition analysis at the "gold-standard" coupled cluster DLPNO-CCSD(T) level. In the study case shown in this paper, nicotine and imidacloprid binding to the nicotinic acetylcholine receptor, our approach provides new insights into how individual amino acids in the active site determine sensitivity and selectivity of the ligands, extending and refining classical pharmacophore hypotheses. By inference, the method is applicable to any kind of host/guest interactions with potential applications in industrial biocatalysis and protein engineering.

KEYWORDS
drug design, intermolecular interactions, local coupled cluster, local energy decomposition, molecular recognition, protein–ligand interactions

1 | INTRODUCTION

Understanding the nature of protein–ligand interactions underlies any attempt to design safe and efficient small molecule drugs as well as agrochemicals. Due to the advances in methodology as well as computing power, computer aided drug discovery has developed from a niche to rather standard in academic as well as industrial life science research. While in former days discovery and optimization of small molecules were focused in the first place toward efficacy against a single target only, since a few years the paradigm is shifting toward a multi-parameter approach, in which adverse effects, for example, affinity to off-targets, species selectivity, resistance development, or environmental effects, are taken into account from very early on.1,2

This trend is further adding on the need for generally applicable methods that lead to a deep understanding of how small molecule
drugs and agrochemicals bind to their biological target, for example, proteins and RNAs. Nowadays, a large number of computational strategies are routinely applied to the study of protein–ligand interactions. For example, free energy perturbation methods have proven instrumental for predicting free energy differences of binding for large numbers of ligands, while machine learning and artificial intelligence techniques have shown their potential to allow highly promising ways to navigate vast chemical biology spaces. In addition, many quantum mechanical (QM) methods can be used to obtain an insight into the nature of protein–ligand interactions, either via calculations on large cluster models or on entire protein–ligand complexes within a hybrid quantum mechanical/molecular mechanical (QM/MM) approach.

QM methods also provide valuable information into the intricate pattern of noncovalent interactions involved in the binding of a ligand to a protein. For example, they give access to molecular properties like electron densities, whose topology can be analyzed in the framework of the Quantum Theory of Atoms in Molecules (QTAIM) to identify the most significant ligand–residue interactions. Unfortunately, although interpretative tools like QTAIM have proven useful in many (bio)chemical applications, they only provide qualitative insights, which significantly limits their usefulness in the rational design and optimization of new drugs. Alternatively, fragmentation methods, in which the energy of a large system is approximated as the sum of contributions from smaller fragments, can be used to obtain more quantitative information, although their accuracy is limited when many-body effects become important.

In this work, we introduce a computational strategy able to quantify accurately the individual ligand–residue and residue–residue interactions that contribute to drug binding. Our approach relies on a local variant of the coupled cluster method with singles, doubles, and perturbatively included triples excitations, CCSD(T), which is generally regarded as the “gold standard” of computational chemistry for its well documented high accuracy and reliability. In particular, we employ the linear-scaling domain-based local pair natural orbital CCSD(T) method, DLPNO-CCSD(T), which allows for the calculation of accurate energies for systems with hundreds of atoms and thousands of basis functions. Importantly, the DLPNO-CCSD(T) energy can be decomposed into fragment-pairwise contributions using the recently introduced Local Energy Decomposition (LED) scheme, which has already found widespread application in the analysis of intermolecular interactions throughout chemistry.

Herein, we demonstrate that the DLPNO-CCSD(T)/LED methodology can be used to exactly dissecting ligand–protein binding energies into ligand–residue and residue–residue contributions, thus providing a quantitative and (bio)chemically readily interpretable framework in which one can discuss ligand-target sensitivity and selectivity. As a prototype case study, this approach is applied to the binding modes of two, biologically very different agonists of the nicotinic acetylcholine receptor (nAChR) and homology models used in the present work. (a) Top and side-view of the structure of a full, heteropentameric nAChR, exemplified by the first such resolved structure of Torpedo marmorata. (b) Homopentameric acetylcholin binding proteins (AChBPs), like the one from Lymnea stagnalis shown here, resemble the structure of the ligand binding domain of nAChRs. (c) Architecture of homology models of insect binding sites, derived from Lymnea s. AChBP. (d) Detailed view of these homology models highlighting an arginine residue from loop D in wild type insects (top, with green carbon atoms) and an important resistance inducing mutation featuring a Threonine residue replacing arginine (bottom, with yellow carbon atoms).
nicotinic Acetylcholine Receptor (nAChR) of insects and mammals, that is, (S)-nicotine and imidacloprid.

