On the Mechanism of Asymmetric Allylation of Aldehydes with Allyltrichlorosilanes Catalyzed by Quinox, a Chiral Isoquinoline N-Oxide

Journal:	*Journal of the American Chemical Society*
Manuscript ID:	ja-2007-11338q
Manuscript Type:	Article
Date Submitted by the Author:	02-Jan-2008

Complete List of Authors:	Malkov, Andrei; University of Glasgow, Chemistry Ramírez-López, Pedro; Consejo Superior de Investigaciones Cientificas, Productos Naturales Biedermannova, Lada; Academy of Sciences of Czech Republic, Inst of Organic Chemistry & Biochemistry Rulisek, Lubomir; Academy of Sciences of Czech Republic, Inst of Organic Chemistry & Biochemistry Dufkova, Lenka; University of Glasgow, Chemistry Kotora, Martin; Faculty of Science, Charles University, Organic and Nuclear Chemistry Zhu, Fujiang; University of Glasgow, Chemistry Kocovsky, Pavel; University of Glasgow, Chemistry

On the Mechanism of Asymmetric Allylation of Aldehydes with Allyltrichlorosilanes Catalyzed by Quinox, a Chiral Isoquinoline *N*-Oxide

Andrei V. Malkov,*[,†] Pedro Ramírez-López,[†,‡] Lada Biedermannová (née Bendová),[†,§] Lubomír Rulíšek,*[,¶] Lenka Dufková,[†,⊥] Martin Kotora,[Δ] Fujiang Zhu,[†] and Pavel Kočovský*[,†]

Department of Chemistry, WestChem, Joseph Black Building, University of Glasgow, Glasgow G12 8QQ, Scotland, UK, Institute of Organic Chemistry and Biochemistry and Gilead Sciences Research Center & IOCB, Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, 16610 Prague 6, Czech Republic, and Department of Organic Chemistry, Charles University, Hlavova 2030, 12840, Prague 2, Czech Republic

E-mail: amalkov@chem.gla.ac.uk; lubos@uochb.cas.cz; pavelk@chem.gla.ac.uk

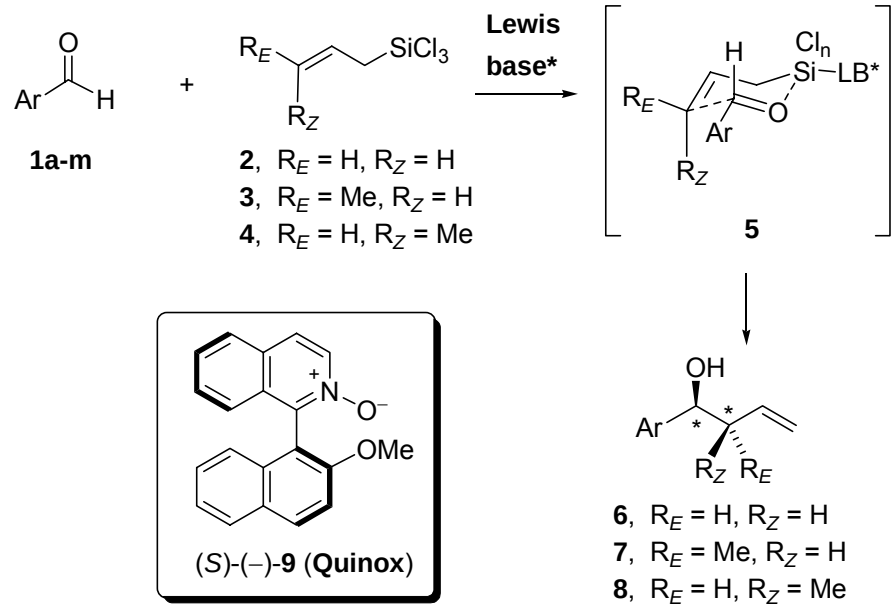
Abstract: Allylation of aromatic aldehydes **1a-m** with allyl- and crotyl-trichlorosilanes **2-4**, catalyzed by the chiral *N*-oxide Quinox (**9**), has been found to exhibit a significant dependence on the electronics of the aldehyde, with *p*-(trifluomethyl)benzaldehyde **1g** and its *p*-methoxy counterpart **1h** affording the corresponding homoallylic alcohols **6g,h** of 96 and 16% ee, respectively, at -40 °C. The kinetic and computational data indicate that the reaction is likely to proceed via an associative pathway involving neutral, octahedral silicon complex **12** with only one molecule of the catalyst involved in the rate- and selectivity-determining step. The crotylation with (*E*) and (*Z*)-crotyltrichlorosilanes **3** and **4** is highly diastereoselective, suggesting the chair-like transition state **5**, which is supported by computational data. High-level quantum chemical calculations further suggest that attractive aromatic interactions between the catalyst **9** and the aldehyde **1** play an important role in the enantiodifferentiation and that the dramatic drop in enantioselectivity, observed with the electron-rich aldehyde **1h**, originates from narrowing the energy gap between the (*R*)- and (*S*)-reaction channels in the associative mechanism (**12**). Overall, a good agreement between the theoretically predicted enantioselectivities for **1a** and **1h** and the experimental data allowed to understand the specific aspects of the reaction mechanism.

Introduction

Allyltrialkylsilanes[1] and stannanes[1,2] react readily at low temperature with aldehydes upon activation of the carbonyl group by a Lewis acid.[1-3] By contrast, allyltrichlorosilane (**2**) and related reagents require activation with a Lewis base that coordinates to the silicon (Scheme 1).[1,4-7] Provided the Lewis base dissociates from the silicon at a sufficient rate, it can act as a catalyst (rather than a stoichiometric reagent). Typical Lewis bases that promote the allylation reaction are the common dipolar aprotic solvents, such as DMF,[4,8] DMSO,[4,5] and HMPA,[5,9] and other substances possessing a strongly Lewis basic oxygen, such as formamides,[4,10,11] urea derivatives,[12] catecholates,[6a] and their chiral modifications;[1] the allylation with allyltrifluorosilane can also be promoted by fluoride.[6a,c,e] Analogous allyl boranes and boronates are more reactive than their Si and Sn counterparts and do not require an activator.[13] Similarly, the chelates, generated from allyltrichlorosilane and a stoichiometric amount of an amino

alcohol, such as pseudoephedrine, do not require further activation and have been shown to produce the corresponding homoallylic alcohols **6** with high enantioselectivity.[14]

Scheme 1. Allylation of aldehydes **1** with allyl and crotyl trichlorosilanes **2**-**4**.[a]

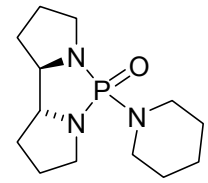


[a]For **a**-**m**, see Table 1.

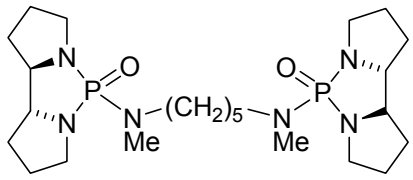
Asymmetric allylation of aldehydes **1** with allyl- and crotyltrichlorosilanes **2**-**4**, catalyzed by chiral Lewis bases (Scheme 1), in particular phosphoramides **10** and **11**,[5,16,17] bipyridine *N,N*-bisoxides (e.g., **12**),[18-20] and terpyridine *N,N,N*-trisoxides[21] (Chart 1) has evolved into an efficient method for the synthesis of enantiomerically enriched homoallylic alcohols **6**-**8** (Scheme 1).[1,7] In general, the reaction displays an excellent diastereocontrol in the case of *trans*- and *cis*-crotylsilanes (**3**/**4**), suggesting the cyclic transition state (TS) **5**. In a detailed investigation into the mechanism of allylation catalyzed by chiral phosphoramides, Denmark has found that the reaction follows a first order kinetics in both silane and aldehyde and a second order in the monodentate phosphoramide promoter (**10**).[16a,22,23] Hence, not surprisingly, bidentate Lewis bases, such as bisphosphoramides (**11**)[16a,22] and bipyridine *N,N*'-dioxides (**12**),[18,19] proved to be superior in reactivity and selectivity compared to monodentate catalysts.[7]

In parallel, we have introduced a family of chiral pyridine *N*-monooxides, such as **9** and **13**, acting as remarkably efficient monodentate catalysts in the allylation reaction (Scheme 1) with up to 98% ee.[24,25] In a preliminary communication,[25] we have demonstrated that, in spite of the structural similarity across the whole series, Quinox (**9**) displayed an unusual reactivity pattern, for which a mechanistic rationale is presented herein.

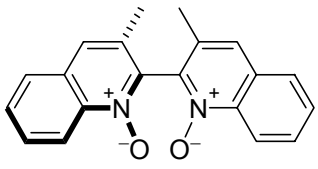
Chart 1. Selected Lewis basic catalysts for allylation of aldehydes with **2a**.



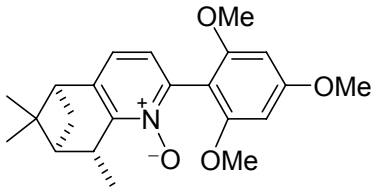
10



11



12



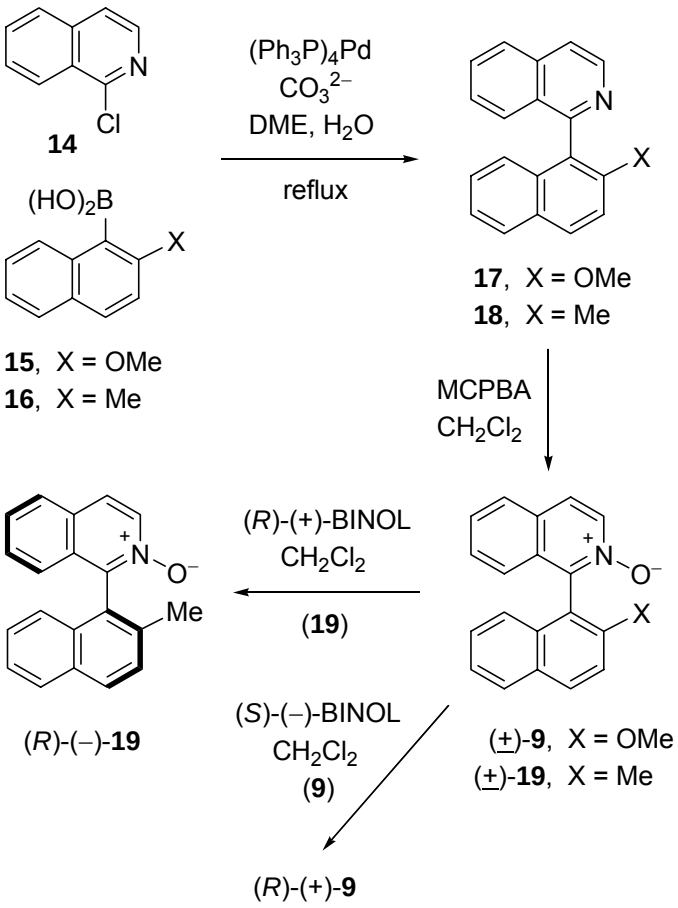
(+)-**13** (**METHOX**)

Results and Discussion

In the past few years, we have developed a series of chiral pyridine *N*-oxide organocatalysts for the enantioselective allylation of aromatic aldehydes **1** with allyltrichlorosilane **2-4** (Scheme 1). Thus, for instance, the terpene-derived Methox (**13**) and its congeners have been shown by us to work generally very well and to tolerate a wide range of electron-poor and electron-rich benzaldehydes, exhibiting little dependence of the reaction rate and enantioselectivity on the nature of the Ar group in **1** (96 and 93% ee for 4-$MeOC_6H_4CHO$ and 4-$CF_3C_6H_4CHO$, respectively, with **13** as catalyst).[24d] However, in contrast to most of the catalysts reported to date, Quinox (**9**), another member of the *N*-oxide family, exhibited a significant preference towards electron-poor aldehydes (96% ee with 4-$CF_3C_6H_4CHO$), whereas electron-rich substrates reacted much slower and were characterized by dramatically lower selectivities (16% ee for 4-$MeOC_6H_4CHO$!).[25] This intriguing behavior suggests that two fundamentally different mechanisms operate in these reactions as a function of the catalyst structure, which prompted us to embark on a detailed mechanistic study.

Synthesis of Quinox (9) and its Methyl Analogue (19). The synthesis of Quinox **9**, briefly described in our preliminary communication (Scheme 2),[25,26] was inspired by the initial steps toward QUINAP.[27] Thus, the Suzuki-Miyaura coupling of 1-chloro-isoquinoline (**14**)[27b] with boronic acid **15**,[27b,28] carried out in refluxing DME in the presence of Cs_2CO_3 and $(Ph_3P)_4Pd$ (3 mol%) overnight, afforded the biaryl derivative **17**[27] (95%), whose treatment with *m*-chloroperoxybenzoic acid provided the racemic *N*-oxide (±)-**9** (99%). The latter racemate was resolved by co-crystallization with (*S*)-(–)-2,2'-dihydroxy-1,1'-binaphthyl (BINOL),[29] which gave the crystalline material containing (*S*)-(–)-BINOL and (+)-**9** in a 1:1 ratio, while (–)-**9** remained in the solution. This co-crystallization, followed by a chromatographic separation of (+)-**9** from (*S*)-BINOL, furnished pure (+)-**9** of 98% ee (as revealed by chiral HPLC) in 89% isolated yield.[26] The absolute configuration of **9** was found to be (*R*)-(+)-**9** by crystallographic analysis of the molecular crystal of (*S*)-(–)-BINOL•(+)-**9** (Figure 1, **A**).[25]

Scheme 2. Synthesis of Quinox and its deoxy analogue.



To asses the electronic role of the methoxy group in Quinox (**9**), we have now synthesized its methyl analogue **19** (Scheme 2). The synthesis was carried out in the same way as that of Quinox: the Suzuki-Miyaura coupling of 1-chloroisoquinoline (**14**)[27b] with boronic acid **16**, prepared from 1-bromo-2-methyl naphthalene by lithiation with *n*-BuLi in THF, followed by a reaction with $(MeO)_3B$ and subsequent hydrolysis of the intermediate boronate ester with HCl.[25,27b,30] The coupling itself was carried out by refluxing a mixture of **14** and **16** in DME with aqueous K_2CO_3 and $(Ph_3P)_4Pd$ as catalyst (3 mol%) for 48 h to afford **18** (71%). Oxidation of the latter derivative with MCPBA gave rise to the racemic *N*-oxide **19** (81%), which was resolved by co-crystallization with (*R*)-(+)-BINOL from CH_2Cl_2.[29] The crystals contained the molecular compound (*R*)-BINOL•(*R*)-**19**, as revealed by single crystal X-ray analysis (Figure 1, **B**). Flash chromatography of the latter material gave (*R*)-(-)-**19** (33%; ≥99.5% ee) and the recovered (*R*)-(+)-BINOL. The mother liquor provided the enantiomerically enriched (*S*)-**19**, which was purified by co-crystallization with (*S*)-(-)-BINOL to produce (*S*)-(+)-**19** (19%; ≥99.8% ee).

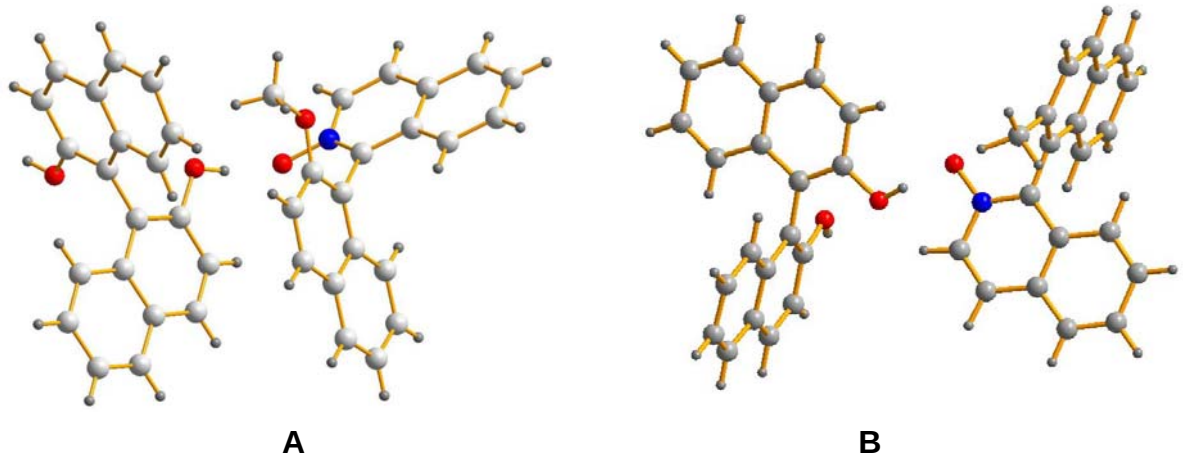


Figure 1. **A:** Disposition of the molecules in the crystal containing (*R*)-(+)-**9** (right) and (*S*)-(-)-BINOL (left) as shown by X-ray crystallography. **B:** Molecules in the crystal of (*R*)-(-)-**19** (right) with (*R*)-(+)-BINOL (left). Both diagrams illustrate the N–O....H–O hydrogen bonding.

