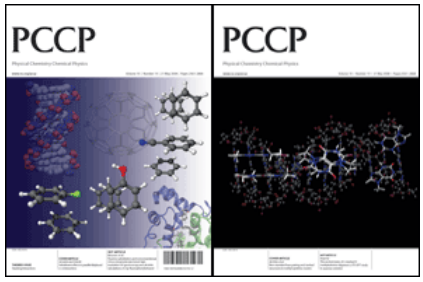
PCCP
Physical Chemistry Chemical Physics



This paper is published as part of a PCCP Themed Issue on:

Stacking Interactions

Guest Editor: Pavel Hobza

Editorial

Perspectives

Papers

Another role of proline: stabilization interactions in proteins and protein complexes concerning proline and tryptophane†

Lada Biedermannova,[ab] Kevin E. Riley,[a] Karel Berka,[a] Pavel Hobza[a] and Jiri Vondrasek*[ab]

Received 27th March 2008, Accepted 25th July 2008
First published as an Advance Article on the web 11th September 2008
DOI: 10.1039/b805087b

Proline–tryptophan complexes derived from experimental structures are investigated by quantum chemical procedures known to properly describe the London dispersion energy. We study two geometrical arrangements: the "L-shaped", stabilized by an H-bond, and the "stacked-like", where the two residues are in parallel orientation without any H-bond. Interestingly, the interaction energies in both cases are comparable and very large ($\sim$7 kcal mol^{-1}). The strength of stabilization in the stacked arrangement is rather surprising considering the fact that only one partner has an aromatic character. The interaction energy decomposition using the SAPT method further demonstrates the very important role of dispersion energy in such arrangement. To elucidate the structural features responsible for this unexpectedly large stabilization we examined the role of the nitrogen heteroatom and the importance of the cyclicity of the proline residue. We show that the electrostatic interaction due to the presence of the dipole, caused by the nitrogen heteroatom, contributes largely to the strength of the interaction. Nevertheless, the cyclic arrangement of proline, which allows for the largest amount of dispersive contact with the aromatic partner, also has a notable-effect. Geometry optimizations carried out for the "stacked-like" complexes show that the arrangements derived from protein structure are close to their gas phase optimum geometry, suggesting that the environment has only a minor effect on the geometry of the interaction. We conclude that the strength of proline non-covalent interactions, combined with this residue's rigidity, might be the explanation for its prominent role in protein stabilization and recognition processes.

Introduction

It is common opinion about L-proline that it plays a distinctive role in the structure and function of proteins because of the specific character of its side chain. The restraints brought by the cyclic structure of proline's side chain give this residue exceptional conformational rigidity compared to other amino acids. Upon folding, the residue loses less conformational entropy which may account, for example, for its higher occurrence in the proteins of thermophilic organisms.[1–3] Therefore, the stabilizing role of the proline residue for a protein tertiary structure is traditionally believed to be the consequence of its extraordinary rigidity.

The role of proline in secondary structure stabilization is, however, still a matter of controversy, as it is very sensitive to the relative position of the residue within a protein and to the surrounding environment. The commonly held opinion is that proline is the most potent alpha-helix breaker, at least in globular proteins soluble in aqueous media. On the other hand, a computational study by Yun *et al.*[4] revealed that when proline is the first residue at the N terminus of an alpha helix, it may be a better helix former than alanine, and thus serves as a helix initiator. Li *et al.*[5] showed in their study that the helical propensity of proline was greatly enhanced in membrane-mimetic environments and proline was found to stabilize the alpha-helical conformation relative to alanine. However, proline was shown to act as an alpha helix disruptor in transmembrane alpha helices[6,7] of various transmembrane proteins. In some antimicrobial peptides proline is an essential structural element of helical motifs and the helical kinks caused by proline play an important role in the disruption of bacterial membranes.[8] The role of proline in transmembrane transport proteins has also been extensively studied and it has been shown that proline residues are important for substrate binding and recognition.[9]

Although proline residue can only serve as a hydrogen bond acceptor and not as a hydrogen bond donor due to the lack of hydrogen on its amide group, it has been shown that the C-delta protons can be involved in C–H$\cdots$O hydrogen bonds with the carbonyl group of the preceding helical turn.[10] Proline residues also frequently occur in turns, which may

[a] *Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, Flemingovo nam. 2, 166 10 Prague 6, Czech Republic. E-mail: jiri.vondrasek@uochb.cas.cz; Fax: (+420) 220-410-320; Tel: +420 220-410-324*
[b] *Laboratory of Ligand Engineering, Institute of Biotechnology, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Prague 4, Czech Republic*

† This article was submitted as part of a Themed Issue on stacking interactions. Other papers on this topic can be found in issue 19 of vol. 10 (2008). This issue can be found from the *PCCP* homepage http://www.rsc.org/Publishing/Journals/CP/index.asp

explain the intriguing fact that, in spite of having an aliphatic side chain, it is very often solvent-exposed.

A remarkable role of the proline residue was suggested by Riley *et al.* in their study of inhibitors of human carbonic anhydrase II.[11] Their calculations revealed that the inhibitor's phenyl ring interacts about as strongly with the proline's pyrrolidine ring of the enzyme as with a phenylalanine phenyl ring in the binding pocket of the enzyme. This result is unexpected, considering the fact that aromatic–aromatic interactions have long been thought to be particularly favorable in protein and protein–ligand structures.[12–14] The authors suggested that a similarly strong interaction might occur between phenylalanine and proline residues in protein structures and that such intramolecular inter-residue interactions are very likely to have a non-negligible influence on protein stability and protein–protein interactions.