## 2 | COMPUTATIONAL ASPECTS

### 2.1 | Cluster models

nAChRs are ligand-gated ion channels that are involved in the rapid neuronal signal transduction in animal nervous systems. The overall architecture is sketched in Figure 1(A); for the scope of this study it is sufficient to realize that nAChRs are homo- or heteromeric pentamers of structurally related subunits that comprise an extracellular N-terminal domain involved in ligand binding, four transmembrane domains forming the actual cation-channel, and an intracellular region.\(^{23}\)

While the exact composition and 3D structure of nAChRs is very well known for many vertebrates, functional architecture, diversity, and three-dimensional (3D) tertiary structure of native insect nAChRs are still poorly understood.\(^{24}\) Homopentameric acetylcholine binding proteins (AChBPs) found in snails (e.g., marine Lymnea stagnalis) resemble nAChRs’ ligand binding domain, Figure 1(B).\(^{25}\) Based on available AChB 3D-structures, numerous homology models of the binding sites in insect nAChRs were constructed\(^{26–32}\) as sketched in Figure 1(C). The calculations in this work are based on an already existing model\(^{24,33}\) of the active site of *Myzus persicae* nAChR. *M. persicae* is a useful model, as the active site of nAChR is highly conserved across insect species.\(^{34}\)

Nicotine and imidacloprid investigated in this work represent the “nicotinoid” and “neo-nicotinoid” classes of nAChR agonists.\(^{36}\) They exhibit reverse patterns of selectivity. (S)-(-)-nicotine is a natural product with only very limited efficacy as an insecticide, reflected by an IC\(_{50}\) ~4000 nM at insect nAChR, but shows high mammalian toxicity with an IC\(_{50}\) of 7 nM.\(^{37}\) The synthetic insecticide imidacloprid, selectively agonizes insect nAChR (IC\(_{50}\) 4.6 nM), binding weaker to vertebrate nAChR by a factor of more than 500.\(^{37}\) The observed

---

**FIGURE 2** Overlay of starting geometries, as derived from the homology models (pink wireframe), and constrained geometry optimizations at the HF-3c level of theory. In the optimized geometries, the ligands imidacloprid (panels a and b) and nicotine (panels c and d) are shown as ball and stick models, with cyan carbon atoms. Panels (a) and (c) show the models of the native insect nAChR binding pocket, featuring an arginine residue with green atoms. In vertebrate as well as resistant insects, this position is occupied by a Threonin, shown with orange carbon atoms in panels (b) and (d). RMSD values comparing starting geometry and relaxed geometry are 0.99, 0.68, 0.75, and 0.66 Å in panels (a), (b), (c), and (d), respectively. The nAChR-ligand models shown in panels (a), (b), (c), and (d) contain 233, 223, 233, and 223 atoms, respectively.
selectivity pattern can be attributed to a single mutation in the binding niche of the respective nAChRs of sensitive insects as well as vertebrates and mammals, that is, the exchange of an arginine-residue by a threonine. Notably, a well-known mutation, commonly referred to as the R81T-mutation, in which a basic aminoacid (that is protonated under physiological condition) from loop D is replaced by an uncharged residue, is found in insects developing target site resistance against neo-nicotinoid insecticides like imidacloprid.28,29,34,38,39

As the numbering of aminoacids is different in our model, we will refer to this mutation as "R55T" in the following, to avoid confusion. Figure 1(D) highlights this mutation in respective homology models of nAChRs in imidacloprid sensitive insects (showing arginine) versus the model of the active site in resistant insects with a threonine residue, similar to the situation in mammals. In particular, we used the two models from refs. [24,33], one representing the binding niche of nAChR as found in neonicotinoid sensitive insects and one introducing the above mentioned R55T mutation, representing the situation found in vertebrates as well as resistant insect mutants. The starting geometries for the protein–ligand adducts, as derived from homology modeling and docking, were optimized by means of constrained geometry optimizations at the HF-3c level of theory (see Figure 2).

During geometry optimization, selected atom positions of the protein model were fixed in order to mimic the steric influence of the protein environment. These selected atoms are the Cα atoms of each protein residue as well as the neighboring hydrogen atoms which are included to saturate the dangling bonds after deleting the protein environment. The atomic positions of the relaxed model systems at the HF-3c level of theory are close to the starting models. The overall RMSDs are 0.99 and 0.68 Å for imidacloprid in the binding pocket model of insect and vertebrate nAChR, respectively, and 0.75 and 0.68 Å for the binding of nicotine. It is worth mentioning that the deviations are marginal for the involved aminoacids. The largest deviation is associated with a "twist" of the pyridine ring’s position, which however does not compromise the hydrogen bond between the pyridinic N-atom of the ligands with the conserved water molecule.