Allylation of Aldehydes 1a-m Catalyzed by Quinox (9). Addition of allyltrichlorosilane (**2**) to benzaldehyde (**1a**) (Scheme 1), carried out in the presence of (*R*)-(+)-**9** (5 mol%) at -40 °C for 12 h, produced (*R*)-(+)-**6a** of 87% ee (Table 1, entry 1). Lowering the catalyst load to 1 mol% slowed the reaction but had no effect on the enantioselectivity (entry 2).

Table 1. The allylation of aldehydes **1a-m** with allylsilane **2**, catalyzed by **9** and **19** (Scheme 1).[a]

entry	catalyst	aldehyde	Ar	solvent	time (h)	t (°C)	yield (%)[b]	ee (%)[c,d]
1	(*R*)-(+)-**9**	**1a**	Ph	CH_2Cl_2	12	-40	68	87 (*R*)
2	(*R*)-(+)-**9**	**1a**	Ph	CH_2Cl_2[e]	12	-40	55	87 (*R*)
3	(*R*)-(+)-**9**	**1a**	Ph	CH_2Cl_2	12	-20	76	84 (*R*)
4	(*R*)-(+)-**9**	**1a**	Ph	CH_2Cl_2	2	18	40	65 (*R*)
5	(*R*)-(+)-**9**	**1a**	Ph	MeCN	12	-40	60	72 (*R*)
6	(*R*)-(+)-**9**	**1a**	Ph	$CHCl_3$	0.5	-40	79	63 (*R*)
7	(*R*)-(+)-**9**	**1a**	Ph	Toluene	12	-40	<5	61 (*R*)
8	(*R*)-(+)-**9**	**1b**	4-NO_2-C_6H_4	CH_2Cl_2	2	-40	73	89 (*R*)
9	(*R*)-(+)-**9**	**1c**	4-Cl-C_6H_4	CH_2Cl_2	2	-40	65	93 (*R*)
10	(*R*)-(+)-**9**	**1d**	2-Cl-C_6H_4	CH_2Cl_2	2	-40	72	91 (*R*)
11	(*R*)-(+)-**9**	**1e**	3-Cl-C_6H_4	CH_2Cl_2	2	-40	49	95 (*R*)
12	(*R*)-(+)-**9**	**1f**	4-F-C_6H_4	CH_2Cl_2	2	-40	79	91 (*R*)
13	(*R*)-(+)-**9**	**1g**	4-CF_3-C_6H_4	CH_2Cl_2	2	-40	85	96 (*R*)
14	(*R*)-(+)-**9**	**1g**	4-CF_3-C_6H_4	CH_2Cl_2	2	-20	88	91 (*R*)
15	(*R*)-(+)-**9**	**1g**	4-CF_3-C_6H_4	CH_2Cl_2	2	0	65	89 (*R*)
16	(*S*)-(–)-**9**	**1h**	4-MeO-C_6H_4	CH_2Cl_2	18	-40	70	16 (*S*)[f]
17	(*S*)-(–)-**9**	**1h**	4-MeO-C_6H_4	CH_2Cl_2	18	-20	45	72 (*S*)[f]
18	(*S*)-(–)-**9**	**1h**	4-MeO-C_6H_4	CH_2Cl_2	18	0	10	45 (*S*)[f]
19	(*R*)-(+)-**9**	**1i**	2-MeO-C_6H_4	CH_2Cl_2	12	-40	73	37 (*R*)
20	(*R*)-(+)-**9**	**1j**	3-MeO-C_6H_4	CH_2Cl_2	12	-40	40	80 (*R*)
21	(*R*)-(+)-**9**	**1k**	3,5-Me_2-C_6H_3	CH_2Cl_2	16	-40	68	81 (*R*)
22	(*R*)-(+)-**9**	**1l**	PhCH=CH	CH_2Cl_2	12	-40	86	51 (*R*)
23	(*R*)-(+)-**9**	**1m**	PhCH=C(Me)	CH_2Cl_2	12	-40	71	55 (*R*)
24	(*R*)-(–)-**19**[g]	**1a**	Ph	CH_2Cl_2	12	-40	19	64 (*S*)
25	(*R*)-(–)-**19**[g]	**1g**	4-CF_3-C_6H_4	CH_2Cl_2	12	-40	26	71 (*S*)
26	(*R*)-(–)-**19**[g]	**1h**	4-MeO-C_6H_4	CH_2Cl_2	12	-40	7	7 (*S*)

[a]The reaction was carried out at 0.4 mmol scale with 1.1 equiv of **2**, in the presence of **9** (5 mol%, ≥98% ee) or **19** (5 mol%, ≥99% ee) as catalyst, and (*i*-Pr)$_2$NEt (1 equiv) as base, unless stated otherwise.

[b]Isolated yield. [c]Determined by chiral HPLC or GC. [d]The configuration of the products **6** was established by the comparison of their optical rotations (measured in $CHCl_3$) and their GC and HPLC retention times with the literature data and with the behavior of authentic samples (ref. **24,25**). [e]With 1 mol% of the catalyst. [f]Note that this experiment was carried out with (*S*)-(-)-**9**. [g]Note that due to the change of substituent preferences in the CIP nomenclature, (*R*)-**19** actually corresponds to (*S*)-**9**.

The solvent effect was investigated briefly. In MeCN, the reaction was slower and rather less enantioselective than that in CH_2Cl_2 (compare entries 1 and 5). Further decrease of enantioselectivity was observed for $CHCl_3$ (63% ee, entry 6) but the reaction proved to be much faster (30 min at -40 ^oC). On the other hand, in toluene, the reaction became very sluggish (entry 7), although moderate asymmetric induction (61% ee) was still observed.

Introduction of electron-withdrawing substituents into the aromatic ring of the aldehyde resulted in a notable increase in both reactivity and enantioselectivity. Thus, *p*-nitrobenzaldehyde **1b** afforded the corresponding product in better yield and with slightly higher ee than benzaldehyde (compare entries 1 and 8); the halo derivatives **1c-f** gave >90% ee (entries 9-12). The highest conversion and enantioselectivity (96% ee) was attained with the *p*-trifluoromethyl derivative **1g** (entry 13).

By contrast, the electron-rich *p*-methoxybenzaldehyde (**1h**) afforded an almost racemic product (entry 16) and its *o*-isomer **1i** also showed low selectivity (entry 19); deceleration was observed in both cases. On the other hand, *m*-methoxybenzaldehyde **1j** (entry 20), where the methoxy group serves as a weak acceptor, again showed a good level of selectivity (80% ee), although the reaction was still quite slow. 3,5-Dimethylbenzaldehyde **1k** exhibited good reactivity and enantioslelectivity (entry 21), while cinnamyl derivatives **1l** and **1m** produced modest results (entries 22 and 23).

Investigation of the effects of temperature on the enantioselectivity of allylation catalyzed by Quinox (**9**) revealed that benzaldehyde **1a** and its electron poor *p*-trifluoromethyl derivative **1g** behaved in an ordinary fashion, showing a slight erosion of the ee at increased temperatures (entries 1, 3, and 4 for **1a** and 13-15 for **1g**). At the same time, the electron-rich *p*-methoxybenzaldehyde **1h** exhibited a significant temperature dependence (entries 16-18), known as isoinversion effect,[31] where the highest value of 72% ee was achieved at -20 ^oC. Furthermore, the increase in temperature had a significant deceleration effect in the case of **1h** (entries 16-18), reflected in a decreased yield (and an increased amount of the unreacted aldehyde recovered).

Allylation Catalyzed by Methyl--Quinox (19). The reactivity of the *N*-oxide **19**, a methyl analogue of Quinox, was explored with the aid of three representative aldehydes **1a,g,h** (Table 1, entries 24-26). Compared to Quinox, the reactions turned out to be much slower and the enantioselectivities lower. Nevertheless, the same trend was observed, i.e., the electron-poor **1g** was the most reactive of the series, while the worst results were obtained with the electron-rich aldehyde **1h**. This behavior clearly shows that the methoxy group in Quinox (**9**) not only serves as steric element preventing the rotation about the chiral axis but also must play an additional, presumably electronic, role.

Crotylation Catalyzed by Quinox (9). The crotylation reaction was found by us to be highly diastereoselective,[25] which is consistent with the cyclic chair-like transition state **5**. Previously, we reported that addition of crotyltrichlorosilane **3** (*trans*/*cis* ratio 87:13) to benzaldehyde (**1a**), catalyzed by Quinox (**9**), produced a ~70:30 mixture of **7a** and **8a**, suggesting a partial loss of diastereoselectivity.[25] We have now repeated these experiments with the isomerically pure **3** (*trans*/*cis* ≥98:2)[32] and **4** (*trans*/*cis* ≤2:98)[32] and found that the reaction, in fact, was highly diastereoselective for both **1a** and **1g** (Table 2, entries 1, 2, 4, and 5);[16b,33]

only with the electron-rich **1h** (entries 3 and 6) it was less stereo-convergent. Significantly, *cis*-crotylsilane **4** exhibited higher rate of allylation compared to the *trans* isomer **3**, which represents a reversal of the trend commonly observed with other *N*-oxide catalysts,[24] in particular Methox.[24d] Based on these results, the seeming loss of diastereocontrol, reported by us previously for **3** with the 87:13 *trans*/*cis* ratio,[25] can now be attributed to a higher reactivity of the minor *cis* isomer **4**, thus showing kinetic preference of Quinox (**9**) towards *cis*-crotylsilane (in the excess of the geometrically impure reagent). Erosion of diastereoselectivity in the case of **1h** was accompanied by low enantioselectivities (entries 3 and 6). Furthermore, the absolute configuration at C(1) of the minor product proved to be opposite to that of the major diastereoisomer (entries 3 and 6), demonstrating an inversion of the facial attack at the C=O bond from *si* to *re*, which may originate from participation of a twisted boat-like transition structure.[34]

Table 2. The allylation of aldehydes **1a,g,h** with allylsilanes **3** ($E/Z \geq 98:2$) and **4** ($E/Z \leq 2:98$) catalyzed by (*S*)-(–)-**9** (Scheme 1).[a]

entry	aldehyde	Ar	silane	yield %[b]	**7** : **8** (*anti* : *syn*)	%ee (configuration)[c,d] *anti*, *syn*
1	**1g**	4-CF_3-C_6H_4	**3**	75	96 : 4	92 (1*S*,2*S*), 90 (1*S*,2*R*)
2	**1a**	Ph	**3**	65	95 : 5	66 (1*S*,2*S*), 77 (1*S*,2*R*)
3	**1h**	4-MeO-C_6H_4	**3**	40	83 : 17	25 (1*S*,2*S*), 12 (1*R*,2*S*)
4	**1g**	4-CF_3-C_6H_4	**4**	85	1 : 99	nd, 94 (1*S*,2*R*)
5	**1a**	Ph	**4**	78	1 : 99	69 (1*S*,2*S*), 79 (1*S*,2*R*)
6	**1h**	4-MeO-C_6H_4	**4**	50	4 : 96	40 (1*R*,2*R*), 60 (1*S*,2*R*)

[a]The reaction was carried out at 0.4 mmol scale with 1.1 equiv of **3**/**4**, in the presence of (*S*)-(−)-**9** (5 mol%, 98% ee) as catalyst and (*i*-Pr)$_2$NEt (1 equiv) as base in CH_2Cl_2 at -40 °C for 24 h. [b]Isolated yield. [c]Determined by chiral GC. [d]The absolute configuration of **7**/**8** was determined by comparison of their optical rotations (measured in $CHCl_3$) and their GC retention times with the literature data (ref. 16b,33) and with the behavior of authentic samples.

A probe into the possible non-linear effect in the allylation of **1a** with **2**, catalyzed by (*R*)-(+)-**9** of 29%, 50%, and 75% ee, respectively, has demonstrated a fully linear relationship between the enantiopurity of the catalyst and the product. In these experiments, the resulting alcohol **6a** was of 27%, 45%, and 71% ee, respectively.

Kinetic Experiments. To shed light on the mechanism, a series of kinetic experiments were carried out. The order in each component was determined by using the method of initial rates. A sampling technique was employed, in which the aliquots taken at certain intervals were analyzed by GC to monitor the product formation. Although the high sensitivity of the reaction to traces of moisture complicated the data collection, acceptable reproducibility was achieved up to the 20% conversion. The reaction followed first-order kinetics in aldehyde (1.12 for **1g**), whereas the order in silane exhibited an unusual behavior, irrespective of the nature of the aldehyde. The lg/lg plot of the initial rates (v_o) vs the [silane]/[catalyst] ratio (Figure 2) consists of two distinct areas. At the silane/catalyst ratio up to 15:1 (lg value 1.18), the order in silane is unity (1.02 for both **1g** and **1h**), whereas above that point the reaction slows down and the order in silane turns negative. At low silane-to-catalyst ratios, the kinetic data obtained for aldehyde **1g** and silane **2** mirror those reported by Denmark for phosphoramides.[16a,22] However, the order in Quinox (**9**), determined at 10-25 mol% loadings (with 1 equiv of **1g** and 1.5 equiv of **2**), was found close to unity (0.82), which contrast with the second order observed for monodentate phosphoramides. Taking into account the reversible nature of coordination of Quinox (**9**) to silane **2** and absence of a non-linear effect (in the case of **9**), combined with the observations that enantioselectivity is not affected by catalyst loading in the range of 1-10 mol%,[6] it appears very

likely that only one molecule of the catalyst is involved in the rate and selectivity determining step.

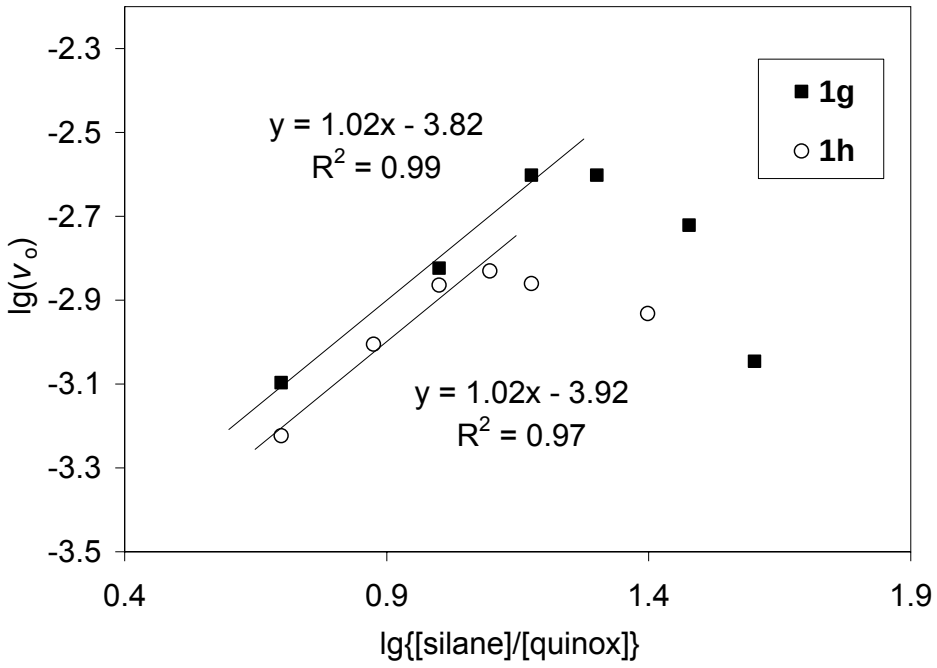


Figure 2. Reaction order in **2**.

The rate limiting step (RLS) was elucidated by using Hammett correlation. The plot in Figure 3 shows that electron-withdrawing substituents accelerated the reaction (ρ = 1.18), consistent with the rate determining C-C bond formation rather than pre-equilibration.[35] This ρ value also suggests that the reaction center in the TS is non-ionic, weakly polar,[36] implying that Lewis acidity of the silicon in **5** is just sufficient to bind aromatic aldehydes but without affecting their electrophilicity. At higher σ values, deviation from linearity was observed,[37,38] indicating a gradual shift in the RLS towards pre-coordination. These data were corroborated by the measurement of the secondary isotope effect (SIE), namely by comparing the relative reactivities of ArCD=O vs ArCH=O. In the case of **1h**, the SIE value of 0.70 indicates a late transition state, closely resembling the sp^3 hybridized product. The sp^2-character in the TS gradually increased to 0.81 (for **1a**) and further to 0.86 (for **1g**), confirming a shift towards pre-coordination RLS, which is reflected in the first order kinetics in **1g**.

Based on the kinetic data, the following mechanistic scenario can be proposed (Scheme 3). Allyltrichlorosilane (**2**) and Quinox (**9**) reversibly form the reactive complex **10**, which is deactivated by a large excess of the silane, presumably via binding a second molecule of **2** to generate a catalytically inactive complex, such as **11** (*vide infra*). The negative order in silane **2** at higher concentrations arises from the need to release the extra silane molecule to return to **10**. The next reaction step, binding the aldehyde, can proceed via associative (**12**) or dissociative (**13**) mechanisms.[22a] Note that the octahedral complex **12** is more crowded than the trigonal bipyramidal **13**. Therefore, participation of the tightly packed transitions state **12** may lead to higher selectivities. The low ρ value, together with the observed higher reaction rates and better selectivities attained in non-polar solvents, are consistent with the reaction pathway via the neutral, octahedral transition state **12**. The observed maximum on the plot of enantioselectivity vs reaction temperature in the case of *p*-methoxybenzaldehyde (the inversion effect) is indicative of either a temperature-dependent shift in the RLS (from pre-equilibrium to C-C-bond

formation) or competition between different mechanism;[31] however at this point the kinetic data do not provide a definite answer.