This could be a chemical–physical explanation for the essential role of proline in various phenomena of complex molecular behavior. To name just a few examples, we point to the SH3 domain. This small protein domain of ~60 residues can be found in various proteins and mediates the interaction with other proteins through a proline patch of consensus sequence -X-P–P-X-P-[15,16] where X is any one of the 20 L-amino acids. A similar example is a polyproline sequence of the mammalian enabled (Mena) protein recognized by WW domains of the FE65 protein responsible for brain development and important as an amyloid precursor.[17] A proline rich conserved region is necessary for the folding of cytochrome P450.[18] Lastly, the interaction sites within a proline rich domain of p53 protein are indispensable for tumor supression.[19]

Up to now the only high level computational study of proline interactions is a comprehensive work by Morozov,[20] describing the pairwise interaction in homodimers of ring-containing aminoacid residues, and the already mentioned work of Riley *et al.*[11] We have recently studied[21] the intramolecular stabilization of a small artificial protein Trp–cage[22] using advanced computational methods. In the structure of this miniprotein we noticed two very strong non-covalent interactions involving proline residues; in both cases the prolines interact with the central tryptophan residue and in both arrangements the stabilization energies are ~8 kcal mol^{-1} at the *ab initio* MP2/aug-cc-pVDZ level. Looking for geometrically similar arrangements of proline-tryptophane interacting residues we found that the review of Ball *et al.*[23] describes interactions of proline rich motifs with EVH1 and GYF domains. Interestingly, the specific arrangements of signal peptide prolines with corresponding proline binding domains resembled the intramolecular stabilizing motif discovered in the Trp cage and gave rise to a question regarding the importance of aromatic residues in maintaining this contact. We therefore could apply the below described strategy and computational methods to the system of practical importance and compare the energetic background of intra- and inter-system interactions.

Therefore, the details of a proline–aromatic residue interaction have not been studied by means of quantum chemistry so far. Hence there are several questions concerning the nature of this interaction, *e.g.* role of dispersion forces, importance of proline cyclic arrangement and presence of heteroatom for the strength of the interaction. The only way to understand the nature of these interactions and to answer these questions is through advanced *ab initio* calculations that accurately treat electron correlation. It is well known[24] that correlation/dispersion is particularly significant for interactions between molecules of non-polar character. As we focus on the interactions of residues having aromatic and aliphatic character, we need to use theoretical methods that take electron correlation into account, and that are at the same time computationally feasible for the size of the systems in focus. Such methods are the Møller–Plesset perturbation method (MP), the density functional theory augmented with empirical dispersion term (DFT-D), and the symmetry adapted perturbation treatment based on density functional theory (DFT-SAPT). Using these methods, we aim to investigate the character of the proline–aromatic residue interaction and to determine its most important structural and electronic properties. In order to validate our results, we also use the highly precise coupled clusters method with complete basis set extrapolation as a benchmark for selected systems.

Methods

Selection of model systems

For the calculation of aromatic–proline interaction we have selected two model systems, both taken from the structure of Trp-cage (PDB code 1L2Y), namely the pairs of residues Pro17–Trp6 and Pro18–Trp6 (Fig. 1). We employ two models of the interacting moieties: the "large model" and the "small model". The large model represents the studied residues as shown in Chart 1; the residue model includes the carbonyl

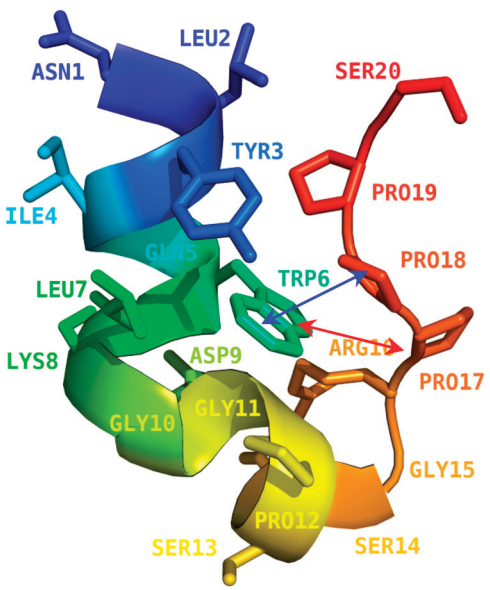


Fig. 1 The geometry of Trp-cage miniprotein. The L-shape arrangement of interaction between TRP6 and PRO17 is represented by double arrow red line whereas the stacked-like arrangement of TRP6 and PRO18 is shown in blue double arrow line.

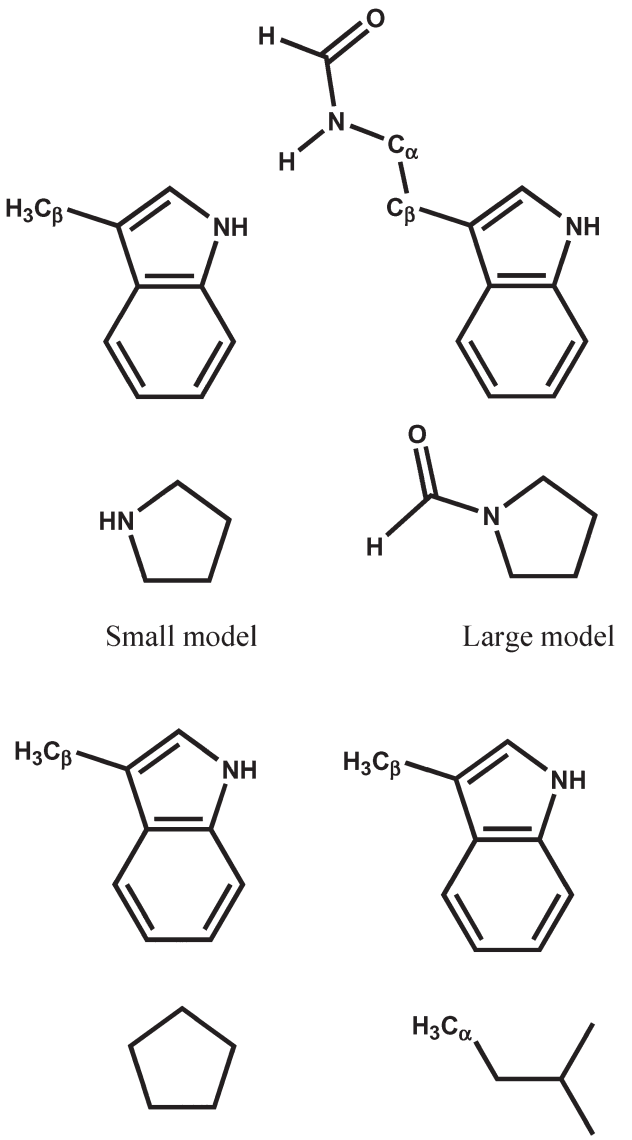


Chart 1 Chemical structure of the molecules used as model systems for the proline–aromatic interaction.

group of the preceding residue in order to take the peptide bond into account (the protein backbone is cut at the C–C_α bond and the resulting fragments capped with hydrogen atoms). This is intended for the determination of the contribution of these polar atoms to the overall stability of a residue–residue complex, and allows for comparison to previous theoretical works on proline interactions, which also included the carboxyl group. In the small model the system is reduced to the side-chain only, starting from the C_β atom; proline is represented as a pyrrolidine molecule to preserve its cyclic structure.