The "empty" protein pockets as well as the isolated ligands were also optimized at the HF-3c level of theory. For the protein, geometry optimizations were carried out using the same constraints used for the protein–ligand adduct.

2.2 | Binding energy calculations and LED analysis

The LED scheme provides an exact decomposition of the DLPNO-CCSD(T) energy into a series of additive contributions corresponding to the interaction of pairs of user-defined fragments. In this work, the LED scheme is used to identify the key noncovalent interactions that contribute to the binding of a ligand L (nicotine and imidacloprid in the present example) to a protein P (insect nAChR and R55T mutated nAChR).

Within a supramolecular approach, the P-L binding energy (ΔEbind) can be computed at the DLPNO-CCSD(T) level as the difference between the energy of the P-L adduct (E_{P}^{L}) and that of the isolated L (E_{L}^{tot}) and P (E_{P}^{tot}) fragments frozen in their in-adduct geometry:

\[ \Delta E_{\text{bind}} = E_{P}^{L} - E_{L}^{tot} - E_{P}^{tot} \]  \hspace{1cm} (1)

In the present case, in order to illustrate our LED approach for the analysis of ΔEbind, we use the cluster models introduced in Section 2.1 (see Section 2.2 for deformation energy, solvation and free energies corrections).

The fragments used in the LED decomposition are the ligand and the key residues in the active site (see Tables S4-S7 for the corresponding cartesian coordinates). E_{P}^{tot} can thus be decomposed into a series of intrafragment contributions, corresponding to the energy of the ligand E_{L}^{tot} and of the residues E_{R}^{tot} when their electronic structure is perturbed due to ligand binding, plus a series of ligand–residue E_{LR}^{tot} and residue–residue E_{RR}^{tot} interaction terms:

\[ E_{P}^{tot} = E_{L}^{tot} + \sum_{R} E_{R}^{tot} + \sum_{LR} E_{LR}^{tot} + \sum_{RR} E_{RR}^{tot} \]  \hspace{1cm} (2)

Similarly, E_{L}^{tot} can be decomposed into intraresidue E_{RL}^{tot} and residue–residue E_{RR}^{tot} contributions:

\[ E_{L}^{tot} = \sum_{R} E_{RL}^{tot} + \sum_{RR} E_{RR}^{tot} \]  \hspace{1cm} (3)

Using Equations (2) and (3), ΔEbind can be expressed as:

\[ \Delta E_{\text{bind}} = \left( E_{L}^{tot} - E_{L}^{tot} \right) + \sum_{R} \left[ E_{R}^{tot} - E_{R}^{tot} \right] + \sum_{LR} \left[ E_{LR}^{tot} - E_{LR}^{tot} \right] + \sum_{RR} E_{RR}^{tot} \]  \hspace{1cm} (4)

Which can be rearranged in a more compact form as:

\[ \Delta E_{\text{bind}} = \Delta E_{L}^{tot} + \sum_{R} \Delta E_{R}^{tot} + \sum_{LR} \Delta E_{LR}^{tot} + \sum_{RR} E_{RR}^{tot} \]  \hspace{1cm} (5)

Where \( \Delta E_{L}^{tot} \) and \( \Delta E_{R}^{tot} \) represent the change in the energy of the ligand and of the residues upon ligand–protein binding; \( \Delta E_{RR}^{tot} \) represents the change in the residue–residue interaction upon ligand binding and \( E_{LR}^{tot} \) denotes the interaction of the ligand with the residues in the active site.

Single point DLPNO-CCSD(T) calculations on the model structures described in Section 2.1 were carried out using the def2-TZVPP basis set with matching def2/C auxiliary basis set. NormalPNO settings were used by setting the TCutPairs threshold to the "TightPNO" value, that is, 10−5. The RJCOSX approximation was used in the HF part (the def2/JK basis set was used) with a fine integration grid (gridx6).