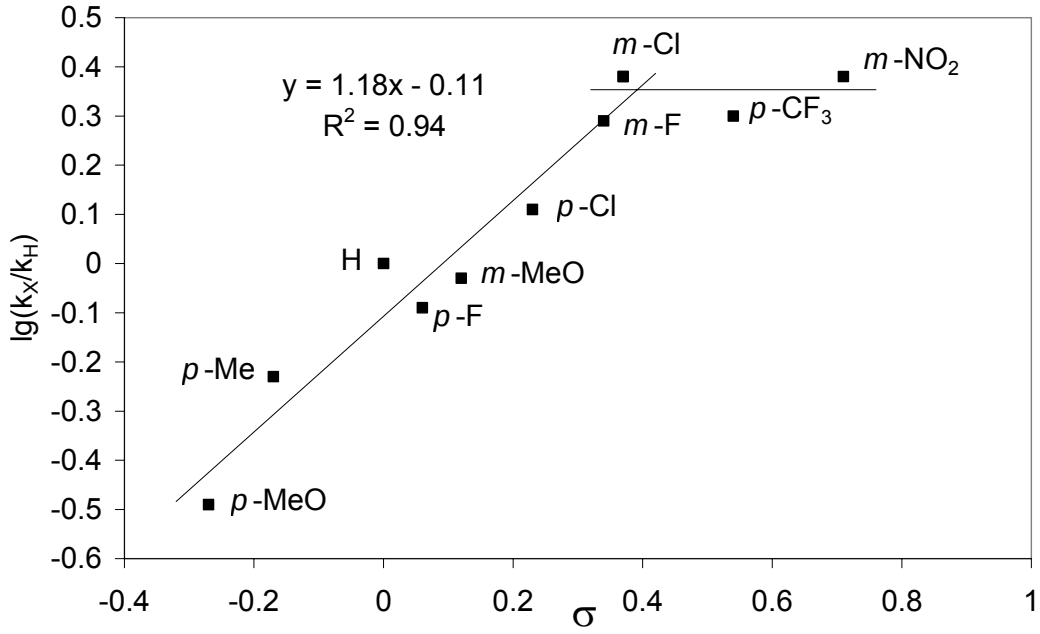
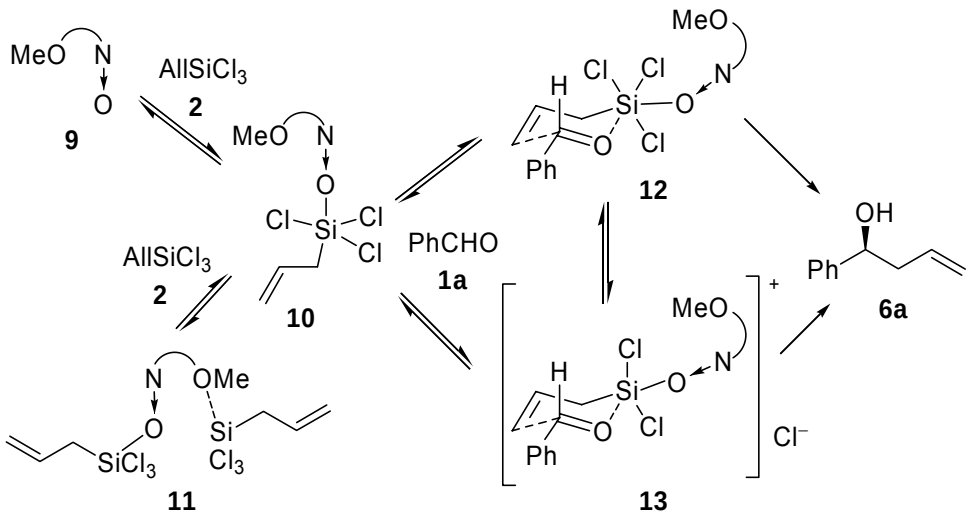


Figure 3. Hammett plot for the allylation of **1** with **2** catalyzed by **9**.

Scheme 3. Allylation of benzaldehyde **1a** catalyzed by Quinox (**9**).



To obtain further insight into the deactivation of Quinox (**9**) by a large excess of allyltrichlorosilane (**2**), for which coordination of Si to the MeO group could be suggested (**11** in Scheme 3), we have investigated the methyl catalyst **19**, where the latter interaction cannot exist. Indeed, in the allylation of *p*-trifluomethylbenzaldehyde (**1g**), increasing the silane/catalyst ratio from 10:1 to 50:1 (while the concentration of aldehyde **1g** was kept constant) resulted in the proportional (~4-fold) increase in the reaction rate; i.e., no catalyst deactivation was taking place. However, it is pertinent to note that at the silane/catalyst ratio 10:1, Quinox (**9**) reacted nearly 3.5 times faster than its methyl analogue **19**. Hence, the MeO group of Quinox can be

regarded as the key feature to attain sufficient reaction rates and high enantioselectivities (with the exception of electron-rich aldehydes, such as **1h**).

Quantum Chemical Calculations. To get a better insight into the mechanism of the allylation catalyzed by Quinox (**9**), a computational analysis of the reaction coordinate and the transition states (TS) was carried out using benzaldehydes **1a** and **1h** as model substrates.

Computational Details. All density functional theory (DFT) calculations were carried out using the program Turbomole 5.7.[39] The Perdew-Burke-Ernzerhof (PBE) functional[40] has been used throughout. The calculations were expedited by expanding the Coulomb integrals in an auxiliary basis set, the resolution-of-identity (RI-J) approximation.[41,42] All geometry optimizations were carried out using the 6-31G(d) basis set,[43] whereas the single point energies were recomputed using the TZVP basis set.[44] To account for the description of dispersion forces (which are not described by the currently used DFT functionals), the empirical dispersion parameters were added to both the energy and gradients during geometry optimization and the calculations of the single point energies [DFT(+D) method].[45,46] Here, default parameters[46] were used, except for the C_6 parameters for Si, where the values for P were used, and Cl, where we used the values of Neumann and Perrin.[47] This method has been shown to essentially yield the data with ~1 kcal.mol^{-1} acuracy in comparison with the reference CCSD(T) values.[46]

To account for solvation effects, the conductor-like screening model (COSMO)[48,49] was employed with the dielectric constant corresponding to acetonitrile ($\varepsilon_r = 36.6$). Gibbs free energy was then calculated as the sum of these contributions (equation 1):

$$G = E_{\mathrm{el(+D)}} + G_{\mathrm{solv}} + E_{\mathrm{ZPE}} - RT \ln(q_{\mathrm{trans}}\, q_{\mathrm{rot}}\, q_{\mathrm{vib}}) \qquad (1)$$

where $\Delta E_{\mathrm{el(+D)}}$ is the *in vacuo* energy of the system [at the RI-PBE(+D)/TZVPP level, with the geometry optimized at the RI-PBE(+D)/6-31G(d) level (vide supra)], ΔG_{solv} is the solvation free energy [at the RI-PBE(+D)/6-31G(d) level] and the $E_{\mathrm{ZPE}} - RT \ln(q_{\mathrm{trans}}\, q_{\mathrm{rot}}\, q_{\mathrm{vib}})$ term is the zero-point energy, thermal corrections to the Gibbs free energy, and entropic term (obtained from a frequency calculation with the same method and software as for the geometry optimizations at the RI-PBE/6-31G(d) level, 298 K and 1 atm pressure, using an ideal-gas approximation[50]).[51]

Transition state (TS) optimizations were performed as follows: first, the TS initial geometry was obtained by the restricted geometry optimization on the interpolated reaction coordinate, defined by Si(1)-C(2) and C(4)-C(5) distances (Figure 4), which were assumed to attain the discrete values between the reactant and product with all the remaining internal coordinates optimized. The maximum on this approximate reaction coordinate was then taken as the initial guess for the saddle point optimization along the eigenvector corresponding to the vibration with the imaginary frequency. Due to the lack of the analytic second derivatives in both DFT(+D) and COSMO solvation methods, the approximate method was used to obtain the TS's in solution: (1) *in vacuo* optimization of transition state according to the above procedure; (2) constraining Si(1)-C(2) and C(4)-C(5) distances to the values corresponding to the TS acquired in step 1 and optimizing all other degrees of freedom by the DFT(+D)/COSMO method. The difference between the DFT(+D)/COSMO energy at the *in vacuo* TS geometry and the relaxed geometry is denoted as the solvent relaxation term and is regarded to represent an accurate approximation to the true transition state in solution, calculated with the DFT(+D) method..

Finally, for benchmarking purposes on the model uncatalyzed reaction, the CCSD(T) method was used[52] with the aug-cc-pVDZ basis set.[53] These calculations were carried out using the MOLPRO program.[54,55]

Results of Computational Analysis. Approximately twenty structural variants for each of the reaction mechanisms using (*R*)-(+)-**9** were examined. The results covering the most stable

structures are summarized in Table 3; the structures of the reactant and product complexes and transition states for the (*R*)- and (*S*)-reaction channels in the associative mechanism are shown in Figure 4. According to this analysis, the reaction commences with the formation of a transient reactant complex (RC). For the associative mechanism (**12**), this complex is higher in energy than the assembly of the isolated reactants and the catalyst by 23.4/23.8 kcal.mol^{-1} (Table 3, entries 1 and 2) and for the dissociative mechanism (**13**) by 19.3/19.8 kcal.mol^{-1} (entries 3 and 4); note the greater stability of the RC with cationic trigonal bipyramidal structure formed in the dissociative pathway. The calculated transition state barriers are $\Delta G^{\neq}_{assoc}$ = 23.9 (*R*) and 25.9 (*S*) kcal.mol^{-1}; and $\Delta G^{\neq}_{dissoc}$ = 26.0 (*R*) and 24.5 (*S*) kcal.mol^{-1} for the associative and dissociative mechanism, respectively.[56] Significantly, a narrow energy gap between the RC and TS in the associative mechanism correlates well with the experimentally observed borderline position of the RLS between pre-coordination of the substrate aldehyde and C-C bond formation. The overall reaction thermodynamics, after the catalyst has dissociated from the product complex (PC), is ΔG = -10.3 kcal.mol^{-1}. The associative mechanism predicts the formation of (*R*)-**6a** in 97% ee (at 233.15 K) for an enantiopure catalyst, which is in a good agreement with the experimental value of 87% ee (Table 1, entry 1) attained with the catalyst of 98% ee. Significantly, the dissociative route favors the formation of the opposite, i.e., (*S*)-enantiomer.[57] Assuming that both mechanisms operated concurrently, the calculations would predict the formation of (*R*)-**6a** in 53% ee.

Table 3. The calculated thermochemical data for the reaction of **1a** with **2** catalyzed by (*R*)-(+)-Quinox (**9**).[a]

entry	mechanism	product configuration	ΔG, reactant complex	$\Delta G^{\neq}$, transition state	ΔG, product complex
1	*associative*	*R*	23.4	23.9	1.1
2	*associative*	*S*	23.8	25.9	1.4
3	*dissociative*	*R*	19.3	26.0	1.1
4	*dissociative*	*S*	19.8	24.5	1.4

[a] All values are in kcal.mol^{-1}

In the case of *p*-methoxybenzaldehyde **1h**, the associative mechanism is even more favored with the TS barriers $\Delta G^{\neq}_{assoc}$ = 26.1 (*R*) and 26.9 (*S*) kcal.mol^{-1}, while the corresponding values for the dissociative pathway are $\Delta G^{\neq}_{dissoc}$ = 29.3 (*R*) and 29.7 (*S*) kcal.mol^{-1}. Significantly, the energy gap between the (*R*)- and (*S*)-reaction channels in the associative mechanism for **1h** is reduced to 0.8 kcal.mol^{-1} and predicts the formation of (*R*)-**6h** in 62% ee (at 273.15 K; in a very good agreement with 45% ee observed experimentally at this temperature). Note that a higher TS barrier for the formation of (*R*)-**6h** compared to (*R*)-**6a** correlates well with the experimentally observed higher reaction rates of **1a** vs **1h**.

The calculations further show that the methoxynaphthalene unit of the Quinox (**9**) and benzaldehyde (**1a**) in the TS are arranged in a parallel orientation, which makes the TS for Quinox lower in energy than that for the simple pyridine *N*-oxide by 3-5 kcal.mol^{-1}.[58] This value is comparable with the strength of the aromatic interaction between benzaldehyde and the catalyst expected in the TS arrangement. Moreover, the difference in the attractive π-π interactions between the (*R*)- and (*S*)-transition states also constitutes the largest contribution to the enantiodifferentiation since the calculated difference in the dispersion energy stabilizations between the (*R*)- and (*S*)-enantiomers amounts to 1.1 kcal.mol^{-1}.[59]

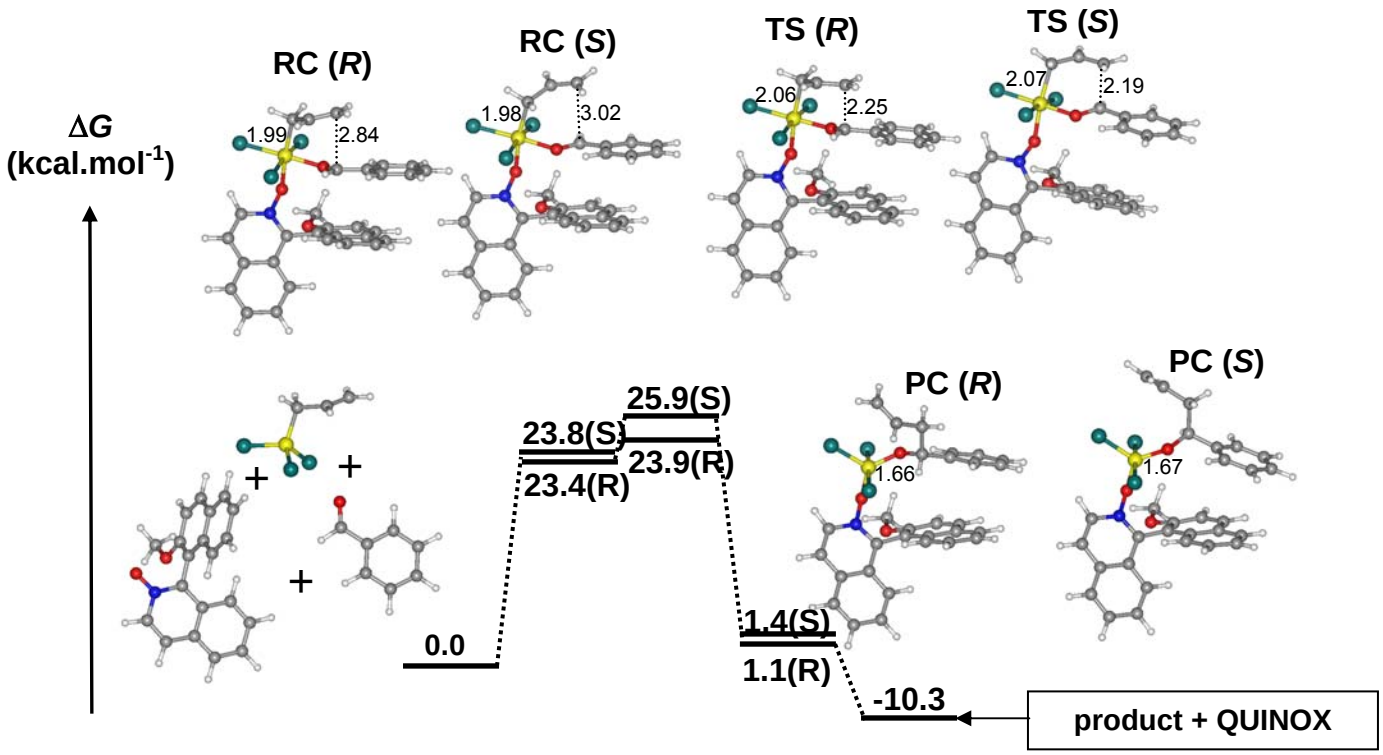


Figure 4. The equilibrium geometries of the most stable reactant complexes (RC), transition states (TS), and product complexes (PC) along the reaction coordinate for the associative pathway of allylation of benzaldehyde (**1a**) catalyzed by (*R*)-(+)-Quinox (**9**). The calculated values for Δ*G* were obtained at the RI-PBE(+D)/TZVP//RI-PBE(+D)/6-31G(d) level; all distances are in Å.