To evaluate the role of proline's cyclic arrangement and nitrogen heteroatom on the strength of stabilization, we use another two models. Replacing the NH group of proline in the small model by a CH_2 group results in a cyclopentane–tryptophan complex, *cf*. Chart 1. For an acyclic model with the same number of heavy atoms, we use the leucine–tryptophan complex (the side chain of leucine starting from C_α atom). The structure of the Leu–Trp complex was taken from the Protein Side-Chain Atlas (http://www.biochem.ucl.ac.uk/bsm/sidechains/Trp/Leu/sindex.html/). We selected the representative cluster No. 1 of Trp–Leu pair, in which Leu and Trp side chains adopt a stacked-like arrangement. (By a stacked-like arrangement we denote a complex where the heavy atoms of a non-aromatic moiety lie roughly in a plane parallel to the plane of the aromatic partner.)

To compare and evaluate the strength of interaction between Trp–Pro intramolecular stabilizing motif found in the Trp cage and the intermolecular interaction motif found in EVH1 and GYF binding domains with proline rich peptides we used another two X-ray structures—1EVH and 1L2Z. There are four intermolecular Pro–Trp complexes: P2-P5···W23 in 1EVH and P66-P69···W28 in 1L2Z which we further explored by means of interaction energy evaluation. Only the large model representation was used because of the mixed character of the existing aromatic–Pro interactions in both protein complexes. Each Pro–Trp pair was cut out from the structure and treated as separate pairwise complex.

Geometry optimizations

In all studied systems we first perform geometry optimization of hydrogen atoms using the DFT/B3LYP/6-31G** method. The geometries of all complexes with only hydrogen atoms optimized were used for the initial energy analysis (Table 1). The full geometry optimization in selected systems is performed using DFT-D/TPSS/TZVP method. For the restricted optimization of the Pro–Trp small model, we use the same method and we fix the position of proline N atom and all Trp residue heavy atoms.

Interaction energies. For the evaluation of the interaction energies the following set of methods is applied. The much used MP2/aug-cc-pVDZ method, the recently introduced DFT-D[25] method with TPSS functional and TZVP basis set, and for benchmark purposes the CCSD(T)/CBS procedure is also applied for selected systems.

The reason for the selection of the aug-cc-pVDZ basis set in connection with MP2 is the well known fact that MP2 procedure overestimates interaction energies when combined with extended basis sets. Because it is specifically the dispersion part that is overestimated, this behavior is most profound in stacked complexes, where dispersion energy represents a dominant attractive term.[26,27] Therefore the MP2 stabilization energies calculated with medium basis sets are, in most cases, more realistic and reliable.

In the DFT-D method the DFT energy is augmented by an empirical term that describes the London dispersion energy. The parameters in the accompanying damping function have been parameterized against accurate CCSD(T)/CBS interaction energies and the method is thus supposed to yield interaction energies close to CCSD(T)/CBS values.[28] The strong point of the method is its computational efficiency, which allows for its use even with extended complexes. Further, we use the RI technique to speed up the calculations. This technique can be also used for geometry optimization.

Table 1 Interaction energies (in kcal mol^{-1}) in the L-shaped and stacked-like arrangements of Pro–complexes in various models of the interaction calculated using CCSD(T), MP2, DFT-D and SAPT methods, and the individual components of the SAPT interaction energy

	L-shaped		Stacked-like			
	Large model	Small model	Large model	Small model	Trp–cyclopentane	Trp–Leu
MP2[a]	−7.8	−0.9	−8.4	−6.5	−5.1	−4.1
DFT-D[b]	−7.6	−1.1	−6.8	−5.4	−4.0	−2.9
D part[c]	−1.9	−1.1	−6.3	−5.0	−5.7	−4.1
CCSD(T)[d]	N/A	−0.9	N/A	−6.0	−4.6	−3.8
SAPT[e]	−7.3	−0.8	−6.9	−5.3	−3.7	−3.3
E(elec.)[f]	−9.9	0.2	−5.6	−3.4	−2.0	−1.5
E(ind.)[f]	−3.7	−0.1	−1.0	−0.5	−0.3	−0.2
E(disp.)[f]	−4.5	−1.2	−8.8	−6.6	−6.8	−4.7
E(exch.)[f]	11.2	0.3	9.2	5.6	5.6	3.5
δHF[f]	−1.8	0.0	−0.7	−0.4	−0.4	−0.3

[a] MP2/aug-cc-pVDZ. [b] DFT-D/TPSS/TZVP results. [c] The dispersion correction term in the DFT-D interaction energy. [d] CCSD(T)/CBS. [e] DFT-SAPT/aug-cc-pVDZ interaction energy. [f] The electrostatic, induction, dispersion, exchange and higher order contribution terms to the SAPT interaction energy, respectively.

To obtain highly precise interaction energies it is necessary to pass to higher level calculations *e.g.* CCSD(T)), or the ΔCCSD(T) corrected MP2 level (the ΔCCSD(T) correction term corresponds to the difference between CCSD(T) and MP2 interaction energies for a given basis set and, in the case of stacking, is always repulsive). We have performed such calculations in this study as a benchmark for selected structures. We calculate the CCSD(T) interaction energy at the complete basis set limit (CBS) as follows:

$$\Delta E_{\mathrm{CBS}}^{\mathrm{CCSD(T)}} = \Delta E_{\mathrm{CBS}}^{\mathrm{MP2}} + \Delta\mathrm{CCSD(T)} \qquad (1)$$

The $\Delta E_{\mathrm{CBS}}^{\mathrm{MP2}}$ term denotes MP2 interaction energy extrapolated to complete basis set limit and was determined using the Helgaker extrapolation scheme.[29] The Hartree Fock and correlation MP2 energies necessary for the extrapolation were determined with aug-cc-pVXZ (X = D, T) basis sets. The ΔCCSD(T) term was calculated with a smaller basis set, 6-31G*(0.25) (exponent of d-functions changed from a standard value of 0.8 to a more diffuse one of 0.25). The use of a smaller basis set is based on the fact that the difference between the MP2 and CCSD(T) interaction energies (contrary to MP2 and CCSD(T) total energies themselves) is much less dependent on the size of the basis set, and the 6-31G*(0.25) basis set already gives satisfactory values of this difference.[30]

All interaction energies were corrected for the basis set superposition error.