For the sake of comparison, binding energies were also computed with the same basis set at various levels of theory, including Hartree–Fock (HF), second-order Møller–Plesset perturbation theory using the resolution of identity approximation (RI-MP2), HF plus London
obtained are compared with those found in previously published computational and experimental works (Section 3.1). Then, various electronic structure methods are used to compute the binding energy between imidacloprid and sensitive and resistant nAChRs and the results are compared with those found at the DLPNO-CCSD(T) level (Section 3.2). Finally, the DLPNO-CCSD(T)/LED approach is used to study the protein–ligand interactions for all the systems studied in the present work (Section 3.3).

2.3 Solvation, entropy and deformation energy

Thermal corrections, entropy and environmental effects also contribute to the affinity of a ligand toward a protein. Moreover, the binding process of the ligand to the receptor might involve a significant structural rearrangement of the protein that also contributes to the binding free energy (the so called “deformation free energy”). The accurate modeling of all these effects has never been attempted before for the systems considered here. Quantitatively modeling these important but very complicated contributions is beyond the scope of the present work. Herein, our aim is to introduce a method able to quantify the key ligand–residues interactions at the DLPNO-CCSD(T) level (see Section 2.2), and to discuss how these results correlate with available experimental data for the case study considered. For the sake of completeness, estimates for the solvation- \( \Delta G_{\text{solv}} \) and deformation energy \( \Delta E_{\text{def}} \) as well as the thermal and entropy \( \Delta G_{\text{corr}} \) corrections were computed and added to \( \Delta E_{\text{bind}} \) in order to provide an approximate estimate for the overall free binding energy in solution:

\[
\Delta E_{\text{bind}} = \Delta E_{\text{bind}}^0 + \Delta E_{\text{def}}^0 + \Delta G_{\text{corr}} + \Delta G_{\text{solv}}
\]

\( \Delta E_{\text{def}}^0 \) is the energy required to distort the protein and the ligand from their equilibrium geometry to that they have in the adduct. It was computed at the same level of theory used for the calculation of \( \Delta E_{\text{bind}} \) (see Section 2.1 and 2.2). \( \Delta G_{\text{corr}} \) incorporates all thermal and entropy corrections at 298 K and was computed at the HF-3c level of theory using the rigid-rotor harmonic oscillator (RRHO) approximation. The Partial Hessian Vibrational Analysis (PHVA) approach was used, as implemented in ORCA. The atoms that were constrained in the geometry optimization (see Section 2.1) were fixed in the PHVA. Harmonic frequencies were computed for the ligand as well as for the protein model relaxed with and without the ligand bound. No imaginary frequencies were observed. Finally, \( \Delta G_{\text{solv}} \) was computed at the B3LYP level of theory using the continuum solvation model C-PCM. The dielectric constant \( \epsilon \) was set to 4, which is usually considered to be a good representation of the dielectric properties of the protein interior. Note that the trend observed for \( \Delta G_{\text{bind}} \) is the same irrespective of the specific implicit solvation method used, as detailed in Supporting Information (see Table S8).

3 RESULTS AND DISCUSSION

In this section, we start by discussing the most important structural features of the homology models just described and the results obtained are compared with those found in previously published works. Herein, our aim is to introduce a method able to quantify the key ligand–residues interactions at the DLPNO-CCSD(T) level (see Section 2.2), and to discuss how these results correlate with available experimental data for the case study considered. For the sake of completeness, estimates for the solvation- \( \Delta G_{\text{solv}} \) and deformation energy \( \Delta E_{\text{def}} \) as well as the thermal and entropy \( \Delta G_{\text{corr}} \) corrections were computed and added to \( \Delta E_{\text{bind}} \) in order to provide an approximate estimate for the overall free binding energy in solution:

\[
\Delta E_{\text{bind}} = \Delta E_{\text{bind}}^0 + \Delta E_{\text{def}}^0 + \Delta G_{\text{corr}} + \Delta G_{\text{solv}}
\]

\( \Delta E_{\text{def}}^0 \) is the energy required to distort the protein and the ligand from their equilibrium geometry to that they have in the adduct. It was computed at the same level of theory used for the calculation of \( \Delta E_{\text{bind}} \) (see Section 2.1 and 2.2). \( \Delta G_{\text{corr}} \) incorporates all thermal and entropy corrections at 298 K and was computed at the HF-3c level of theory using the rigid-rotor harmonic oscillator (RRHO) approximation. The Partial Hessian Vibrational Analysis (PHVA) approach was used, as implemented in ORCA. The atoms that were constrained in the geometry optimization (see Section 2.1) were fixed in the PHVA. Harmonic frequencies were computed for the ligand as well as for the protein model relaxed with and without the ligand bound. No imaginary frequencies were observed. Finally, \( \Delta G_{\text{solv}} \) was computed at the B3LYP level of theory using the continuum solvation model C-PCM. The dielectric constant \( \epsilon \) was set to 4, which is usually considered to be a good representation of the dielectric properties of the protein interior. Note that the trend observed for \( \Delta G_{\text{bind}} \) is the same irrespective of the specific implicit solvation method used, as detailed in Supporting Information (see Table S8).