Conclusions

The kinetic and computational data indicate that the allylation of aldehydes **1** with allyltrichlorosilanes **2-4**, catalyzed by Quinox (**9**), is likely to proceed via an associative pathway involving neutral, octahedral silicon complex **12**. In contrast to catalysis by chiral monodentate phosphoramides, such as **10**, only one molecule of Quinox (**9**) is involved in the rate- and selectivity-determining step. The reaction presumably proceeds via a cyclic chair-like[60] transition state **5** and is characterized by high enantio- and diasteroselectivity. Attractive aromatic interactions between the catalyst and the substrate aldehyde appear to play an important role in the enantiodifferentiation. A drop in enantioselectivity observed in the case of electron-rich substrates apparently originates from narrowing the energy gap between the (*R*)- and (*S*)-reaction channels in the associative mechanism (**12**).

Acknowledgement: The paper is dedicated to Professor Victor Sniečkus on the occasion of his 70th birthday. We thank the EPSRC for grant No. GR/T27051/01, the Ministry of Education and Science of Spain for a fellowship to P. R.-L., the Ministry of Education of the Czech Republic (MŠMT ČR) for grant No. LC512, the European Socrates-Erasmus exchange program for scholarships to L. B. and L. D., Dr. Alfred Bader for an additional support and Dr. DeLiang Long for the X-ray structure determination. We also thank Profs. Guy C. Lloyd Jones and Pavel Hobza for helpful discussions.

Supporting Information Available: General experimental methods, representative allylation and crotylation procedures, NMR spectra of new compounds, all kinetic data and the equilibrium structures and energies of all the computationally studied species. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

† University of Glasgow.
‡ Current address: Departamento de Química Orgánica, Universidad Complutense de Madrid, Madrid 28040, Spain.
§ Exchange student from the Institute of Organic Chemistry and Biochemistry, AVČR.
¶ Institute of Organic Chemistry and Biochemistry, AVČR.
⊥ Exchange student from Charles University.
Δ Charles University.

(1) For leading reviews, see: (a) Denmark, S. E.; Stavenger, R. A. *Acc. Chem. Res.* **2000**, *33*, 432. (b) Denmark, S. E.; Fu, J. *Chem. Commun.* **2003**, 167; (c) Denmark, S. E.; Fu, J. *Chem. Rev.* **2003**, *103*, 2763. (d) Kennedy, J. W. J.; Hall, D. G. *Angew. Chem., Int. Ed.* **2003**, *42*, 4732.

(2) (a) Marshall, J. A. *Chem. Rev.* **1996**, *96*, 31. (b) Marshall, J. A. In *Organometallics in Synthesis* (M. Schlosser, Ed.); J. Wiley and Sons: Chichester, 2002, p. 399.

(3) For the coordination of the C=O group to Lewis acids (η^1 vs η^2 fashion), see, e.g.: (a) Shambayati, S.; Crowe, W, E.; Schreiber, S. L. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 256. (b) Lenges, C. P.; Brookhart, M.; White, P. S. *Angew. Chem., Int. Ed.* **1999**, *38*, 552.

(4) (a) Kobayashi, S.; Nishio, K. *Tetrahedron Lett.* **1993**, *34*, 3453. (b) Kobayashi, S.; Nishio, K. *J. Org. Chem.* **1994**, *59*, 6620. (c) Kobayashi, S.; Nishio, K. *Synthesis* **1994**, 457.

(5) Denmark, S. E.; Coe, D. M.; Pratt, N. E.; and Griedel, B. D. *J. Org. Chem.* **1994**, *59*, 6161.

(6) (a) Kira, M.; Sato, K.; Sakurai, H. *J. Am. Chem. Soc.* **1988**, *110*, 4599. (b) Kira, M.; Kobayashi, M.; Sakurai, H. *Tetrahedron Lett.* **1987**, *28*, 4081. (c) Sakurai, H. *Synlett* **1989**, 1. (d) Kira, M.; Sato, K.; Sakurai, H. *J. Am. Chem. Soc.* **1990**, *112*, 257. (e) Kira, M.; Zhang, L. C.; Kabuto, C.; Sakurai, H. *Organometallics* **1998**, *17*, 887.

(7) (a) Malkov, A. V.; Kočovský, P. *Eur. J. Org. Chem.* **2007**, 29. (b) Kočovský, P.; Malkov, A. V. In *Enantioselective Organocatalysis* (P. I. Dalko, Ed.); Wiley-VCH: Weinheim, 2007, p 255.

(8) Short, J. D.; Attenoux, S.; Berrisford, D. J. *Tetrahedron Lett*. **1997**, *38*, 2351.

(9) Wang, Z.; Xu, G.; Wang, D.; Pierce, M. E.; Confalone, P. N. *Tetrahedron Lett.* **2000**, *41*, 4523.

(10) (a) Iseki, K.; Mizuno, S.; Kuroki, Y.; Kobayashi, Y. *Tetrahedron Lett.* **1998**, *39*, 2767. (b) Iseki, K.; Mizuno, S.; Kuroki, Y.; Kobayashi, Y. *Tetrahedron* **1999**, *55*, 977. For analogous activation of Cl_3SiH by chiral tertiary amides derived from amino acids, see: (c) Iwasaki, F.; Omonura, O.; Mishima, K.; Kanematsu, T.; Maki, T.; Matsumura, Y. *Tetrahedron Lett*. **2001**, *42*, 2525. (d) Malkov, A. V.; Mariani, A.; MacDougall, K. N.; Kočovský, P. *Org. Lett.* **2004**, *6*, 2253.

(11) (a) Kobayashi, S.; Nishio, K. *Chem. Lett.* **1994**, 1773. (b) Kobayashi, S.; Nishio, K. *J. Org. Chem*. **1994**, *59*, 6620. (c) Kobayashi, S.; Nishio, K. *J. Org. Chem*. **1994**, *59*, 6620.

(12) Chataigner, I.; Piarulli, U.; Gennari, C. *Tetrahedron Lett.* **1999**, *40*, 3633.

(13) Smith, K. In *Organometallics in Synthesis*; Schlosser, M., Ed.; J. Wiley and Sons; Chichester, 2002, p. 514.

(14) (a) Wang, X.; Meng, Q.; Nation, A. J.; Leighton, J. L. *J. Am. Chem. Soc.* **2002**, *124*, 10672. (b) Kinnaird, J. W. A.; Ng, P. Y.; Kubota, K.; Wang, X.; Leighton, J. L. *J. Am. Chem. Soc* **2002**, *124*, 7920. (c) Kubota, K.; Leighton, J. L. *Angew. Chem. Int. Ed.* **2003**, *42*, 946.

(15) (a) Iseki, K.; Kuroki, Y.; Takahashi, M.; Kishimoto, S.; Kobayashi, Y. *Tetrahedron* **1997**, *53*, 3513. (b) Wadamoto, M.; Ozasa, N.; Yanagisawa, A.; Yamamoto, H. *J. Org. Chem.* **2003**, *63*, 5593. (c) Iseki, K.; Kuroki, Y.; Takahashi, M.; Kobayashi, Y. *Tetrahedron Lett.* **1996**, *37*, 5149.

(16) (a) Denmark, S. E.; Fu, J. *J. Am. Chem. Soc.* **2000**, *122*, 12021. (b) Denmark, S. E.; Fu, J. *J. Am. Chem. Soc.* **2001**, *123*, 9488.

(17) Hellwig, J.; Belser, T.; Müller, J. F. K. *Tetrahedron Lett.* **2001**, *42*, 5417.

(18) (a) Nakajima, M.; Saito, M.; Shiro, M.; Hashimoto, S. *J. Am. Chem. Soc.* **1998**, *120*, 6419. (b) Nakajima, M.; Saito, M.; Hashimoto, S. *Chem. Pharm. Bull.* **2000**, *48*, 306.

(19) (a) Shimada, T.; Kina, A.; Ikeda, S.; Hayashi, T. *Org. Lett.* **2002**, *4*, 2799. (b) Shimada, T.; Kina, A.; Hayashi, T. *J. Org. Chem.* **2003**, *68*, 6329. (c) Kina, A.; Shimada, T.; Hayashi, T. *Adv. Synth.. Catal.* **2004**, *346*, 1169.

(20) For reviews, see: (a) Chelucci, G.; Thummel, R. P. *Chem. Rev.* **2002**, *102*, 3129. (b) Malkov, A. V.; Kočovský, P. *Curr. Org. Chem.* **2003**, *7*, 1737. (c) Chelucci, G.; Murineddu, G.; Pinna, G. A. *Tetrahedron: Asymmetry* **2004**, *15*, 1373.

(21) Wong. W.-L.; Lee, C.-S.; Leung, H.-K.; Kwong, H.-L. *Org. Biomol. Chem.* **2004**, *2*, 1967.

(22) (a) Denmark, S. E.; Fu, J.; Coe, D. M.; Su, X.; Pratt, N. E.; Griedel, B. D. *J. Org. Chem*. **2006**, *71*, 1513. For mechanistic studies on the related aldol reaction, see: (b) Denmark, S.E., Bui, T., *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5439. (c) Denmark, S.E., Bui, T., *J. Org. Chem*. **2005**, *70*, 10393. (d) Denmark, S. E.; Pham, S. M.; Stavenger, R. A.; Su, X.; Wong, K.-T.; Nishigaichi, Y. *J Org. Chem*. **2006**, *71*, 3904. For mechanistic studies on Lewis-base catalyzed opening of epoxides with chlorosilanes, see: (e) Denmark, S. E.; Barsanti, P. A.; Beutner, G. L.; Wilson, T. W. *Adv. Synth. Catal.* **2007**, *349*, 567.

(23) In the allylation, unlike in the related aldol reaction (ref 2e), the minor, one catalyst manifold could not be characterized.(ref 22a).

(24) (a) Malkov, A. V.; Orsini, M.; Pernazza, D.; Muir, K. W.; Langer, V.; Meghani, P.; Kočovský, P. *Org. Lett.* **2002**, *4*, 1047. (b) Malkov, A. V.; Bell, M.; Orsini, M.; Pernazza, D.; Massa, A.; Herrmann, P.; Meghani, P.; Kočovský, P. *J. Org. Chem.* **2003**, *68*, 9659. (c) Malkov, A. V.; Bell, M.; Vassieu, M.; Bugatti, V.; Kočovský, P. *J. Mol. Catal. A* **2003**, *196*, 179. (d) Malkov, A. V.; Bell, M.; Castelluzzo, F.; Kočovský, P. *Org. Lett.* **2005**, *7*, 3219.

(25) Malkov, A. V.; Dufková, L.; Farrugia, L.; Kočovský, P. *Angew. Chem. Int. Ed.* **2003**, *42*, 3674.

(26) (+)-Quninox has been first prepared by Nakajima, including the same Suzuki-Miyaura coupling and resolution. However, the synthesis was mentioned as a footnote without revealing the conditions, and the absolute configuration was not determined: Nakajima, M.; Saito, M.; Uemura, M.; Hashimoto, S. *Tetrahedron Lett.* **2002**, *43*, 8827.

(27) (a) Brown, J. M.; Hulmes, D. I.; Layzell, T. P. *J. Chem. Soc., Chem. Commun.* **1993**, 1673. (b) Alcock, N. W.; Brown, J. M.; Hulmes, D. I. *Tetrahedron: Asymmetry* **1993**, *4*, 743.

(28) Vyskočil, Š.; Meca, L.; Tišlerová, I.; Císařová, I.; Polášek, M.; Harutyunyan, S. R.; Belokon, Y. N.; Stead, R. M. J.; Farrugia, L.; Lockhart, S. C.; Mitchell, W. L.; Kočovský, P. *Chem. Eur. J.* **2002**, *8*, 4633.

(29) For the previous use of BINOL as a resolving agent, see refs **18, 19a**, **26**, and: Nakajima, M.; Saito, M.; Hashimoto, S. *Tetrahedron: Asymmetry* **2002**, *13*, 2449.

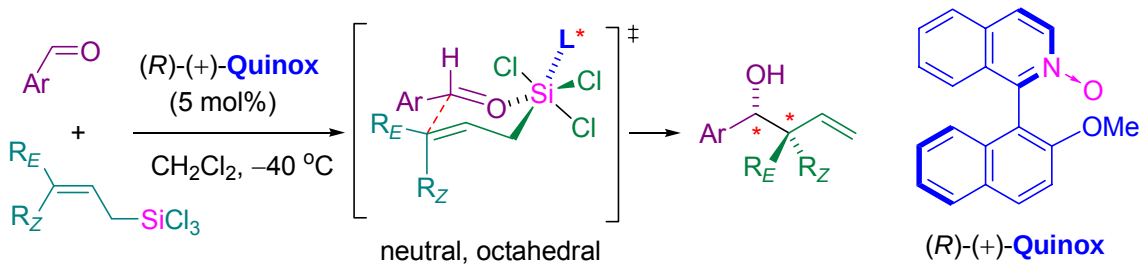
(30) Pathak, R.; Nhlapo, J. M.; Govender, S.; Michael, J. P.; van Otterlo, W. A. L.; de Koning, C. B. *Tetrahedron* **2006**, *62*, 2820.

(31) For a discussion of isoinversion principle, see: (a) Schmid, R.; Sapunov, V. N. *Non-formal Kinetics*, VCH, Weinheim, 1982. (b) Gypser, A.; Norrby, P.-O. *J. Chem. Soc., Perkin Trans. 2,* **1997**, 939.

(32) Iseki, K.; Kuroki, Y.; Takahashi, M.; Kishimoto, S.; Kobayashi, Y. *Tetrahedron* **1997**, *53*, 3513.
(33) (a) Hackman, B. M.; Lombardi, P. J.; Leighton, J. L. *Org. Lett.* **2004**, *6*, 4375. (b) McManus, H. A.; Cozzi, P. G.; Guiry, P. J. *Adv. Synth. Catal.* **2006**, *348*, 551.
(34) For a similar effect in the related aldol reaction, ref 22d.
(35) In the allylation, the C-C bond formation was assumed to be a rate-limiting step (ref 22a) but not confirmed by kinetic experiments.
(36) A similar value (ρ = 1.18) has been reported for the allylation of aldehydes with allysiliconates: Hosomi, A.; Kohra, S.; Ogata, K.; Yanagi, T.; Tominaga, Y. *J. Org. Chem.* **1990**, *55*, 2415.
(37) For an analogous nonlinear behavior at high σ values, observed for the silicon-directed aldol condensation, see: Myers, A. G.; Widdowson, K. L. Kukkola, P. J. *J. Am. Chem. Soc.* **1992**, *114*, 2765.
(38) The plot incorporating all the data points showed lower statistical significance of R^2 = 0.87.
(39) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989** *162*, 165.
(40) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.*, **1996**, *77*, 3865.
(41) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283.
(42) Eichkorn, K.; Weigen, F.; Treutler, O.; Ahlrichs, R. *Theor. Chim. Acta* **1997**, *97*, 119.
(43) Hehre, W. J.; Radom, L.; Schleyer, P. I. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley-Interscience: New York, 1986.
(44) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.
(45) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.
(46) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. *J. Comput. Chem.* **2007**, *28*, 555.
(47) Neumann, M. A.; Perrin, M. A. *J. Phys. Chem. B* **2005**, *109*, 15531.
(48) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.
(49) Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. *Phys. Chem. Chem. Phys.* **2000,** *2*, 2187.
(50) Jensen, F. *Introduction to Computational Chemistry*; John Wiley & Sons, 1999.
(51) The free energy calculated according to Eq. (1) is regarded as a good approximation to Δ*G* in diluted solution.
(52) (a) Čížek, J. *J. Chem. Phys.* **1966**, *45*, 4256-4266. (b) Čížek, J. *Adv. Chem. Phys.* **1969**, *14*, 35-89. (c) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.
(53) Woon, D. E.; Dunning, Jr., T. H. *J. Chem. Phys.* **1993**, *98*, 1358.
(54) MOLPRO is a package of *ab initio* programs written by Werner, H.-J. et al.
(55) Hampel, C.; Peterson, K.; Werner, H.-J. *Chem. Phys. Lett.* **1992**, *190*, 1.
(56) Lowering the barrier by 4.2 kcal.mol^{-1} for the (*R*) reaction channel (compared to the uncatalyzed reaction) corresponds to a ~10^3 increase of the reaction rate, which is in agreement with the experimental observations.
(57) There is an uncertainty in the translational entropy of the Cl^- ion, which has to be taken into account in the dissociative reaction mechanism; however, this factor is unlikely to affect the overall picture significantly.
(58) The TS barrier calculated for the reaction of benzaldehyde and allyltrichlorosilane catalyzed by pyridine *N*-oxide is $\Delta G^{\neq}$ = 27.5 kcal.mol^{-1}.
(59) Decomposition of the energy differences in the free energy barriers in the associative mechanism between the (*R*) and (*S*) enantiomers (in kcal.mol^{-1}) are as follows: in vacuo energies ΔE_{gp} = 0.3; solvation free energies ΔG_{solv} = -0.7; in vacuo entropic terms $\Delta(-T\Delta S)_{gp}$ = -0.5; and dispersion energy stabilizations ΔE_{disp} = -1.1.

(60) According to the calculations carried out for **1a** and **1h**, the chair conformation is preferred over the twisted boat for both (*R*)- and (*S*)-channels by about 1.5-2.0 kcal.mol^{-1}.