Interaction energy decomposition

The above mentioned methods provide total interaction energies. For the purpose of evaluating the origin of complex stabilization it is desirable to also have energy components. Partial splitting can be obtained from RI-DFT-D calculations, which yield DFT interaction energies and dispersion energies separately. Full energy decomposition is obtained by using the DFT-SAPT method, which allows for the separation of interaction energies into physically meaningful components such as those arising from dispersion, electrostatics, induction, and exchange-repulsion.[31–34]

The SAPT interaction energy is given as:

$$E_{\mathrm{int}} = E_{\mathrm{pol}}^{(1)} + E_{\mathrm{ex}}^{(1)} + E_{\mathrm{ind}}^{(2)} + E_{\mathrm{ex\text{-}ind}}^{(2)} + E_{\mathrm{disp}}^{(2)} + \delta\mathrm{HF} \qquad (2)$$

Some of these terms can be combined in order to define values that correspond to commonly understood physical quantities. In this work we define the following equalities:

$$E(\mathrm{elec.}) = E_{\mathrm{pol}}^{(1)},$$

$$E(\mathrm{ind.}) = E_{\mathrm{ind}}^{(2)} + E_{\mathrm{ex\text{-}ind}}^{(2)},$$

$$E(\mathrm{disp.}) = E_{\mathrm{disp}}^{(2)} + E_{\mathrm{ex\text{-}disp}}^{(2)},$$

and

$$E(\mathrm{exch.}) = E_{\mathrm{ex}}^{(1)}.$$

These four quantities refer to the electrostatic, induction, dispersion, and exchange-repulsion contributions (respectively) to the overall interaction energy. The δHF term is a Hartree–Fock correction for higher order contributions to the interaction energy that are not included within the second order DFT-SAPT terms. In this work we have performed DFT-SAPT computations using Dunning's aug-cc-pVDZ basis set. The computational details of the DFT-SAPT calculations performed are analogous to those described by Jansen *et al.*[35]

Calculations were performed in Gaussian 03,[36] Molpro v.2006.1[37] and Turbomole[38] v.5 programs as well as our own code for the DFT-D calculations.[39]

Results and discussion

For the examination of proline–tryptophan interactions we select the two interacting Pro–Trp complexes, which have shown extraordinarily high stabilization in the Trp-cage mini-protein:[21] the Pro17–Trp6 residue pair, hereafter referred to as the L-shaped arrangement, and the Pro18–Trp6 residue pair, referred to as the stacked-like arrangement (see Fig. 1). We adopt several models of these interacting complexes, as described in the methods section.

First we examine both the large and small models of the L-shaped complex (see Fig. 2A and B). The results summarized in Table 1 show that there is a very strong interaction in the large model (-7.6 kcal mol^{-1} at DFT-D level and -7.8 kcal mol^{-1} at MP2 level). The large magnitude of this interaction is not surprising as there is a hydrogen bond between the tryptophan ring NH group and the proline backbone carbonyl group. Notice also the very good agreement between MP2 and DFT-D interaction energies, which is characteristic of H-bonded complexes, also confirmed by the CCSD(T) value calculated for the small model. The key role of the H-bond for stabilization in this arrangement is confirmed by the evaluation of the interaction energy in the small model, where no hydrogen bond is present and the interaction energy is almost negligible (-1.1 kcal mol^{-1} at DFT-D level, -0.9 kcal mol^{-1} at MP2 level, -0.9 kcal mol^{-1} at CCSD(T)/CBS level). The backbone portion of the tryptophan residue can be assumed to contribute negligibly to the interaction due to its distance from proline ($\sim$7 Å). It can be further seen that in the case of the large model the dispersion energy plays only a minor role (*cf*. the *D part* of the DFT-D energy). In the case of the small model the situation is different, because a majority of the stabilization comes from dispersion energy.

The results of the DFT-SAPT interaction energy decomposition for the L-shaped complex are also shown in Table 1. It can be seen that for the hydrogen bonded large model the interaction energy is slightly underestimated compared to both MP2 and DFT-D. This strong (-7.3 kcal mol^{-1}) interaction is largely dominated by the electrostatic term (-9.9 kcal mol^{-1}). This result is expected for a hydrogen bonding pair; the relatively large induction (-3.7 kcal mol^{-1}) and dispersion (-4.5 kcal mol^{-1}) terms are also not surprising. For the small Pro17–Trp6 model, which contains no hydrogen bonding interaction, the attraction between the two residues is weak (-0.8 kcal mol^{-1}) and can be chiefly ascribed to the dispersion term. To sum up, the L-shaped arrangement of Pro–Trp residue pair seems to be of interest because of its unusually large interaction energy. The strength of the interaction is,