In contrast to imidacloprid, nicotine is a positively charged ligand and is expected to interact strongly with polarizable residues. Consistent with this picture, nicotine seems to interact significantly with W143 in both sensitive (Figure 3(B)) and resistant (Figure 3(D)) models. However, neither the geometries nor the QTAIM analysis provides any direct evidence for relevant interactions to either T55 (in the resistant model) or R55 (in the sensitive model). Moreover, only a very weak interaction was found with C187–C188 in the sensitive model, but shows a noticeable interaction with the same residue in the resistant model. For both models, imidacloprid shows a significant interaction with a water molecule and with a tryptophan residue (W143), while no significant interaction was detected with the tyrosine Y185 and Y192 residues.

These results appear to be consistent with the widely-accepted selectivity mechanism proposed in seminal works by the groups of Tomizawa and Casida, as well as Satelle, and are also consistent with previous three-dimensional quantitative structure–activity relationships (3D-QSAR) and pharmacore models developed for neonicotinoid binding to nAChRs. For sensitive insect nAChRs, it is commonly argued that the nitro-moiey of imidacloprid is capable of binding strongly to protonated residues like lysin or, as in our case, arginine. In contrast to imidacloprid, nicotine is lacking an
electron-rich group analogous to the nitro-group and it is furthermore protonated under physiological conditions. These two factors are expected to have negative impact on the ligand binding to sensitive insect nAChR. For vertebrate nAChR, the lack of protonated residues is thus advocated as a major reason for the higher vertebrate toxicity of nicotine compared to its rather moderate insecticidal activity, in analogy to acetylcholine. Moreover, the Tomizawa and Casida model predicts that the positively charged center in protonated nicotine can undergo a cation-\(\pi\) interaction with a key tryptophan. If and how the partially positively charged guanidine-center of imidacloprid is able to "mimic" the protonated nitrogen of nicotine, thus giving rise to a cation-\(\pi\)-like interaction to tryptophan as well, is still under debate. Based on atomic charge calculations, the Sattelle model of imidacloprid binding assumes that binding of the guanidinium-moiety to tryptophane is synergistically enhanced by the binding of the nitro group to a basic aminoacid. In any model, the importance of the tryptophan to neonicotinoid-binding to nAChRs is out of question.27,30,32,56

3.2 Comparison between different electronic structure methods

Before starting our discussion on the nature of the ligand-nAChR interaction, we calculated the binding energies \(\Delta E_{\text{bind}}\) (Equation (1)) at various levels of theory in order to compare the performance of different electronic structure methods in this context. This preliminary benchmark study was conducted on the systems containing imidacloprid. The resulting \(\Delta E_{\text{bind}}\) values are reported in Table 1. Interestingly, dynamic electron correlation is found to play a fundamental role in these systems. The electron correlation contribution to the binding energy at the DLPNO-CCSD(T) level is always larger than 60 kcal/mol, whether calculated at the coupled cluster or RI-MP2 level. These results are reminiscent of those recently found by some of us in the calculation of enzymatic energy barriers57 and highlight the importance of incorporating electron correlation accurately when studying protein–ligand interactions.

For these systems, RI-MP2 overestimates binding energies with respect to the DLPNO-CCSD(T) reference values, while HFLD provides results that are closer to the reference coupled cluster data.41 The failure of RI-MP2 and the good performance of HFLD provide evidence that dispersion forces between the ligands and the residues constitute the dominant electron correlation contribution to the binding energy.41,58,59 These results suggest that the HFLD approach could be used as a cost-effective alternative to DLPNO-CCSD(T) in this context. Interestingly, B3LYP-D3 results slightly overestimate binding energies with respect to the DLPNO-CCSD(T) reference, providing an accuracy between that of RI-MP2 and HFLD. The -D3 dispersion correction greatly contributes to the accuracy of the DFT predictions, which is also consistent with the above-mentioned importance of dispersion forces in this context. In terms of computational time, the calculation of the protein–ligand adduct energy involving the sensitive nAChR model system took 2, 2, 17, 29, 77, and 130 h for HF, B3LYP, HFLD, RI-MP2, DLPNO-CCSD, and DLPNO-CCSD(T), respectively. In all cases, 16 cores from a single cluster node equipped with 4 Intel Xeon CPUs were used. In the following, only DLPNO-CCSD(T) binding energies are discussed.
the selectivity trend. Hence, for the sake of simplicity, only with a differential binding energy $\Delta E_{\text{bind}}^R$ and thus obtained are reported in Table 2.