TOC entry


Ar
O
+
R_E
R_Z
$SiCl_3$
(*R*)-(+)-**Quinox**
(5 mol%)
CH_2Cl_2, –40 °C
H
Ar
O
Si
Cl
Cl
Cl
L*
‡
R_E
R_Z
neutral, octahedral
OH
Ar
*
*
R_E
R_Z
N
O
OMe
(*R*)-(+)-**Quinox**

Asymmetric Reduction

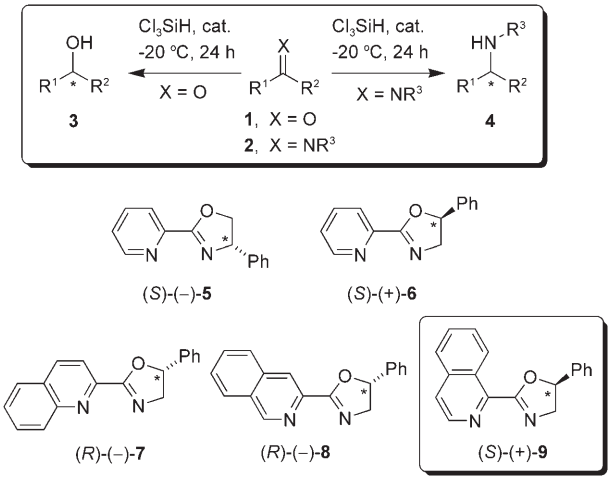
DOI: 10.1002/anie.200503941

Remote Chiral Induction in the Organocatalytic Hydrosilylation of Aromatic Ketones and Ketimines**

Andrei V. Malkov, Angus J. P. Stewart Liddon, Pedro Ramírez-López, Lada Bendová, David Haigh, and Pavel Kočovský**

Dedicated to Professor Henri Kagan on the occasion of his 75th birthday

Asymmetric hydrogenation, hydroboration, and hydrosilylation are the most frequently used catalytic methods for the reduction of prochiral ketones **1** and imines **2** (Scheme 1).[1]



Scheme 1. Catalytic reduction of ketones and ketimines; for R^1–R^3, see Tables 1 and 2.

[*] Dr. A. V. Malkov, A. J. P. S. Liddon, Dr. P. Ramírez-López, L. Bendová, Prof. Dr. P. Kočovský
Department of Chemistry, WestChem
Joseph Black Building
University of Glasgow, Glasgow G12 8QQ (UK)
Fax: (+44) 141-33-488
E-mail: amalkov@chem.gla.ac.uk
p.kocovsky@chem.gla.ac.uk
Dr. D. Haigh
Department of Medicinal Chemistry
Metabolic and Viral Centre of Excellence for Drug Discovery
GlaxoSmithKline Pharmaceuticals
Medicines Research Centre
Stevenage, Hertfordshire, SG1 2NY, (UK)

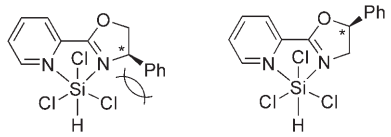
[**] We thank the GSK and EPSRC for an industrial CASE awards to A.J.P.S.L., the Ministry of Education and Science of Spain for a postdoctoral fellowship to P.R.-L., the Socrates Exchange program, and Dr. Alfred Bader and the University of Glasgow for an additional support.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

Although asymmetric hydrogenation remains favored by industry in general, it is not free of problems, namely, those associated with metal leaching, high pressure, and the cost of the catalyst and its regeneration. Stoichiometric borane reduction, catalyzed by chiral oxazaborolidine, avoids most of these problems and offers high levels of enantioselection,[1] but its cost is prohibitive for large-scale industrial application. The recently developed reduction of imines, which uses the Hantsch dihydropyridine as a stoichiometric reducing agent and a chiral Brønsted acid as an organocatalyst,[2] is also tainted by the cost implications. Transition-metal-catalyzed hydrosilylation[1,3,4] on the other hand relies on much cheaper silanes as reducing agents but shares the problem of metal leaching with hydrogenation. Furthermore, the current methods usually perform well with either ketones or imines, but rarely with both classes.[4] The alternative metal-free methods[2,5] are rare and considerably less effective;[6] nevertheless, a promising development has recently been reported,[7] which relies on Cl_3SiH, an inexpensive and easy-to-handle reducing agent.[8] Herein, we report an organocatalytic hydrosilylation applicable to both ketones and ketimines.

As part of our program focusing on the activation of organosilicon reagents by Lewis bases,[9] we examined 2-pyridyloxazolines of type **5**, derived from (*S*)-phenylglycinol, as potential organocatalysts in the hydrosilylation of aromatic ketones with Cl_3SiH (Scheme 1, **1**→**3**). Initial attempts to reduce acetophenone (**1a**) in the presence of **5** (20 mol%, $CHCl_3$, −20°C, 24 h) resulted in a rather low conversion into alcohol **3a** (Table 1, entry 1). By contrast, the isomeric oxazoline **6**, obtained from (*S*)-mandelic acid, proved much more reactive and furnished **3a** in 78% *ee* (entry 2).[10] Reduction of a series of aromatic ketones **1b–e**, catalyzed by **6**, proceeded in a similar fashion, with 70–80% *ee* (entries 3–6). On the other hand, reduction of the non-aromatic ketone **1f** resulted in the formation of a racemic product (entry 7). Other solvents, such as toluene, diethyl ether, and THF, proved inferior.

The lower reactivity of **5** relative to **6** can be conjectured to originate from an increase in the steric radius of the extra-coordinate silicon moiety by coordination to the ligand. Thus, the coordination of Cl_3SiH to **5** is impaired by the adjacent phenyl group (Scheme 2, **A**), whereas this adverse effect is absent in chelate **B**, derived from **6**.[11]



Scheme 2. Coordination of Cl_3SiH to (2-pyridyl)oxazolines.

To improve the efficacy of the catalyst and to shed more light on the mechanism, related oxazolines were prepared with the quinoline **7** or isoquinoline **8** and **9** fragments in place of the pyridine moiety (Scheme 1). No reduction was observed with **7** ($CHCl_3$, −20°C; Table 1, entry 8), presumably because of the steric constraints imposed by the quinoline moiety, thus mirroring the behavior of **5**. The 3-isoquinoline catalyst **8** was approximately as efficient as the parent (pyridyl)oxazoline **6** (entry 9; compare with entry 2). A real improvement over a range of substrates (**1a–e, g, h**) was attained with the 1-isoquinolyl catalyst **9** (entries 10–16), which allowed the catalyst loading to be lowered to 10 mol% and provided by far the best enantioselectivities in the metal-free hydrosilylation of ketones to date (≤94% *ee*).

Importantly, catalyst **9** was also found to exhibit high efficiency in the reduction of *N*-aryl imines **2a–f** (Table 2). Therefore, it can be regarded as the first organocatalyst that

Table 1: Reduction of ketones (**1**→**3**) with trichlorosilane catalyzed by **5–9**.[a]

Entry	Catalyst (mol%)	Ketone	R^1, R^2	Yield [%]	*ee* [%][b,c]
1	(*S*)-**5** (20)	**1a**	Ph, Me	29[d]	66[g]
2	(*S*)-**6** (20)	**1a**	Ph, Me	85[d]	78
3	(*S*)-**6** (20)	**1b**	2-MeO-C_6H_4, Me	100[d]	77
4	(*S*)-**6** (20)	**1c**	2-F-C_6H_4, Me	30[d]	70
5	(*S*)-**6** (20)	**1d**	4-Me-C_6H_4, Me	100[d]	80
6	(*S*)-**6** (20)	**1e**	Ph, Et	91[d]	80
7	(*S*)-**6** (20)	**1f**	*c*-C_6H_{11}, Me	67[d,e]	0
8	(*R*)-**7** (20)	**1a**	Ph, Me	0	–
9	(*R*)-**8** (20)	**1a**	Ph, Me	41[e,f]	73[g]
10	(*S*)-**9** (10)	**1a**	Ph, Me	85[f]	84
11	(*S*)-**9** (10)	**1b**	2-MeO-C_6H_4, Me	50[f]	87
12	(*S*)-**9** (10)	**1c**	2-F-C_6H_4, Me	35[f]	70
13	(*S*)-**9** (10)	**1d**	4-Me-C_6H_4, Me	90[f]	85
14	(*S*)-**9** (10)	**1e**	Ph, Et	55[f]	86
15	(*S*)-**9** (10)	**1g**	2-naphth, Me	93[f]	94
16	(*S*)-**9** (10)	**1h**	6-Me-2-naphth, Me	93[f]	92

[a] The reactions were carried out on a 0.4-mmol scale with Cl_3SiH (2.1 equiv) in $CHCl_3$ in the presence of the catalyst at −20°C for 24 h. [b] Determined by chiral GC or HPLC. [c] All products **3** were *R* configured (unless stated otherwise), as revealed by comparison of their optical rotation with the literature data. [d] Conversion determined by GC of the reaction mixture after a standard work up. [e] The reaction was carried out in CH_2Cl_2. [f] Yield of the isolated product (shown to be pure by ^1H NMR spectroscopic analysis). [g] The product was *S* configured. naphth = napthyl.

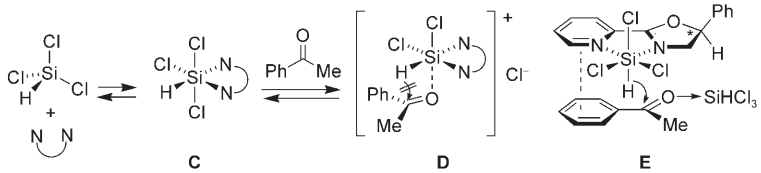
Table 2: Reduction of ketimines (**2–4**) with trichlorosilane catalyzed by **9**.[a]

Entry	Imine	R^1, R^2, R^3	Yield [%][b]	*ee* [%][c,d]
1	**2a**	Ph, Me, Ph	65	87
2	**2b**	Ph, Me, PMP[e]	60	85
3	**2c**	2-naphth, Me, Ph	67	87 (98[f])
4	**2d**	2-naphth, Me, PMP[e]	67	86
5	**2e**	4-MeO-C_6H_4, Me, PMP[e]	51	87
6	**2f**	4-CF_3-C_6H_4, Me, PMP[e]	65	87

[a] The reactions were carried out on a 0.4-mmol scale with $HSiCl_3$ (2.0 equiv) in $CHCl_3$ in the presence of catalyst **9** (20 mol%) at −20°C for 24 h. [b] Yield of the isolated product (product shown to be pure by ^1H NMR spectroscopic analysis). [c] Amines **4** were *R* configured, as revealed by comparison of their optical rotation and their HPLC retention times with the literature data. [d] Enantiomeric excess was determined by HPLC. [e] PMP = *p*-methoxyphenyl. [f] After a single recrystallization from methanol.

allows reduction of both ketones and ketimines with a practically useful enantioselectivity.

To rationalize the high selectivity of catalysts **6**, **8**, and **9**, the following issues have to be taken into account: 1) Table 1 reveals that the order of reactivity of the ketones displays a good correlation with their Brønsted basicity[12] (pK_{HB}): 2-MeO-C_6H_4COMe (1.34) > PhCOMe (1.11) > 2-F-C_6H_4COMe (0.90). If we accept that this trend can be extrapolated to Lewis basicity, the results would suggest that the carbonyl oxygen atom is coordinated to the weakly Lewis acidic Cl_3SiH in the transition state.[13] 2) Trichlorosilane does not reduce carbonyl or heterocarbonyl compounds unless activated by Lewis bases through the formation of extra-coordinate silicon species, such as **C**, in the case of bidentate activators (Scheme 3).[7] Coordination of the carbonyl group



Scheme 3. Mechanism of hydrosilylation.

(**D**) might be presumed (upon replacement of a chloride atom at the silicon center) but can be regarded as unproductive, since the corresponding four-membered transition state (TS) for the reduction is unlikely.[13,14] 3) A linear relationship between the enantiopurity of catalyst **6** and the product was observed for the reduction of acetophenone (**1a**→**3a**), thus suggesting that only one molecule of the catalyst is actively involved in the enantiodifferentiating event. 4) The role of the substituent on the oxazoline rings in **6**, **8**, and **9** can be attributed to the shielding of one of the faces of the catalyst by the Ph group.[15]

Accordingly, the following mechanistic picture can be drawn: the *N*,*N*-chelation of Cl_3SiH by the catalyst creates an activated hydrosilylating species, while another molecule of Cl_3SiH is likely to activate the ketone by coordination to the oxygen atom in the *E* fashion.[16] The ketone–Cl_3SiH complex will then approach the catalyst–Cl_3SiH complex from the less-hindered side (as dictated by the remote chiral center in the catalyst). Transition state **E**, which accommodates all these effects, is consistent with the experimentally observed *si*-facial selectivity of the reaction and with the features discussed herein. It can also be hypothesized that **E** will be stabilized by arene–arene interactions between the heteroaromatic systems of the catalyst and the substrate.[17]

The potential role of the latter π–π interactions in shaping the TS[17] was assessed by a computational analysis of the pyridine–acetophenone model complex (Figure 1).[18] The successive scans, according to the steering angle (θ), the distance between the planes (r_1), and the parallel displacement of the two molecules (r), were carried out (each scan starting at the geometry of interaction energy minimum of the previous scan). The lowest interaction energy found by the potential-energy surface scans was at $-6.3\ kcal\,mol^{-1}$. The

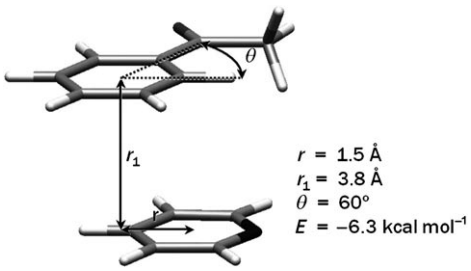


Figure 1. Arene–arene interactions of acetophenone (**1a**) with pyridine as the model TS **E**.

magnitude of this interaction lends credence to the proposed involvement of π stacking in the transition state **E**.

A similar mechanism can be envisaged for the reduction of imines. However, the coordination of Cl_3SiH to the nitrogen atom is unlikely, as this would create a very crowded environment. Hence, the R^3 group of imine **2** can be assumed to fill the space that the $SiHCl_3$ group fills in **E**. Experimental data support this hypothesis: Note that the electron-rich ketones react faster than their electron-poor counterparts (Table 1, entries 3 versus 4 and 11 versus 12), which is consistent with the carbonyl basicity (see below) and, therefore, its propensity to coordinate the Lewis acidic Cl_3SiH. By contrast, imines exhibit the opposite trend (Table 2, entries 5 versus 6).

In conclusion, we have developed a new, practical, metal-free protocol for the enantioselective reduction of aromatic ketones (≤94% *ee*) and ketimines (≤87% *ee*). The reaction is characterized by an unusual, long-ranging chiral induction. The enantiodifferentiation is presumed to be aided by aromatic interactions between the catalyst and the substrate.

Experimental Section

General procedure for the asymmetric reduction of ketones **1** and ketimines **2** with trichlorosilane: Trichlorosilane (86 μL, 0.84 mmol, 2.1 equiv) was slowly added dropwise to a solution of the catalyst (11.0 mg, 0.04 mmol or 21.9 mg, 0.08 mmol) and the corresponding ketone or imine (0.40 mmol, 1.0 equiv) in $CHCl_3$ (2 mL) at −20 °C. The reaction mixture was stirred at −20 °C for 24 h, after which time saturated aqueous $NaHCO_3$ (1 mL) was added to quench the reaction. The mixture was extracted with CH_2Cl_2 (3 × 10 mL) and the combined organic extracts were dried over $MgSO_4$. Concentration in vacuo followed by flash chromatography on silica gel (2 × 15 cm) with petroleum ether/ethyl acetate (20:1) or CH_2Cl_2 as the eluent afforded alcohols **3a**–**3h** or amines **4a**–**4f**.

Received: November 7, 2005
Published online: January 20, 2006

Keywords: arene–arene interactions · asymmetric catalysis · N ligands · organocatalysis · reduction

[1] E. N. Jacobsen, A. Pfaltz, H. Yamamoto, *Comprehensive Asymmetric Catalysis, Vol. I–III*, Springer, Heidelberg, **1999**.
[2] a) S. Singh, U. K. Batra, *Indian J. Chem.* **1989**, *27*, 1; b) M. Rueping, E. Sugiono, C. Azap, T. Theissmann, M. Bolte, *Org. Lett.* **2005**, *7*, 3781; c) S. Hoffmann, A. M. Seayad, B. List,

Angew. Chem. **2005**, *117*, 7590; *Angew. Chem. Int. Ed.* **2005**, *44*, 7424.

[3] O. Riant, N. Mostefai, J. Courmarcel, *Synthesis* **2004**, 2943.

[4] A highly efficient Cu-catalyzed hydrosilylation of ketones and ketimines has been developed: a) B. H. Lipshutz, K. Noson, W. Chrisman, A. Lower, *J. Am. Chem. Soc.* **2003**, *125*, 8779; b) B. H. Lipshutz, H. Shimizu, *Angew. Chem.* **2004**, *116*, 2278; *Angew. Chem. Int. Ed.* **2004**, *43*, 2228.