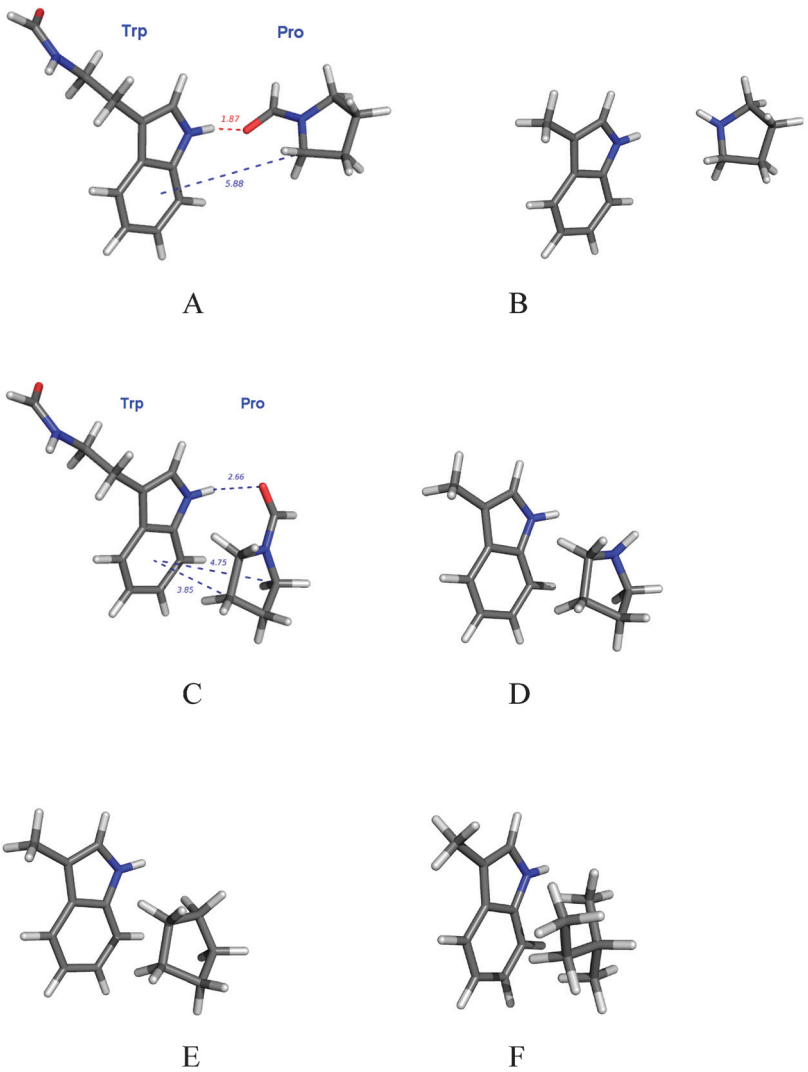


Fig. 2 Large (A) and small (B) models of interaction between Pro17–Trp6, large (C) and small (D) model of interaction between Pro18–Trp6, Trp6 interacting with cyclopentane (E) based on geometry of Pro18 and Trp–Ile complex (F) in stacking orientation obtained from the Atlas of Protein Side-Chain Interactions.

however, not particularly surprising because it is chiefly attributable to classical H bonding.

The situation is more interesting in the complex with a stacked-like arrangement (see Fig. 2C and D). The results for this complex are also summarized in Table 1. There is no H-bond between the two interacting systems, and the truncation of the carbonyl group leads to only a moderate decrease in the interaction energy in this case (−6.8 *vs.* −5.4 kcal mol^{-1} at the DFT-D level, and −8.4 *vs.* 6.5 kcal mol^{-1} at MP2 level). Comparing to the reference CCSD(T) values, the MP2 energies seem to be relatively overestimated, while DFT-D (as well as SAPT) values are slightly underestimated. As already discussed above, the MP2 method is known to overestimate interaction energies, especially of stacked complexes, when used with extended basis sets. The use of smaller basis sets, like the aug-cc-pVDZ used here, usually leads to cancellation of errors, which, however, cannot be relied on in all cases. The underestimation of the interaction energy by the DFT-D method, on the other hand, might be caused by the fact that the S22 set of complexes used for parameterization of this method includes mainly H-bonded and classical stacked complexes.

As follows from the dispersion part of the total stabilization energy obtained by the DFT-D method the dispersion term constitutes a major portion of the interaction energy in both the large and the small model (−6.3 and −5.0 kcal mol^{-1}, respectively). The prevailing character of the interaction in the stacked-like arrangement is therefore dispersive.

The strength of the interaction is surprising since it is not a classical stacked interaction of two aromatic systems, such as the interaction observed in the stacked phenylalanine dimer, neither is it a H-bonded complex. Here the proline molecule is an aliphatic system without any π electrons, and so this interaction can be classified as a mixed π/aliphatic type. In the present case the MP2 and DFT-D interaction energies differ more than in the previous L-shape arrangement.

In terms of DFT-SAPT analysis, the interaction in the stacked-like complex is also very different in nature from the L-shaped complex, as demonstrated by the data given in Table 1. It would be probably worth to stress here that dispersion term obtained by DFT-SAPT method and that coming from DFT-D method are quite different in their nature. The DFT-D is DFT augmented by dispersion which is constructed to fit a sum of DFT energy and empirical term containing dispersion correction on reference CCSD(T) values. The dispersion term in SAPT is evaluated as one of the second order corrections which is an independent term in the perturbational treatment of the interacting complex besides the induction term. The only thing worth mentioning here is that the E_{disp} term obtained by SAPT better captures the nature of the interaction. Based on the results of the analysis one can determine a ratio between the major stabilizing terms. First considering the results obtained for the large model, the overall strength of interaction is comparable with the L-shaped complex (−6.9 *vs.* −7.3 kcal mol^{-1}), but significantly more dispersive in its character (−8.8 *vs.* −4.5). The fact that this interaction is dominated by the dispersion term is not unusual, as this is typically found for stacked structures.[40] However, the electrostatic term is also relatively large (−5.6 kcal mol^{-1}). A possible explanation for this observation could be the presence of two C–H/pi interactions between hydrogens located on the proline ring and each of the aromatic rings on tryptophan.

Looking at the results for the small model complex, each of the DFT-SAPT interaction energy terms is smaller than in the case of the large model. This weakening can be explained by two factors: a smaller number of atoms that can participate in dispersive contacts (*cf.* decrease of dispersion and exchange terms) and the change of the electronic character of the proline and tryptophan rings in the absence of main chain atoms (lower electrostatic term). Despite this decrease, the resulting interaction energy is still relatively high (−5.3 kcal mol^{-1} at the DFT-SAPT level and −6.0 kcal mol^{-1} at the CCSD(T) level). This is an unusual interaction strength considering that only one interacting partner is of aromatic character and the interaction thus cannot be classified as a classical stacking.

In order to uncover the key structural features responsible for the large interaction energy observed even in the small model of the stacked-like complex, we investigate the role of the nitrogen heteroatom and the importance of the cyclic arrangement. To this end, we compare the already discussed results for tryptophan–proline stacked complexes with two other models of the interaction, the cyclopentane–tryptophan and the leucine–tryptophan models (see Fig. 2E and F). For details on the selection of these systems see the Methods section.

The data given in Table 1, columns 5 to 7, show that the replacement of the NH group by a CH_2 group reduces the interaction energy non-negligibly by about 1.5 kcal mol^{-1}, and passing to the acyclic Leu–Trp model further decreases the interaction energy by about 1.0 kcal mol^{-1}. Nevertheless, the resulting interaction energy is still relatively large.

Comparing DFT-SAPT data for the cyclopentane–tryptophan model with the small model given in Table 1, the most prominent aspect is that the electrostatic term is substantially smaller for the former one, while the induction, dispersion, and exchange terms are all very similar. It is obvious that the presence of the N heteroatom contributes significantly to the large interaction energy value in the small model of proline–tryptophan complex. Passing from proline to cyclopentane evidently reduces the molecular dipole moment, resulting in a smaller electrostatic component of the interaction energy.