An important point to note is that the differential binding energy, $\Delta E_{\text{bind}}^R = \Delta E_{\text{bind}}^\text{Resist} - \Delta E_{\text{bind}}^\text{Sens}$, is obtained by analyzing the ligand-nAChR binding energies at the DLPNO-CCSD(T)/LED studies of protein–ligand interactions.

New insights into the nature of the selectivity differences discussed in Section 3.1 can be obtained by analyzing the ligand-nAChR binding energies $\Delta E_{\text{bind}}$ and $\Delta G_{\text{bind}}$ obtained using the DLPNO-CCSD(T)/LED methodology outlined in Sections 2.2 and 2.3. The $\Delta E_{\text{bind}}$ and $\Delta G_{\text{bind}}$ thus obtained are reported in Table 2.

Although deformation, free energy and solvation corrections provide significant "repulsive" contributions to the binding energy, $\Delta E_{\text{bind}}$ and $\Delta G_{\text{bind}}$ show similar trends. These results suggest that the nature of the ligand–protein interaction plays a crucial role in determining the selectivity trend. Hence, for the sake of simplicity, only $\Delta E_{\text{bind}}$ values are discussed in the following.

For imidacloprid, we found an interaction energy of $-75.7$ kcal/mol with the sensitive nAChR model and of $-52.4$ kcal/mol with the resistant model. For nicotine, we found interaction energies of $-50.2$ and $-81.9$ kcal/mol for sensitive and resistant models, respectively. Thus, the binding of imidacloprid to insect nAChR is stronger than that of nicotine, with a differential binding energy $\Delta E_{\text{bind}}^R = -25.5$ kcal/mol. By contrast, nicotine interacts preferably with the nAChR featuring the "R55T" resistance mutation, showing a differential binding energy $\Delta E_{\text{bind}}^R = 29.5$ kcal/mol.

Switching perspective, imidacloprid interacts preferentially with the R55 variant of the binding site over the "R55T" mutated by $\Delta E_{\text{bind}}^R = -23.3$ kcal/mol, while nicotine prefers the "R55T" binding site by $\Delta E_{\text{bind}}^R = 31.7$ kcal/mol. These results clearly reflect the experimentally observed differences in selectivity, that is, the effect of the "R55T" resistance mutation to the interaction of the ligand with the receptor, decreasing its capacity of binding with nicotine and increasing its affinity with imidacloprid.

These energetic considerations provide a qualitatively consistent framework for the observed behavior of the two drugs in their interactions with the natural occurring and mutated receptors. However, by itself, such calculations do not provide any insight into how and why the energetics play out the way they do beyond the level of qualitative speculation. Below, we demonstrate that the DLPNO-CCSD(T)/LED methodology allows for a precise breakdown of the (differential) interaction energies in terms of individual drug/receptor contact points. This provides quantitative insights into the observed selectivity differences since it allows one to quantitatively judge the importance of each individual interaction and how it changes by mutation. (Figure 3).

The results of the DLPNO-CCSD(T)/LED binding energy decomposition into additive, fragment-pairwise contributions are illustrated in terms of "LED interaction maps" shown in panels (A)–(D) of Figure 4. It emerges from Figure 4 that, upon binding of the ligand to the receptor, a new pattern of intermolecular interactions is established between the ligand and the protein target’s residues in the active site. The energy associated to the individual ligand–residue interactions is shown in the "Ligand" row of each panel. As a consequence of these newly formed interactions, the electronic structure of the ligand and of the residues is perturbed. The diagonal elements of each panel denote the energy changes associated with these perturbations. Finally, ligand binding also affects the interactions between the residues in the active site, as a consequence of the changes in their electronic structure. The associated changes in the residue–residue interactions are also shown in each panel. The difference between the various panels provides the "LED ligand perturbation maps" (the changes in interaction upon changing the ligand) and the "LED mutation maps" (the changes in interaction upon mutating a residue in the active site), which provide a decomposition of the differential binding energies discussed above. The numerical values obtained from this analysis are given in Supporting Information. In particular, Figure S1 illustrates how the LED values are broken down into electrostatic, exchange and correlation energy contributions (where the latter, for the systems studied here, is dominated by the London dispersion energy).