[5] a) P. I. Dalko, L. Moisan, *Angew. Chem.* **2001**, *113*, 3840; *Angew. Chem. Int. Ed.* **2001**, *40*, 3726; b) P. I. Dalko, L. Moisan, *Angew. Chem.* **2004**, *116*, 5248; *Angew. Chem. Int. Ed.* **2004**, *43*, 5138; c) E. R. Jarvo, S. J. Miller, *Tetrahedron* **2002**, *58*, 2481; d) A. Berkessel, H. Gröger, *Asymmetric Organocatalysis*, Wiley-VCH, Weinheim, **2005**.

[6] a) R. Schiffers, H. B. Kagan, *Synlett* **1997**, 1175; b) H. Nishikori, R. Yoshihara, A. Hosomi, *Synlett* **2003**, 561; c) F. J. LaRonde, M. A. Brook, *Tetrahedron Lett.* **1999**, *40*, 3507; d) F. J. LaRonde, M. A. Brook, *Inorg. Chim. Acta* **1999**, *296*, 208; e) M. D. Drew, N. J. Lawrence, W. Watson, S. A. Bowles, *Tetrahedron Lett.* **1997**, *38*, 5857.

[7] a) R. A. Benkeser, D. Snyder, *J. Organomet. Chem.* **1982**, *225*, 107; b) S. Kobayashi, M. Yasuda, I. Hachiya, *Chem. Lett.* **1996**, 407; c) M. A. Brook, *Silicon in Organic, Organometallic and Polymeric Chemistry*, Wiley, New York, **2000**, pp. 133 – 136; d) F. Iwasaki, O. Onomura, K. Mishima, T. Maki, Y. Matsumura, *Tetrahedron Lett.* **1999**, *40*, 7507; e) F. Iwasaki, O. Onomura, K. Mishima, T. Kanematsu, T. Maki, Y. Matsumura, *Tetrahedron Lett.* **2001**, *42*, 2525.

[8] A very efficient organocatalytic protocol for the hydrosilylation of ketimines (with $\leq 92\%$ *ee*) has been developed by us recently: a) A. V. Malkov, A. Mariani, K. MacDougall, P. Kočovský, *Org. Lett.* **2004**, *6*, 2253; b) A. V. Malkov, S. Stončius, K. N. MacDougall, A. Mariani, G. D. McGeoch, P. Kočovský, *Tetrahedron* **2006**, *62*, 264.

[9] a) A. V. Malkov, M. Orsini, D. Pernazza, K. W. Muir, V. Langer, P. Meghani, P. Kočovský, *Org. Lett.* **2002**, *4*, 1047; b) A. V. Malkov, M. Bell, M. Orsini, D. Pernazza, A. Massa, P. Herrmann, P. Meghani, P. Kočovský, *J. Org. Chem.* **2003**, *68*, 9659; c) A. V. Malkov, M. Bell, M. Vassieu, V. Bugatti, P. Kočovský, *J. Mol. Catal. A* **2003**, *196*, 179; d) A. V. Malkov, L. Dufková, L. Farrugia, P. Kočovský, *Angew. Chem.* **2003**, *115*, 3802; *Angew. Chem. Int. Ed.* **2003**, *42*, 3674; e) A. V. Malkov, M. Bell, F. Castelluzzo, P. Kočovský, *Org. Lett.* **2005**, *7*, 3219.

[10] Hydrosilylation of **1a**, catalyzed by Rh complexes of **5** and **6**, produced alcohol **2a** in 69% *ee* and 9% *ee*, respectively: H. Brunner, U. Obermann, *Chem. Ber.* **1989**, *122*, 499.

[11] Similar effects have been observed by us for substituted bipyridines, in which the increased bulk prevented coordination of $[Mo(CO)_6]$ and other metal complexes: A. V. Malkov, I. R. Baxendale, J. Fawcett, D. R. Russel, V. Langer, D. J. Mansfield, M. Valko, P. Kočovský, *Organometallics* **2001**, *20*, 673.

[12] F. Besseau, M. Luçon, C. Laurence, M. Berthelot, *J. Chem. Soc. Perkin Trans. 2* **1998**, 101.

[13] T. Kudo, T. Higashide, S. Ikedate, and H. Yamataka, *J. Org. Chem.* **2005**, *70*, 5157.

[14] Transfer hydrogenation and oxazaborolidine-catalyzed reduction of ketones with borane have been shown to proceed through a six-membered transition state: a) M. Yamakawa, H. Ito, R. Noyori, *J. Am. Chem. Soc.* **2000**, *122*, 1466; b) S. Itsuno in *Comprehensive Asymmetric Catalysis, Vol. 1*, Springer, Heidelberg, **1999**, pp. 290 – 315.

[15] Replacing the Ph in **6** with *t*Bu led to virtually identical results, which rules out arene–arene interactions of the substrate with the Ph in **6**.

[16] For a review on the carbonyl coordination (*E* vs *Z*; η^1 vs η^2), see: S. Shambayati, W. E. Crowe, S. L. Schreiber, *Angew. Chem.* **1990**, *102*, 273; *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 256.

[17] Arene–arene interactions have been suggested by us to account for the stereochemical outcome of allyl silane addition to aromatic aldehydes;[9d,e] Hayashi and co-workers arrived at a similar conclusion: a) T. Shimada, A. Kina, T. Hayashi, *J. Org. Chem.* **2003**, *68*, 6329; Birman et al. recently invoked arene – arene interactions to rationalize the kinetic resolution in the acylation of racemic alcohols: b) V. B. Birman, H. Jiang, *Org. Lett.* **2005**, *7*, 3445.

[18] The geometries of monomers were determined by the cc-pVTZ/RI-MP2 gradient optimization in the Turbomole v.5.6 package (R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Lett.* **1989**, *162*, 165); the optimized geometries of monomers were kept frozen throughout the scans. Interaction energies were computed with aug-cc-pVDZ basis set at RI-MP2 level in Turbomole v.5.6 and are BSSE-corrected. Only the face-to-face geometry was examined, thus keeping the planes of the monomers parallel throughout the scans.

Another role of Proline: Stabilization interactions in proteins concerning proline and residues of aromatic characters.

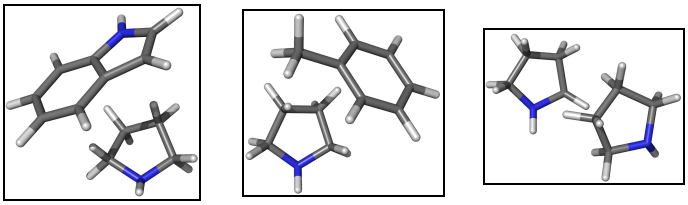
Lada Biedermannova, Kevin E Riley, Pavel Hobza* and Jiri Vondrasek*

Institute of Organic Chemistry and Biochemistry Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems
Flemingovo nám. 2, Prague, Czech Republic

* Dr Jiri Vondrasek
Institute of Organic Chemistry and Biochemistry Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems
Flemingovo nám. 2
166 10 Prague 6 (Czech Republic)
Fax: (+420) 220-410-320, tel +420 220-410-324
E-mail: jiri.vondrasek@uochb.cas.cz

* Prof. Pavel Hobza
Institute of Organic Chemistry and Biochemistry Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems
Flemingovo nám. 2
166 10 Prague 6 (Czech Republic)
Fax: (+420) 220-410-320, tel +420 220-410-311
E-mail: pavel.hobza@uochb.cas.cz

Graphical contents entry



Computational study of tryptophan-proline, proline-proline, and phenylalanine-proline complexes proved that cyclic arrangement of proline allows for the largest amount of dispersive contact with studied amino acids..

Summary

Tryptophan-proline, proline-proline, and phenylalanine-proline complexes derived from experimental structures were investigated by various quantum chemical procedures known to properly describe the London dispersion energy. The MP2/aug-cc-pVDZ procedure overestimates the binding for stacked complexes while the RI-DFT-D procedure parameterized to CCSD(T) benchmark binding energies provides reliable interaction energies for H-bonded and stacked structures. The tryptophan-proline stacking energies are very large (about 7 kcal/mol) and DFT-SAPT calculations demonstrate the important role of dispersion energy in these interactions. The large stabilization of the complex mentioned above is surprising since only one partner is aromatic. The role of the cyclic structure, dipole moment, and polarizability of proline is discussed and it is concluded that all these factors are responsible for the surprisingly large stabilization of the tryptophan-proline complex. The most critical feature determining the strong binding energy of this complex is the fact that the cyclic arrangement of proline allows for the largest amount of overlap or, in other words, dispersive contact, with tryptophan. Comparing the proline-tryptophan, proline-phenylalanine, and proline-proline complexes based on real protein structures we found that the former cluster in the stacked structure is the most stabilizing. This is in agreement with the fact that the Trp-Pro motifs are populated more frequently than the other ones in real proteins.

Introduction

Common opinion about L-Proline is 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 gives this residue exceptional conformational rigidity compared to other amino acids. Upon folding the residue thus loses less conformational entropy, which may account, for example, for its higher occurrence in the proteins of thermophilic organisms.

The structural role of proline is, however, still a matter of controversy, as it is very sensitive to the relative orientation 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. It can also act as an alpha helix disruptor in transmembrane alpha helices of various transmembrane proteins. On the other hand, a computational study by Yun et. al. 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 be a helix initiator. Li et. al 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. In some antimicrobial peptides proline is an essential structural element of helical motifs and helical kinks caused by proline plays an important role in the disruption of bacterial membranes.

Although proline can only serve as a hydrogen bond acceptor and not 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 ... O hydrogen bonds with the carbonyl group of the preceding helical turn. Proline also frequently occurs in turns, which may explain the intriguing fact that, in spite of having an aliphatic sidechain, it is very often solvent-exposed.

The role of proline in transmembrane transport proteins has also been extensively studied and it has been shown that proline is important for substrate binding and recognition. Proline can also work as a stabilizer in the folding process, during which it effectively inhibits protein aggregation. Experiments suggest that proline forms an ordered supramolecular assembly at high concentrations, resembling a protein folding chaperone, due to the formation of an ordered, amphipathic, supramolecular assembly. It is worth mentioning here that proline itself can also function as an agent against oxidative stress in plants to contribute to osmotic adjustment of a species

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

The above mentioned facts might also be a chemical-physical explanation for various sorts of phenomena in which proline acts as an essential element of complex molecular behavior. To name just a few among many examples, we point to the SH3 domain - a small protein domain of ~60 residues that can be found in various proteins and that mediates the

interaction with other proteins through a proline patch of consensual sequence -X-P-p-X-P-. 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. A proline rich conserved region is necessary for the folding of cytochrome P450. Lastly, the interaction sites within a proline rich domain of p53 protein are indispensable for tumor supression.

Up until now the only high level computational studies of proline interactions are an excellent work by Morozov and the already mentioned work of Riley et. al, We have recently studied the intramolecular stabilization of a small artificial protein Trp–cage using advanced computational methods. In the structure of this miniprotein we noticed two very strong non-covalent interactions involving proline residues and the central tryptophan amino acid. Both of these distinct arrangements of proline and tryptophan have shown significant stabilization energies at ab initio MP2/aug-cc-pVDZ level (~ 8kcal/mol).

The nature of the interaction between proline and an aromatic moiety has not been studied in details so far. Hence there are several questions that arise concerning this type of non-covalent interaction: Why are these interactions so strong? What is the role of aromaticity in these interactions? Are these proline-aromatic interactions chiefly attributable to dispersion forces? 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 that correlation/dispersion is particularly significant for interactions between molecules of non-polar character. The theoretical methods that take electron correlation into account, and that are computationally feasible for the size of the systems in focus are the Møller-Plesset (MP) perturbation methods, the density functional method augmented by empirical London dispersion energy (DFT-D), and the density functional theory combined with symmetry adapted perturbation theory (DFT-SAPT). It is our aim to use these methods to investigate the character of the interaction between proline and an aromatic moiety and to determine the most important structural and electronic properties of this type of interaction.

Methods

Initially we selected two systems for the calculation of aromatic-proline interaction, both taken from the structure of Trp-cage (pdb code 1L2Y), namely the pairs of residues Trp6-Pro17 and Trp6-Pro18 (Figure 1). In order to discover the source of stabilization in these two contacts, we employed several models of the interacting moieties. Because previous theoretical works describing interactions involving proline have always included the carboxyl group of the amino acid, we also aim to determine how important the presence of these polar atoms in is in terms of their contribution to the overall stability of a proline-tryptophan complex.

In our first model (referred herein as the “large model”; cf Chart 1) we represented the molecules in the same way as in our previous paper i.e. with the peptide bond between a residue and the preceding amino acid (we cut the protein backbone at the C-C_α bond and cap the resulting fragments with hydrogens) to take the backbone atoms, and especially the peptide bond, into account. Next, we reduced the system to the side-chain only (refereed herein as the “small model”), starting from the C_β atom. In the case of proline we represented it as a pyrrolidine molecule to preserve its cyclic structure. In order to evaluate the role of the

nitrogen heteroatom on the strength of stabilization, we replaced it with a carbon atom, representing the proline with a cyclopentane model (cf Chart 1)..

Last but not least, in order to assess the significance of the cyclic arrangement of proline we calculate the interaction in an acyclic model with the same number of heavy atoms – the C_α model of leucine. We took the structure of one of the representative clusters of the Trp-Leu interaction taken from the electronic version of the side chain atlas, in which the Leu adopts a "stacked" arrangement relative to the Trp sidechain.

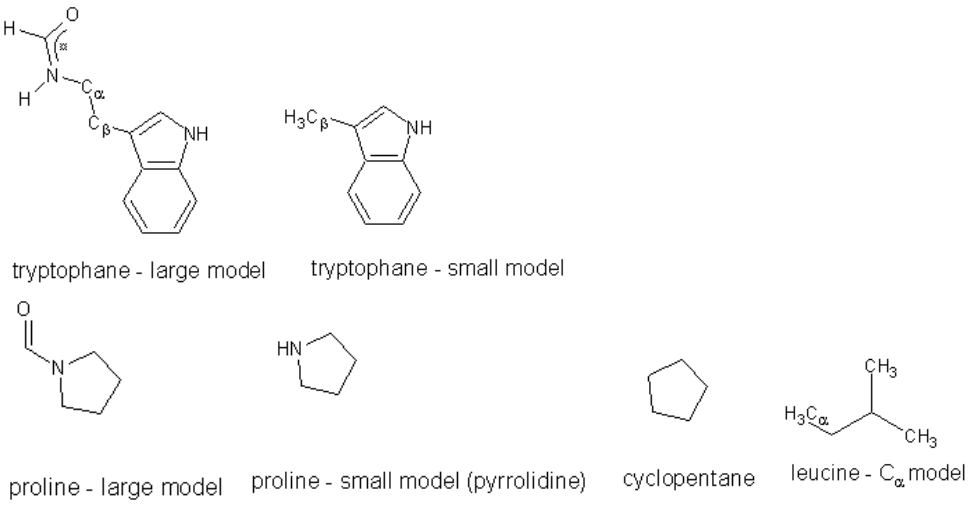


Chart 1. Chemical structure of the molecules used as model systems for the aromatic-proline interaction.

To more clearly understand the role of proline in protein strucure more generally, we utilized an updated version of the Protein Side-Chain Atlas (http://www.biochem.ucl.ac.uk/bsm/sidechains/), which presents all 3D contacts between the amino acids side chains within the known protein structures taken from PDB set. We studied interaction of proline with phenylalanine, proline, and tryptophan. There are 6 structural clusters for each of the selected pair-wise interaction based on clustering method utilized for classification of the all interacting pairs from PDB involving proline. By carefully checking we found that only 3 clusters of Pro-Phe, 4 clusters of Pro-Pro, and 5 clusters of Pro-Trp represent stacking-like arrangement. We have evaluated the interaction energies for the stacking arrangements of these complexes.

In all studied systems we first perform geometry optimizations of hydrogen using the DFT/B3LYP/6-31G** method. Interaction energies were evaluated at two levels. First, the MP2/aug-cc-pVDZ method was applied. It is well known that the MP2 procedure, when combined with extended basis sets, overestimates the interaction energy. this is especially true for stacking interactions. To obtain reliable interaction energies it is necessary to pass to higher level calculations (e.g. CCSD(T)) and 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 now collected an extensive body of evidence suggesting that MP2/medium basis set interaction energies are close to the CCSD(T)/complete basis set limit ones. This is evidently due to a compensation of errors. In the following text we used the MP2/aug-cc-pVDZ method, which provides reliable interaction energies. In the second step we applied the

recently introduced DFT-D/TPSS/TZVP method. Here the DFT method is augmented by an empirical term that describes the London dispersion energy. The parameters in the accompanying damping function were parameterized against accurate interaction energies of a set of selected complexes. The method thus yields interaction energies that are close to CCSD(T)/complete basis set values. This technique was also used for geometry optimization. The strong point of the method is its computational efficiency, which allows for its use even with extended complexes. We use the RI technique to speed up the calculations. Calculations were performed in Gaussian 03 and Turbomole v.5 programs as well as our own code for the DFT-D calculations.

Both of the above mentioned methods provide total interaction energies. For the purpose of evaluating the origin of complex stabilization it is desirable to have also energy components. Partial splitting can be obtained from RI-DFT-D calculations, which yields 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.
The SAPT interaction energy is given as:

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

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:

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

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

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

and

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

These four quantities refer to the electrostatic, induction, dispersion, and exchange contributions (respectively) to the overall interaction energy. The δ_{HF} term is a Hartree-Fock correction for higher order contributions to the interaction energy that are not included within the other DFT-SAPT terms.