The role of cyclicity was explored using the Leu–Trp model. It would be expected that this interaction should be weaker because of the fact that, for leucine, there is one fewer atom in dispersive contact with the tryptophan heterocyclic ring than in the case of cyclopentane. In Table 1 it can be seen that the Trp–Leu binding energies (for each of the methods used here) are about 1 kcal mol^{-1} lower (less stable) than those for the Trp–cyclopentane complex, indicating that our original supposition was correct. The fact that the dispersion and exchange terms are much lower in this case indicates that there is less overlap between the interacting molecules. This corresponds to the fact that the density of atoms within a cyclic molecular arrangements should lead to more contact between two molecules.

To sum up, the large stabilization in the stacked-like arrangement of Trp–Pro complex is a result of several factors,

Table 2 Interaction energies (in kcal mol^{-1}) calculated using DFT-D in the stacked-like arrangement of Pro–Trp complex in various models of the interaction after their full geometry optimization by DFT-D method; and the SAPT interaction energy decomposition in these complexes

	Stacked-like			
	Small model	Small model[a]	Trp–cyclopentane	Trp–Leu
DFT-D[b]	−11.6 (−5.4)	−5.5 (−5.4)	−4.1 (−4.0)	−3.2 (−2.9)
D part[c]	*−5.0 (−5.0)*	*−4.9 (−5.0)*	*−5.4 (−5.7)*	*−3.7 (−4.1)*
SAPT[d]	−9.5 (−5.3)	−5.4 (−5.3)	N/A	N/A
E(elec.)[e]	*−16.0*	*−3.4*	*N/A*	*N/A*
E(ind.)[e]	*−3.1*	*−0.5*	*N/A*	*N/A*
E(disp.)[e]	*−8.6*	*−6.5*	*N/A*	*N/A*
E(exch.)[e]	*21.4*	*−5.4*	*N/A*	*N/A*
δHF[e]	*−3.2*	*−0.4*	*N/A*	*N/A*

[a] Restricted position of N atom. [b] DFT-D/TPSS/TZVP results. [c] The dispersion correction term in the DFT-D interaction energy. [d] DFT-SAPT/aug-cc-pVDZ interaction energy. [e] The electrostatic, induction, dispersion, exchange and higher order contribution terms to the SAPT interaction energy, respectively; the value of interaction energy in the unoptimized complex, given in parentheses.

of which the role of the heteroatom is the most important but the cyclic arrangement also plays a non-negligible role.

In the next step of this study we examine how far the arrangements derived from protein structures are from their optimum geometric positions in the gas phase. The strength of the interaction in the stacked arrangement is so large that it seems very unlikely that a more stable geometry of a similar arrangement could exist in the context of protein structure. In order to determine the minimum energy conformations we perform geometry optimization of the following model complexes of the stacked arrangement: Pro–Trp small model, Trp–cyclopentane model and Trp–leucine model. We selected the DFT-D method for this purpose, because of its computational efficiency and its correct description of trends in the interactions of these complexes.

The interaction energies calculated using the DFT-D method in the optimized geometries are shown in Table 2. We observed that the Pro–Trp small model complex relaxes into a geometric position in which an H-bond is formed between the NH group within the Trp heterocycle (H-bond donor) and the N atom in the Pro ring (H-bond acceptor), with a very large interaction energy of −11.6 kcal mol^{-1} (*cf.* the value of interaction energy in the unoptimized complex, given in parentheses). This H-bond formation is, however, not realistic with respect to the initial structure from which the motif has been taken and does not reflect the accessible geometric possibilities for a proline residue within the protein structure, where its movement would be restricted by the presence of the backbone.

Therefore, we repeated the geometry optimization with the proline N atom position fixed relative to the Trp molecule to mimic the constraints imposed by the protein backbone (for details, see the Methods section). This time, the geometry of the complex changed only slightly during the course of optimization and the interaction energy for the optimized complex is very similar to that of the unoptimized system, *cf.* column 3 in Table 2.

On the other hand, the geometries changed very little in the cases of the cyclopentane and leucine models, where no H-bond formation is possible, even without any restraints.

When examining the DFT-SAPT results in Table 2, it can be seen that the interaction occurring for the fully optimized complex is predominantly of an electrostatic nature, consistent with a hydrogen bonding interaction. DFT-SAPT results for the restrictively optimized tryptophan–proline complex are remarkably similar to the results obtained for the unoptimized structure (differences in all energy terms are within 0.2 kcal mol^{-1}, and the total interaction energy is only 0.1 kcal mol^{-1} higher than that of the unoptimized structure), confirming that the two structures are very similar.

We therefore conclude, based on these geometry optimizations, that the residue contacts within protein structures are very close to their backbone constrained geometry optima. This also suggests that it is the "gas phase" interaction energies/enthalpies between protein building blocks that mostly determine the structure of the protein interior, while other influences (*e.g.* solvent) play only modulating roles.

As follows from the detailed study of the Pro–Trp intramolecular interaction in Trp cage the energy content of the interaction is quite large. Therefore it is interesting to compare intermolecular and intramolecular representations of Pro–Trp pairs and characterize their energy/geometry similarities and differences. We evaluated the interaction of the Trp residue with the proline polypeptide in two different protein–ligand complexes. The first representing the EVH1 binding domain

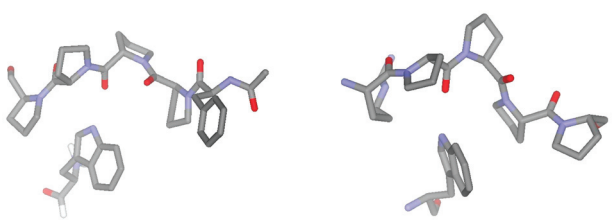


Fig. 3 Detailed intermolecular interactions of two representative members of protein rich motives (PRM)-binding families with their peptide ligands. Single binding modes known to date for the GYF (a) and one of the binding modes for the EVH1 domains (b).

Table 3 Interaction energies (in kcal mol^{-1}) for Trp–Pro complexes from two protein–ligand structures. Interaction energy is calculated by DFT-D/TPSS/TZVP methods

Interacting pairs/systems	pdb ID 1EVH	pdb ID 1L2Z
P2_W23 and P66_W28	*−0.5*	*−2.5*
P3_W23 and P67_W28	*−1.5*	*−7.7*
P4_W23 and P68_W28	*−8.1*	*−4.2*
P5_W23 and P69_W28	*−5.0*	*−1.2*

and Pro rich sequence is shown in Fig. 3a, the second representing the GYF binding domain with polyproline peptide is shown in Fig. 3b. Only the central Trp residue of the protein is shown and the proline polypeptide is truncated to the length of interacting Pro residues. As follows from Table 3 the interaction energies of each Pro–Trp pair in large model representation evaluated for both structures are quite different. Interestingly, two largest energy stabilizations (−7.7 and −8.1 kcal mol^{-1}) come from the "L-shape" like arrangement but the next largest contributions (−5.0 and −4.2 kcal mol^{-1}) are from the "stacked" like Pro–Trp pairs. The striking fact is that in both instances the energies as well as geometries are comparable or similar, pointing to some general feature of such interaction pattern.