The first eye-catching feature of Figure 4 is that the magnitude of the binding energy is dominated by the ligand–residue interactions and by the diagonal elements, while the cooperative effect between ligand–residue and residue–residue interactions is much smaller. This is an important finding, as it demonstrates that the binding energy is

### Table 2

| nAChR model | Ligand       | $\Delta E_{\text{bind}}$ | $\Delta G_{\text{bind}}$ |
|-------------|--------------|--------------------------|---------------------------|
| Sensitive   | Imidacloprid | $-75.7$                  | $-49.2$                   |
| Resistant   | Imidacloprid | $-52.4$                  | $-18.2$                   |
| Sensitive   | Nicotine     | $-50.2$                  | $-18.5$                   |
| Resistant   | Nicotine     | $-81.9$                  | $-45.3$                   |
essentially determined by a limited number of ligand–residue interactions, which can be easily quantified with our approach. Let us start by analyzing the binding of imidacloprid to sensitive nAChR (panel A). The strongest interaction partner for imidacloprid in sensitive nAChR is R55. As expected for an H-bonded system, the interaction is predominantly electrostatic in nature (81%, see Figure S1 and Table S1 in Supporting Information for the individual contributions to the interaction energies). In addition, and consistent with the Sattelle as well as the Tomizawa and Casida model, a noticeable attractive interaction between imidacloprid’s guanidine moiety and the tryptophan residue W143 is also found, which is stronger compared to the one that is occurring with all the other aromatic residues. Finally, the imidacloprid’s chlorine atom also exhibits a non-negligible interaction with L112, which is composed of 55% electrostatics, 16% exchange, and 29% correlation contribution. Similar interactions of chlorine with the L112-M114 fragment have been discussed for Chloro-thiazole-neonicotinoids. Consistent with these findings, the electronic structure of imidacloprid, R55, W143, and L112-M114 is significantly perturbed by the ligand–protein binding, as shown by the intense blue color in the corresponding diagonal elements of the LED interaction maps.

Let us start by analyzing the binding of imidacloprid to sensitive nAChR (panel A). The strongest interaction partner for imidacloprid in sensitive nAChR is R55. As expected for an H-bonded system, the interaction is predominantly electrostatic in nature (81%, see Figure S1 and Table S1 in Supporting Information for the individual contributions to the interaction energies). In addition, and consistent with the Sattelle as well as the Tomizawa and Casida model, a noticeable attractive interaction between imidacloprid’s guanidine moiety and the tryptophan residue W143 is also found, which is stronger compared to the one that is occurring with all the other aromatic residues. Finally, the imidacloprid’s chlorine atom also exhibits a non-negligible interaction with L112, which is composed of 55% electrostatics, 16% exchange, and 29% correlation contribution. Similar interactions of chlorine with the L112-M114 fragment have been discussed for Chloro-thiazole-neonicotinoids. Consistent with these findings, the electronic structure of imidacloprid, R55, W143, and L112-M114 is significantly perturbed by the ligand–protein binding, as shown by the intense blue color in the corresponding diagonal elements of the LED interaction maps.

The interaction of imidacloprid with "R55T" mutated nAChR (panel C) retains some of the key features just discussed. For instance, a strong interaction between the ligand and W153 is observed. Sattelle suggested that, in sensitive nAChR, such an interaction is further enhanced by the interaction of the nitro head group with R55.
LED shows, however, that the reduction in the interaction strength upon R55T mutation is only on the order of 12%. Moreover, the nature of the interaction, in terms of the contributions of electrostatics, exchange and correlation, is not affected by the presence or absence of R55, but remains at 62% electrostatics, 13% exchange, and 25% correlation. Moreover, the interaction with the L112-M14 fragment is also unaffected by the R55T mutation. Hence, the difference in binding of imidacloprid to sensitive nAChR and to the resistance mutant can thus be mostly attributed to the presence, or absence, of a positively charged residue; cooperative effects in the sense of the Sattelle model deliver only modest modulation.

All these effects are clearly illustrated by the corresponding LED mutation map reported in panel (A)–(C). Moreover, this plot reveals some additional information: when binding to the resistance mutant, imidacloprid intensifies binding to the conserved water molecule, to Y185, and, albeit weaker, to the C187–C188-bridge. However, all these newly formed stabilizing interactions do not compensate for the loss of interaction with R55.