In this work we have performed DFT-SAPT computations using Dunning's aug-cc-pVDZ basis set. It is known that when this, relatively small, basis set is employed in these types of calculations the dispersion contribution to the interaction energy is generally underestimated by approximately 10%-20%. In order to estimate the effect of this underestimation we have simply increased the magnitude of the aug-cc-pVDZ dispersion term by 15% (ie multiplied it by 1.15). The computational details of the DFT-SAPT calculations performed are analogous to those described by Jansen et al

Results

Initially we selected the two interacting Trp-Pro complexes, which show surprisingly high stabilization in the Trp-cage miniprotein – the Trp6-Pro17 and Trp6-Pro18 (see Figure 1). We adopt two models for the interacting residues, the "small" and "large" models described in the Methods section. First we examine both the large and small model of the Trp6-Pro17 complex (see Figure 2 A,B), which assumes a perpendicular arrangement, which we will refer to as the "L-shaped" conformation. The main difference between the small and large model complexes studied here is that the large model contains a hydrogen bond between the tryptophan cyclic amine and proline's carbonyl group, whereas for the small model there is no hydrogen bond. In Table 1 it can be seen that there is a very strong interaction in the large model, whereas the binding energy for the small model is almost negligible.

Table 1. Interaction energies of the Trp6-Pro17 pair in large and small models using MP2 and DFT-D methods.

	Trp6-Pro17	
	large model	small model
MP2/aDZ	-7.8	-0.9
DFT-D/TPSS/TZVP	-7.6	-1.1
D part	-1.9	-1.1

Evidently the key reason for the higher interaction strength in the large model is the presence of the hydrogen bond, as we assume that the backbone portion of the tryptophan residue contributes negligibly to the interaction due to their distance. It is further seen that, in the case of large model, the interaction energy originates from the DFT interaction energy and that the dispersion energy plays only a minor role. In the case of small model the situation is different and now all stabilization comes from dispersion energy. Notice also the very good agreement between MP2 and DFT-D interaction energies, which is characteristic for H-bonded complexes.

The situation is more interesting in the case of the Trp6-Pro18 (see Figure 2 C,D) contact, where the two molecules are in a "stacked" arrangement. The results for this complex are summarized in Table 2. There is no H-bond between the two interacting systems, and, in this case, the truncation of the carbonyl group leads to only a moderate decrease in the interaction energy. The interaction is in fact dispersion-driven; the dispersion term constitutes a major portion of the interaction energy in both the large and the small models (See the last line in Table 2).

Table 2. Interaction energies in the "stacked" arrangement of Trp and Pro residues in various models of the interaction calculated using MP2 and DFT-D methods.

	Trp6-Pro18		Trp-Leu
	large model	small model	cyclopentane	leucine
MP2/aDZ	-8.4	-6.5	-5.1	-4.7
DFT-D/TPSS/TZVP	-6.8	-5.4	-4.0	-3.3
D part	-6.3	-5.0	-5.7	-4.6

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. Here the proline molecule is an aliphatic system without any π electrons, and so this interaction can be classified as mixed π/aliphatic type. Evidently, rather significant stacking interaction energies can be attained even if a non-aromatic system stacks with an aromatic one. The prerequisite for this is clearly the cyclic character of the proline system. In the present case the MP2 and DFT-D interaction energies differ more than previously and, as discussed above, the MP2 stabilization energy is overestimated. The present DFT-D energies are expected to be more reliable.

The size of the models used to describe these tryptophan-proline complexes has a relatively large impact on the stabilization energies obtained. This can be attributed to two key factors, the number of atoms that are contained in the dispersion contact region and the effects of the side chains modifying the electronic density within the tryptophan and proline rings. Contribution of the carbonyl group of proline to the total stabilization energy is relatively large (~2 kcal/mol,) indicating that not only the presence of the polar oxygen is important but also its influence on the electron structure of the interacting complexes. This carbonyl oxygen, which is within proteins a part of main chain, thus modulates the proline affinity to its interacting partners depending on its precise geometrical arrangement. Nevertheless, the dominant part of stabilization originates in the proline moiety.

It is interesting to ask where this large stabilization energy comes from, i.e. what is the role of the nitrogen heteroatom and how important is the cyclic arrangement. To this end we compare these tryptophan-proline results to data obtained for two other models, the tryptophan-cyclopentane and the tryptophan-leucine complexes (see Figure 2 E,F). For details on the selection of these systems see the Methods section. The data given in Table 2, column 4 and 5th show that the replacement of the NH group by a CH_2 group reduces the interaction energy by about 1.5 kcal/mol, but the resulting interaction energy is still relatively large. The presence of the heteroatom contributes to, but is not essential for, the large interaction strength in the tryptophan-proline complexes. Passing from proline to cyclopentane evidently reduces the molecular dipole moment, which leads to a lowering of the DFT stabilization energy. Also following expectation, the dispersion energy becomes larger for the cyclopentane model (see the last line in Table 2). Results obtained for the leucine model further demonstrate that the same is true for the cyclic arrangement of the atoms, that is to say that having a cyclic arrangement enhances the interaction but is not a defining characteristic. Changing the arrangement from being cyclic (cyclopentane) to acyclic (leucine) only decreases the interaction energy by about ~0.5 kcal/mol. To sum up, the large stabilization in the Trp-Pro complex is due to a number of contributions, none of which is clearly dominant.

The next step in this study is to examine how far the arrangements derived from protein structures are from their optimum geometrical positions. The strength of the stacked tryptophan-proline interaction in this system is so large that it seems very unlikely that a more stable interaction arrangement could exist. In order to determine the minimum energy conformations we have performed geometry optimizations on the complexes of stacked tryptophan (small model) with proline (small model), cyclopentane, and leucine. The interaction energies calculated using the DFT-D method in the optimized geometries are shown in Table 3. Here it can be seen that, indeed, the geometries changed very little in the

cases of the cyclopentane and leucine models (cf. the value of interaction energy in the unoptimized complex).

Table 3. Interaction energies in the "stacked" arrangement of Trp and Pro residues in various models of the interaction after full geometry optimization calculated using DFT-D.

	Trp6-Pro18			Trp-Leu
	small model	small model*	cyclopentane	leucine
DFT-D/TPSS/TZVP	-11.6	-5.5	-4.1	-3.6
D part	-5.0	-4.9	-5.4	-4.4

* restricted position of N atom

On the other hand, the Trp-Pro small model complex relaxed into a geometry in which an H-bond is formed between the NH group of Trp and the N atom of Pro, with a very large interaction energy of -11.6 kcal/mol. This H-bond formation is, however, an artifact caused by the fragmentation and does not reflect the accessible geometric possibilities for a proline residue within a protein structure, where its movement would be restricted by the presence of the backbone. If we repeat the geometry optimization with the proline N atom's position fixed relative to the Trp molecule, the geometry of the complex changes only slightly during the course of optimization and the interaction energy for the optimized complex is very similar to that of the unoptimized system. We therefore conclude that the residue contacts within protein structures are very close to their geometry optima. This conclusion is essential since it supports the fact that it is interaction energies/enthalpies between protein building blocks that determine the structure of the protein interior.

Another question we ask is: "how large is the interaction energy between proline and an aromatic residue, or another proline, within a real protein?". We focused on three interacting systems: Pro-Pro, Pro-Phe and, Pro-Trp, representing interactions between cyclic systems of aromatic and non-aromatic nature. Table 4 gives the interaction energies of the aforementioned complexes in their representative "stacking" geometry arrangements obtained from the PDB (see Figure 3). It is quite clear that interaction energies rise with the size of the interacting systems eg. Pro-Pro < Phe-Pro < Trp-Pro. The highest interaction energy for a given type of complex is -2.3 kcal/mol for Pro-Pro, -2.7 kcal/mol for Pro-Phe and -3.7 kcal/mol for Pro-Trp. The average binding energy values varies from -2.0 kcal/mol for Pro-Pro to -3.3 kcal/mol for Pro-Trp. It can be seen in Table 4 that it is the dispersion that is chiefly responsible for such strong stabilization energies. It should also be kept in mind that the cluster representative geometries studied here could generally be quite far from the actual geometries found in real protein structures and that these values do not represent the most ideal arrangements for these interactions. As such , the interaction energies obtained for these structures most likely represent near minimum bounds for these types of interactions. What is quite stunning is the fact that, in all cases, the Pro-Trp interaction energy values approach those of typical hydrogen bonds, and that this type of interaction is relatively well populated in proteins.

Table 4. Interaction energies of Trp-Pro, Phe-Pro and Pro –Pro pairs taken from side chain atlas, their interaction energies in geometries from crystal and then optimized

		DFT-D/TPSS/TZVP	D part
Phe-Pro	clust 1	-2.3	-2.9
	clust 2	-2.7	-2.8
	clust 3	-2.4	-3.9
Pro-Pro	clust 1	-1.6	-2.7
	clust 2	-2.3	-2.5
	clust 4	-2.2	-2.9
Trp-Pro	clust 1	-3.0	-5.2
	clust 2	-3.7	-3.8
	clust 3	-3.3	-5.1
	clust 4	-3.4	-3.4
	clust 5	-3.2	-6.0

In order to decompose the interaction energies into their physically meaningful components we performed SAPT analyses of the Pro-Trp complexes. As mentioned above, all DFT-SAPT calculations have been carried out using the aug-cc-pVDZ basis set. DFT-SAPT/aug-cc-pVTZ results have been estimated by increasing the dispersion contribution to the interaction energy by 15%. In the following discussion, unless otherwise noted, we will refer only to the estimated aug-cc-pVTZ results.

Table 5 gives the DFT-SAPT interaction energy decomposition results for the Trp6-Pro17 (L-shaped) complex. Here it can be seen that for the, hydrogen bonding, large model, which includes the residue backbone atoms, the interaction is very strong (-11.28 kcal/mol) and is largely dominated by the electrostatic term (-9.85 kcal/mol). This result is expected for a hydrogen bonding pair; as would also be expected the induction (-5.61 kcal/mol) and dispersion (-5.15 kcal/mol) terms are relatively large. For the small Trp-6-Pro17 model, which contains no hydrogen bonding interaction, the attraction between the two residues is weak (-1.01 kcal/mol) and can be chiefly ascribed to the dispersion term. It is interesting to note that the electrostatic term for this interaction is positive, indicating that repulsive electrostatic forces exist for this configuration of the small model L-shaped complex.

Table 5. SAPT interaction energy decomposition for the L-shaped configuration of the tryptophan-proline complex (kcal/mol, a-pVxZ refers to aug-cc-pVxZ).

	Small model		large model	
	a-pVDZ	est. a-pVTZ	a-pVDZ	est. a-pVTZ
$E(elec.)$	0.17		-9.85	
$E(ind.)$	-0.12		-5.61	
$E(disp.)$	-1.17	-1.35	-4.48	-5.15
$E(exch.)$	0.30		11.16	
δHF	-0.01		-1.82	
E_{int}	-0.83	-1.01	-10.61	-11.28

In terms of DFT-SAPT analysis, the interaction in the stacked Trp6-Pro18 complex is much different in nature than the interaction occurring for the L-shaped complex Trp6-Pro17, as demonstrated by the data given in Table 6. First considering the results obtained for the large model, it can be seen that the overall strength of interaction is weaker and more dispersive in character than the hydrogen bonding interaction for the L-shaped complex. The fact that this interaction is dominated by the dispersion term is not surprising, as this is typically found for stacked structures . It is somewhat surprising, however, that the electrostatic term (-5.64 kcal/mol) is also relatively large. A possible explanation for the strength of the electrostatic interaction is 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 it can be seen that each of the DFT-SAPT interaction energy terms is smaller than in the case of the large model. The fact that there are fewer atoms that can dispersively interact within the small model probably accounts for most of the weakening of the interaction. Another factor that may play a role in making the interaction seen for the small model weaker than that of the large model is the fact that the presence of the side chains may somewhat change the electronic character of the proline and tryptophan rings.

Table 6. SAPT interaction energy decomposition for the stacked arrangement of the tryptophan-proline complex (kcal/mol, a-pVxZ refers to aug-cc-pVxZ).

	Small model		large model	
	a-pVDZ	est. a-pVTZ	a-pVDZ	est. a-pVTZ
$E(elec.)$	-3.37		-5.64	
$E(ind.)$	-0.45		-0.97	
$E(disp.)$	-6.62	-7.61	-8.79	-10.11
$E(exch.)$	5.57		9.21	
δHF	-0.43		-0.71	
E_{int}	-5.30	-6.29	-6.89	-8.21

As mentioned above, the full optimization of the tryptophan-proline (small model) complex yields a structure with a hydrogen bond occurring between the two residues. It can be seen in Table 7 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, with the magnitude of the interaction energy for the optimized structure being only 0.08 kcal/mol higher than that of the unoptimized structure. It is interesting to note that each of the DFT-SAPT interaction energy terms for the (restrictively) optimized complex is within 0.20 kcal/mol of its counterpart for the unoptimized complex, confirming that the two structures are very similar.

Table 7. SAPT interaction energy decomposition for optimized structures of the tryptophan-proline complexes (kcal/mol, a-pVxZ refers to aug-cc-pVxZ).

	full optimization		restricted optimization	
	a-pVDZ	est. a-pVTZ	a-pVDZ	est. a-pVTZ
$E(elec.)$	-16.02		-3.39	
$E(ind.)$	-3.10		-0.47	
$E(disp.)$	-8.55	-9.84	-6.50	-7.48
$E(exch.)$	21.40		5.39	
δHF	-3.19		-0.42	
E_{int}	-9.46	-10.74	-5.39	-6.37

One of the most prominent aspects of the DFT-SAPT data given in Table 8 is the fact that, considering the results for the tryptophan-cyclopentane complex, the induction, dispersion, and exchange terms are all very similar to those obtained for the tryptophan-proline complex. The main difference between the tryptophan-cyclopentane and the tryptophan-proline interactions is that the electrostatic term is substantially smaller for the former one. This indicates that there is a relatively strong electrostatic interaction connected to the presence of the nitrogen heteroatom on the pyrrolidine ring that contributes non-negligibly to the overall tryptophan-proline interaction. As discussed above passing from proline to cyclopentane reduces the molecular dipole moment, which leads to a lowering of the electrostatic attraction. Looking at the tryptophan-leucine data it can be seen that this interaction bears little resemblance, in terms of its DFT-SAPT interaction energy components, to the tryptophan-proline (or tryptophan-cyclopentane) interaction. The fact that the dispersion and exchange terms are much lower for this type of interaction indicates that there is less contact (overlap) between the leucine and tryptophan than for the other complexes. This seems reasonable because the density of atoms within cyclic molecular arrangements should lead to more contact between two molecules.

Table 8. SAPT interaction energy decomposition for optimized structures of the tryptophan-cyclopentane and tryptophan-leucine complexes (kcal/mol, a-pVxZ refers to aug-cc-pVxZ).

	Cyclopentane		leucine	
	a-pVDZ	est. a-pVTZ	a-pVDZ	est. a-pVTZ
$E(elec.)$	-2.02		-1.47	
$E(ind.)$	-0.28		-0.20	
$E(disp.)$	-6.67	-7.67	-4.86	-5.58
$E(exch.)$	5.61		3.53	
δHF	-0.38		-0.26	
E_{int}	-3.74	-4.74	-3.26	-3.99

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 might be of similar importance, as suggested by the recent works by Morozov et. al. and by Riley et. al. In this work we have performed the correlated ab initio, DFT-D, and DFT-SAPT calculations on tryptophan-proline complexes derived from the experimental structure of the Trp–cage miniprotein. The most important result of this study is that the tryptophan-proline interaction, when it is arranged in a stacked configuration, is bound very strongly (-6.8 kcal/mol at the DFT-D/TPSS/TZVP level). The fact that the dispersion term in the DFT-D method generated 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.

Comparison of the tryptophan-proline and tryptophan-cyclopentane interactions (with the same geometries) reveals that the proline complex is bound more strongly than the cyclopentane one by about 1.5 kcal/mol. 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 tryptophan-proline complex electrostatically. In order to study the role of cyclicity on the interaction that occurs for the tryptophan-proline complex, we carried out calculations on the tryptophan-leucine complex. It was found that leucine binds more weakly to tryptophan than does proline (or cyclopentane), with a binding energy of -3.33 kcal/mol (DFT-D/TPSS/TZVP – compared with -6.4 kcal for Trp-Pro complex). The most likely explanation for this decreased interaction energy is the fact that the cyclic arrangements of proline and cyclopentane allow for the highest degree of dispersive contact with tryptophan.

When we carried out a (restricted) geometry optimization of the tryptophan-proline complex it was found that the structure's geometry does not change very much and that the interaction energy of the optimized structure is very close to that of the initial structure. This indicates that the structure taken from the tryptophan cage crystal structure is geometrically very close to the potential energy minimum. 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. Optimization of cyclopentane-tryptophan and leucine-tryptophan complexes from initial geometries based on the structure of the proline-tryptophan stacked complex does not show any notable geometrical rearrangement. This indicates that the interaction energy surface of tryptophan is probably very shallow and that various interacting partners, regardess of their character (generally aliphatic moieties), can be bound.