Conclusions

It has long been known that interactions between aromatic residues play an important role in protein structure and function. Only recently has it been recognized that mixed interactions between aromatic and aliphatic residues, such as proline–tryptophan, might be of similar importance, as suggested by the recent works by Morozov *et al.*[20] and by Riley *et al.*[11]

In this work we have investigated two proline–tryptophan complexes derived from the experimental structure of the Trp–cage miniprotein, one in a "L-shaped" arrangement and the other in a "stacked-like" arrangement. The L-shaped arrangement features an H-bond proline and tryptophan residues, whereas in the "stacked-like" arrangements the residues are in a parallel geometry without any H-bond between them. We have performed correlated MP2 and DFT-D calculations of interaction energies, including benchmark CCSD(T)/CBS calculations for selected complexes, as well as a DFT-SAPT interaction energy decomposition. Our calculations showed that The L-shaped arrangement is very strongly stabilized and the main source of stabilization is the classical H-bond, as the truncation of the system leads to a dramatic interaction energy decrease (−7.6 *vs.* −1.1 kcal mol^{-1} for the large and small models, respectively, at DFT-D/TPSS/TZVP level). To put the obtained results into some broader context, let us have a look on a typical strength of non-covalent interactions between different types of systems determined at the highest level of theory.[41] For example, the typical hydrogen bonded complexes as, *e.g.*, NH3–NH3 or water–water amount −3.2 kcal mol^{-1} and −5.0 kcal mol^{-1}, respectively. Complexes with predominant dispersion contribution as, for example, benzene dimer or indole–benzene (representing Trp–Phe interaction in proteins) amount to −2.7 and −5.2 kcal mol^{-1}, respectively. In this context the stabilization energy of the studied Trp–Pro complex is large and most probably very important in proteins and their interactions.

The most important result of this study is, however, that the stacked-like arrangement of tryptophan–proline interaction is also bound very strongly, even without the presence of any classical H-bond, and the truncation of the system does not diminish the interaction energy as profoundly as in the previous case. (−6.8 *vs.* −5.4 kcal mol^{-1} and −8.4 *vs.* −6.5 kcal mol^{-1} for the large and small models, at DFT-D and MP2 levels, respectively) The fact that the dispersion term in the DFT-D method is responsible for most of the attractive force within this complex indicates that the strong interaction found therein is principally attributable to dispersion forces. This was confirmed by DFT-SAPT analysis, which predicts dispersion to be the most dominant interaction energy component of this interaction. However, it should be noted that the electrostatic contribution to this interaction is not negligible and is about half as strong as that of dispersion.

To investigate the role of proline's cyclic arrangement and the presence of heteroatom on the unusual strength of this interaction, we applied two other models of the stacked-like arrangement. Comparison of the small model of Pro–Trp interaction with the cyclopentane model revealed that replacing the NH group of proline with a CH_2 group leads to an interaction energy decrease of about 1.4 kcal mol^{-1} (at MP2, DFT-D as well as CCSD(T) levels). SAPT analysis reveals that the interaction found within the cyclopentane complex is much less electrostatic in nature than that of the proline complex, indicating that the proline dipole moment plays a role in stabilizing the proline–tryptophan complex. In order to study the role of cyclicity on the interaction, we carried out calculations on the leucine–tryptophan complex. It was found that compared to cyclopentane, leucine binds tryptophan about 0.9 kcal mol^{-1} more weakly (at MP2, DFT-D and CCSD(T) levels). The explanation for this decrease is the fact that the cyclic arrangements of proline and cyclopentane allow for the highest degree of dispersive contact with tryptophan. Thus, we conclude that for the unusual strength of this interaction, the presence of the N heteroatom is especially important, but the cyclic arrangement of proline also plays a non-negligible role.

In order to study how far the described geometries, derived from protein experimental structures, are from their geometry optima, we performed geometry optimizations of the models of the stacked-like arrangement. The small model of Pro–Trp stacked-like interaction underwent a significant geometry rearrangement, with a formation of an H-bond between the NH group of proline and the Trp nitrogen. As such a geometry shift would not be allowed by the constraints of backbone in real protein, we repeated the optimization with a constraint on the position of proline N atom. When we carried out this restricted geometry optimization, the geometry did not change significantly and also the interaction energy of the optimized structure was very close to that of the initial structure. It should also be noted that the DFT-SAPT

interaction energy components for the optimized structure are very similar to those of the crystal structure derived arrangement. Also, the optimization of cyclopentane–tryptophan and leucine–tryptophan complexes, where an H-bond cannot be formed, did not show any notable geometric rearrangement.

These results indicate that the structure of the interacting side chains taken from the Trp-cage experimental structure are geometrically very close to the potential energy minimum restrained by the backbone. Since this result is the same in all three models of Pro–Trp stacked-like arrangement, the interaction energy surface of tryptophan is probably quite shallow, in the sense that various interacting partners, regardless of their character (generally aliphatic moieties), can be bound with a similar strength and without precise requirements of their interacting positions.

We are aware of the fact that results obtained for the gas phase cannot be generally applied to side-chain interactions in protein environment, because the effect of solvent and protein environment cannot be neglected. Nevertheless, the fact that the crystal geometries of pair-wise structures containing proline were practically not changed upon restrained "gas-phase" optimization indicates a strong correlation between gas phase and environment behavior of the system in the particular case described here. A direct consequence of the described results is the important role of interaction energy/enthalpy compared to the entropy part of the interaction.

The results presented in this work indicate that contact between proline and an aromatic residue can result in very stable interactions, able to stabilize structural elements in proteins and protein complexes. These findings present possible explanation for why proline often plays not only a structural role in a protein, but is also often found to be a crucial element in various molecular interaction and recognition phenomena. The relatively strong interactions of this residue, together with its exceptional rigidity, might account, for example, for the formation of spatial clutches in proline rich regions at the protein surface, recognizable by various biomolecules and responsible for cell signaling properties. Study of complexes representing interactions of signal peptides (or Pro rich regions) with protein receptor confirmed the hidden importance of the aromatic–Pro interactions in context of a living cell.