Most computational studies focus on the binding behavior of neonicotinoids, or other synthetic agonists. Despite a lot of insight can come from understanding the binding of nicotine to nAChR, comparably few contributions deal with its binding behaviour. Our results reveal two main striking features for nicotine binding to sensitive nAChR, as evidenced from the analysis of Figure 4, panel B.

First, a very strong nicotine-W143 interaction, contributing to 67% of the total interaction energy of nicotine with all active site residues, was detected. Importantly, our analysis reveals that this strong interaction cannot be considered as a pure cation-π interaction, as would be possibly expected from the classical binding mode models. In fact, the protonated nitrogen of nicotine forms a very strong H-bond to a backbone carbonyl of residue W143, with electrostatics, exchange and correlation contributions being 75%, 11%, and 13%, respectively (Table S1). These results are in line with structural biology results on the binding of nicotine to human nAChR as well as earlier theoretical models. A comparison to the values discussed above for the respective interaction of imidacloprid (62%, 13%, and 25%) underlies the severe difference of the nature of how imidacloprid and nicotine interact with this tryptophan residue.

Second, a purely electrostatic repulsion between nicotine and the basic residue R55 was detected, thus providing quantitative evidence to the assumptions on the electrostatic repulsion of nicotinoids discussed above. In absolute values, the strength of this repulsion is comparable to the remaining attractive nicotine–protein interactions and amounts to +134.1 kcal/mol. In the resistance mutant, the positively charged arginine is replaced by a neutral threonine and thus the repulsive interaction is not observed. Interestingly, the interaction with W143 is weakened, while there is a slight increase in interaction with the water molecule and the L112-M114 fragment (LED mutation map “B–D”) in Figure 4.

Finally, as nicotine is lacking a feature analogous to the nitro-group of imidacloprid, R55 is lacking an interaction partner in the nicotine-nAChR complex and finds weak compensation in an interaction with the cysteine-bridge C187-C188. This is reflected in the LED ligand perturbation map (panel “A–B”) by the respective, prominently blue off-diagonal element.

4 CONCLUSIONS

We have demonstrated here that the DLPNO-CCSD(T)/LED scheme is an extremely powerful tool for the analysis of any kind of host/guest interaction. The methodology is capable to deliver in-depth, chemically intuitive insights into the nature and mechanisms of ligand binding, irrespective of whether it is electrostatic or dispersive in origin or a combination of both. Our LED maps appear particularly powerful for an in-depth understanding of hot spots of binding between ligands and the individual residues of the receptors, concerning: (i) the differences between different ligands binding to the same receptor; (ii) the differences between binding of one ligand to different mutations of a receptor. We have illustrated the power of this approach by analyzing the differential binding of nicotine and imidacloprid to insect and mammalian nAChRs. Our calculations did not only predict trends in binding strengths that correlate well with experimental selectivity differences, but also provide a quantitative analysis of which specific residues are most important for the interaction. We believe that the DLPNO-CCSD(T)/LED methodology opens up new avenues in the analysis of drug/receptor interactions and we hope that it will ultimately aid the design of new, highly specific drugs.

ACKNOWLEDGMENT

GB and Frank Neese gratefully acknowledge the Priority Program “Control of Dispersion Interactions in Molecular Chemistry” (SPP 1807) of the Deutsche Forschungsgemeinschaft for financial support. Generous support by the Max-Planck-Society is also gratefully acknowledged. Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

Michael Edmund Beck is an employee of Bayer AG, which is marketing products containing imidacloprid as active ingredient; Christoph Riplinger and Frank Neese are founders of FAccTs GmbH, which is marketing the ORCA software suite, for commercial usage (https://www.faccsts.de). For academic use, ORCA remains free (https://orcaforum.kofo.mpg.de/app.php/portal).

DATA AVAILABILITY STATEMENT

The data generated during this study are available within the article or its supplementary materials.

ORCID

Michael Edmund Beck https://orcid.org/0000-0001-9095-3881
Christoph Riplinger https://orcid.org/0000-0002-0166-6272
Giovanni Bistoni https://orcid.org/0000-0003-4849-1323

REFERENCES

[1] M. Danhof, K. Klein, P. Stolk, M. Aitken, H. Leufkens, Drug Discov. Today 2018, 23(12), 1990.
[2] C. A. Nicolaou, N. Brown, Drug Discov. Today Technol. 2013, 10(3), e427.
[3] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W.