Acknowledgements

This work was supported by grants No. 203/06/1727 and 203/05/H001 from the Czech Science Foundation, grant No. A400550510 from the Grant Agency of the Academy of Sciences of the Czech Republic and grant No. LC512 from the Ministry of Education, Youth and Sports (MSMT) of the Czech Republic. It was also part of research projects No. Z40550506, AV0Z50520701 and MSM6198959216. K.E.R. and P.H. acknowledge support from Praemium Academiae, Academy of Sciences of the Czech Republic dedicated to P.H. in 2007. Some of the calculations were performed using supercomputers at the Pacific Northwest National Laboratory, USA.

References

1 R. B. Greaves and J. Warwicker, *BMC Struct. Biol.*, 2007, **7**.
2 K. Ishii and K. Marumo, *Resource Geol.*, 2002, **52**, 135–146.
3 R. Jaenicke, *J. Biotechnol.*, 2000, **79**, 193–203.
4 R. H. Yun, A. Anderson and J. Hermans, *Proteins: Struct., Funct., Genet.*, 1991, **10**, 219–228.
5 S. C. Li, N. K. Goto, K. A. Williams and C. M. Deber, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 6676–6681.
6 I. Nilsson, A. Saaf, P. Whitley, G. Gafvelin, C. Waller and G. von Heijne, *J. Mol. Biol.*, 1998, **284**, 1165–1175.
7 S. Arnold, A. Curtiss, D. H. Dean and O. Alzate, *FEBS Lett.*, 2001, **490**, 70–74.
8 J. Y. Suh, Y. T. Lee, C. B. Park, K. H. Lee, S. C. Kim and B. S. Choi, *Eur. J. Biochem.*, 1999, **266**, 665–674.
9 T. G. Consler, O. Tsolas and H. R. Kaback, *Biochemistry*, 1991, **30**, 1291–1298.
10 P. Chakrabarti and S. Chakrabarti, *J. Mol. Biol.*, 1998, **284**, 867–873.
11 K. E. Riley, G. L. Cui and K. M. Merz, *J. Phys. Chem. B*, 2007, **111**, 5700–5707.
12 M. L. Waters, *Biopolymers*, 2004, **76**, 435–445.
13 S. Aravinda, N. Shamala, C. Das, A. Sriranjini, I. L. Karle and P. Balaram, *J. Am. Chem. Soc.*, 2003, **125**, 5308–5315.
14 S. K. Burley and G. A. Petsko, *Science*, 1985, **229**, 23–28.
15 S. S. C. Li, *Biochem. J.*, 2005, **390**, 641–653.
16 A. Palencia, E. S. Cobos, P. L. Mateo, J. C. Martinez and I. Luque, *J. Mol. Biol.*, 2004, **336**, 527–537.
17 M. Meiyappan, G. Birrane and J. A. A. Ladias, *J. Mol. Biol.*, 2007, **372**, 970–980.
18 B. Kemper, *Toxicol. Appl. Pharmacol.*, 2004, **199**, 305–315.
19 F. Toledo, C. J. Lee, K. A. Krummel, L. W. Rodewald, C. W. Liu and G. M. Wahl, *Mol. Cell. Biol.*, 2007, **27**, 1425–1432.
20 A. V. Morozov, K. M. S. Misura, K. Tsemekhman and D. Baker, *J. Phys. Chem. B*, 2004, **108**, 8489–8496.
21 L. Bendova, P. Hobza and J. Vondrasek, *Proteins: Struct., Funct., Bioinform.*, 2008, **72**, 402–413.
22 L. L. Qiu, S. A. Pabit, A. E. Roitberg and S. J. Hagen, *J. Am. Chem. Soc.*, 2002, **124**, 12952–12953.
23 L. J. Ball, R. Kuhne, J. Schneider-Mergener and H. Oschkinat, *Angew. Chem., Int. Ed.*, 2005, **44**, 2852–2869.
24 P. Hobza, R. Zahradnik and K. Muller-Dethlefs, *Collection Czech. Chem. Commun.*, 2006, **71**, 443–531.
25 P. Jurecka, J. Cerny, P. Hobza and D. R. Salahub, *J. Comput. Chem.*, 2007, **28**, 555–569.
26 P. Jurecka, J. Sponer, J. Cerny and P. Hobza, *Phys.Chem. Chem. Phys.*, 2006, **8**, 1985–1993.
27 K. E. Riley and P. Hobza, *J. Phys. Chem. A*, 2007, **111**, 8257–8263.
28 P. Jurecka, J. Cerny, P. Hobza and D. R. Salahub, *J. Comput. Chem.*, 2007, **28**, 555–569.
29 A. Halkier, T. Helgaker, P. Jorgensen, W. Klopper, H. Koch, J. Olsen and A. K. Wilson, *Chem. Phys. Lett.*, 1998, **286** 243–252.
30 P. Jurecka and P. Hobza, *Chem. Phys. Lett.*, 2002, **365**, 89–94.
31 R. Podeszwa, R. Bukowski and K. Szalewicz, *J. Chem. Theory Comput.*, 2006, **2**, 400–412.
32 A. Hesselmann and G. Jansen, *Chem. Phys. Lett.*, 2002, **362**, 319–325.
33 A. Hesselmann and G. Jansen, *Chem. Phys. Lett.*, 2002, **357**, 464–470.
34 A. Misquitta, R. Podeszwa, B. Jeziorski and K. Szalewicz, *J. Chem. Phys.*, 2005, **123**, 214103.
35 A. Hesselmann, G. Jansen and M. Schutz, *J. Chem. Phys.*, 2005, **122**, 14103.
36 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, Jr, J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas,

D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03 (Revision C.02)*, 2004.

37 H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby and M. a. o. Schutz, *MOLPRO, Version 2006.1*, 2006.

38 R. Ahlrichs, M. Bar, M. Haser, H. Horn and C. Kolmel, *Chem. Phys. Lett.*, 1989, **162**, 165–169.

39 P. Jurecka, J. Cerny, P. Hobza and D. R. Salahub, *J. Comput. Chem.*, 2007, **28**, 555–569.

40 M. O. Sinnokrot and C. D. Sherrill, *J. Am. Chem. Soc.*, 2004, **126**, 7690–7697.

41 P. Jurecka, J. Sponer, J. Cerny and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993.