When proline, phenylalanine, and tryptophan are compared as interacting partners of proline based on arrangements in real protein structures, a hierarchical picture of the studied interactions was obtained: the most stabilizing is the Trp-Pro interaction, and the most important fact is that most of the Trp-Pro clusters were found in stacking arrangements. This is not the case for Phe-Pro or Pro-Pro arrangements, where only three stacked clusters have been found for each interaction. The average stabilization energy value for these complexes

varies from -1.7 fro Pro-Pro to -3.3 for Trp-Pro. This is quite a significant contribution considering that the interaction energies are evaluated only for side-chain - side-chain interactions of the studied amino acid residues.

The results presented in this work indicate that close contacts between an aromatic residue and proline may result in very stable interactions that may function to stabilize structural elements in proteins and protein complexes. These findings might be an explanation of various phenomena where proline plays not only a structural, but also an interaction role. Together with the exceptional rigidity of this residue, it 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.

Acknowledgement

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 Nos. Z40550506 and MSM6198959216.

References: 1 R. B. Greaves and J. Warwicker, Bmc Structural Biology, 2007, 7.
2 K. Ishii and K. Marumo, Resource Geology, 2002, 52, 135-146.
3 R. Jaenicke, Journal of Biotechnology, 2000, 79, 193-203.
4 I. Nilsson, A. Saaf, P. Whitley, G. Gafvelin, C. Waller, and G. von Heijne, Journal of Molecular Biology, 1998, 284, 1165-1175.
5 S. Arnold, A. Curtiss, D. H. Dean, and O. Alzate, Febs Letters, 2001, 490, 70-74.
6 R. H. Yun, A. Anderson, and J. Hermans, Proteins-Structure Function and Genetics, 1991, 10, 219-228.
7 S. C. Li, N. K. Goto, K. A. Williams, and C. M. Deber, Proceedings of the National Academy of Sciences of the United States of America, 1996, 93, 6676-6681.
8 J. Y. Suh, Y. T. Lee, C. B. Park, K. H. Lee, S. C. Kim, and B. S. Choi, European Journal of Biochemistry, 1999, 266, 665-674.
9 P. Chakrabarti and S. Chakrabarti, Journal of Molecular Biology, 1998, 284, 867-873.
10 T. G. Consler, O. Tsolas, and H. R. Kaback, Biochemistry, 1991, 30, 1291-1298.
11 F. G. Meng, Y. D. Park, and H. M. Zhou, International Journal of Biochemistry & Cell Biology, 2001, 33, 701-709.
12 T. K. S. Kumar, D. Samuel, G. Jayaraman, T. Srimathi, and C. Yu, Biochemistry and Molecular Biology International, 1998, 46, 509-517.
13 H. B. C. Molinari, C. J. Marur, E. Daros, M. K. F. de Campos, J. F. R. P. de Carvalho, J. C. Bespalhok, L. F. P. Pereira, and L. G. E. Vieira, Physiologia Plantarum, 2007, 130, 218-229.
14 K. E. Riley, G. L. Cui, and K. M. Merz, Journal of Physical Chemistry B, 2007, 111, 5700-5707.
15 M. L. Waters, Biopolymers, 2004, 76, 435-445.
16 S. Aravinda, N. Shamala, C. Das, A. Sriranjini, I. L. Karle, and P. Balaram, J. Am. Chem. Soc., 2003, 125, 5308-5315.
17 S. K. Burley and G. A. Petsko, Science, 1985, 229, 23-28.
18 S. S. C. Li, Biochemical Journal, 2005, 390, 641-653.
19 A. Palencia, E. S. Cobos, P. L. Mateo, J. C. Martinez, and I. Luque, Journal of Molecular Biology, 2004, 336, 527-537.
20 M. Meiyappan, G. Birrane, and J. A. A. Ladias, Journal of Molecular Biology, 2007, 372, 970-980.
21 B. Kemper, Toxicology and Applied Pharmacology, 2004, 199, 305-315.
22 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.
23 A. V. Morozov, K. M. S. Misura, K. Tsemekhman, and D. Baker, Journal of Physical Chemistry B, 2004, 108, 8489-8496.
24 L. Bendova, P. Hobza, and J. Vondrasek, Proteins-Structure Function and Bioinformatics, 2007, in press.
25 L. L. Qiu, S. A. Pabit, A. E. Roitberg, and S. J. Hagen, J. Am. Chem. Soc., 2002, 124, 12952-12953.
26 P. Hobza, R. Zahradnik, and K. Muller-Dethlefs, Collection of Czechoslovak Chemical Communications, 2006, 71, 443-531.
27 P. Jurecka, J. Sponer, J. Cerny, and P. Hobza, Phys. Chem Chem Phys., 2006, 8, 1985-1993.

28 P. Jurecka, J. Cerny, P. Hobza, and D. R. Salahub, J Comput. Chem, 2007, 28, 555-569.
29 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.
Ref Type: Computer Program
30 R. Ahlrichs, M. Bar, M. Haser, H. Horn, and C. Kolmel, Chemical Physics Letters, 1989, 162, 165-169.
31 P. Jurecka, J. Cerny, P. Hobza, and D. R. Salahub, J Comput. Chem, 2007, 28, 555-569.
32 R. Podeszwa, R. Bukowski, and K. Szalewicz, Journal of Chemical Theory and Computation, 2006, 2, 400-412.
33 A. Hesselmann and G. Jansen, Chem. Phys. Lett., 2002, 362, 319-325.
34 A. Hesselmann and G. Jansen, Chem. Phys. Lett., 2002, 357, 464-470.
35 A. Misquitta, R. Podeszwa, B. Jeziorski, and K. Szalewicz, J. Chem. Phys., 2005, 123, 214103.
36 A. Hesselmann, G. Jansen, and M. Schutz, J. Chem. Phys., 2005, 122, 14103.
37 M. O. Sinnokrot and C. D. Sherrill, J. Am. Chem. Soc., 2004, 126, 7690-7697.

Figures:

Figure 1. The geometry of Trp-cage miniprotein.

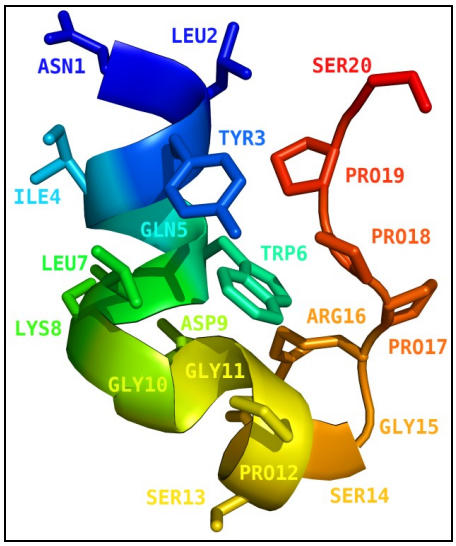
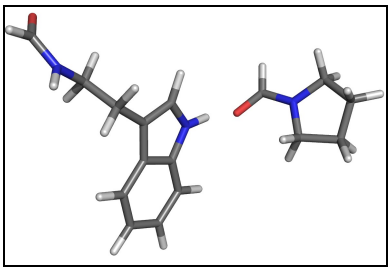
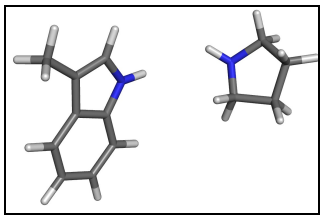


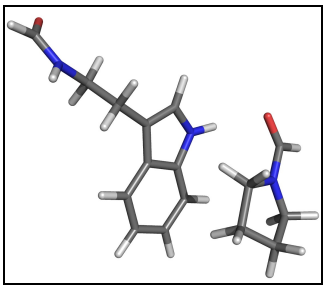
Figure 2 Large (A) and small (B) model of interaction between Pro17-Trp6, large (C) and small (D) model of interaction between Pro18-Trp6, Trp6 interacting with cyclopentane (E) based on geometry of Pro18 and Trp-Ile complex (F) in stacking orientation obtained from Atlas of Protein Side-Chain Interactions



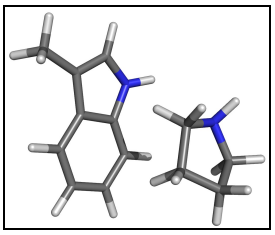
A



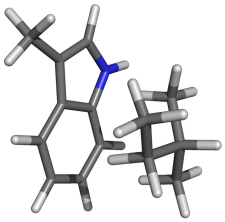
B

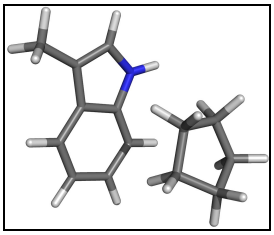


C



D

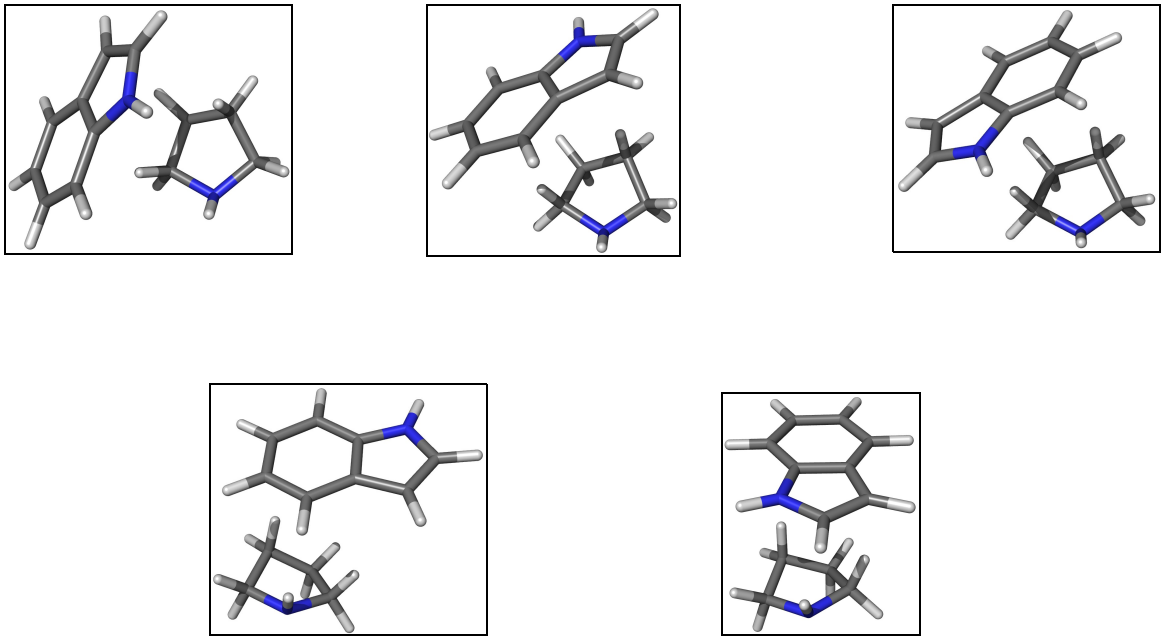




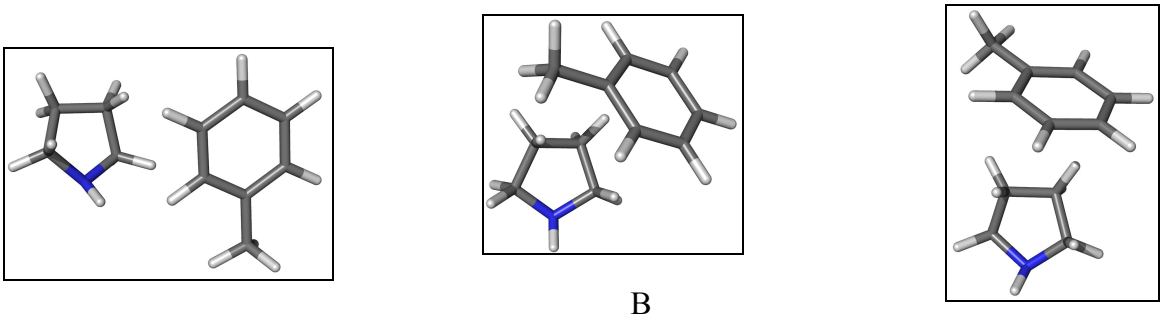
E

F

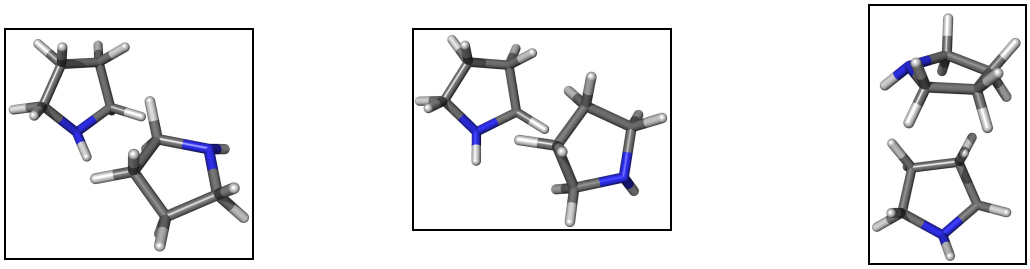
Figure 3. Representative clusters (cluster 1,2,3,4,5) of Trp-Pro (A), (cluster 1,2,4) Phe-Pro (B) and (cluster 1,2,3) Pro-Pro(C) complexes from Atlas of Protein Side-Chain Interactions



A



B



C

Model of Peptide Bond−Aromatic Ring Interaction: Correlated Ab Initio Quantum Chemical Study

Lada Bendová,[†] Petr Jurečka,[‡] Pavel Hobza,*[,†,‡] and Jiří Vondrášek*[,†]

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and Center for Biomolecules and Complex Molecular Systems, Flemingovo namesti 2, 166 10 Prague 6, Czech Republic, and Department of Physical Chemistry and Center for Biomolecules and Complex Molecular Systems, Palacký University, Tr. Svobody 26, 771 46 Olomouc, Czech Republic

Received: April 12, 2007; In Final Form: June 14, 2007

Aromatic ring−peptide bond interactions (modeled as benzene and formamide, *N*-methylformamide and *N*-methylacetamide) are studied by means of advanced computational chemistry methods: second-order Möller−Plesset (MP2), coupled-cluster single and double excitation model [CCSD(T)], and density functional theory with dispersion (DFT-D). The geometrical preferences of these interactions as well as their interaction energy content, in both parallel and T-shaped arrangements, are investigated. The stabilization energy reaches a value of over 5 kcal/mol for the *N*-methylformamide−benzene complex at the CCSD(T)/complete basis set (CBS) level. Decomposition of interaction energy by the DFT-symmetry-adapted perturbation treatment (SAPT) technique shows that the parallel and T-shaped arrangements, although similar in their total interaction energies, differ significantly in the proportion of electrostatic and dispersion terms.

Introduction

The stability of protein tertiary structure is a result of an interplay between various noncovalent interactions, which contribute to the overall stability of the protein molecule because of their frequent occurrence. Besides the long-recognized forces, for example, H-bonds and salt bridges, there are abundant van der Waals interactions, among which the aromatic interactions such as $\pi-\pi$ stacking and XH$-\pi$ H-bonding have been shown to also play an important role for protein structure,[1] as well as protein−ligand recognition.[2]

The first to point out the importance of aromatic interactions in proteins were Burley and Petsko[3] in a work on interaction between phenylalanine residues. The strength of the stabilization energy in Phe pairs was estimated to be 1−2 kcal/mol[4−6] by gas-phase calculations of benzene and toluene dimer model systems. In the benzene dimer model, two minima exist—parallel displaced arrangement and T-shaped arrangement—both stabilized by favorable dispersion and quadrupole−quadrupole electrostatic interaction.[7,8]

It is worth mentioning that $\pi-\pi$ stacking is not strictly defined as an interaction of aromatic systems; it is a more general phenomenon that includes interactions of planar systems with delocalized orbitals, such as peptide bonds. Analogically, the XH$-\pi$ bonding is not limited to aromatic−aromatic interaction, because the aromatic ring can serve as an acceptor for nonaromatic H-bond donors.[9] It has been shown by gas-phase calculations[10] that both electrostatic and dispersion terms are important in the XH$-\pi$ interaction. The directionality of the interaction is mainly controlled by the electrostatic term; however, the potential energy surface is very flat near the minimum due to the long-range dispersion term.

In proteins, the interactions of an aromatic ring with a nonaromatic moiety have been subject of an exhaustive Protein Data Bank (PDB) mining study,[11] which revealed a relatively high occurrence of such interactions (1 per 10.8 aromatic amino acids). A large fraction of these interactions are NH$-\pi$ contacts. This finding can be attributed to the abundance of potential NH donors in the protein backbone as well as in the side chains of some residues. Moreover, the size of the aromatic ring together with the broad minimum of the potential energy surface makes it a good substitute in the case where conventional H-bond acceptors are locally absent. Therefore the aromatic−amide interaction is one of the most common weak interactions in proteins. Several studies[4,12,13] have shown that the N bearing groups prefer to lie above the aromatic ring. However, only a fraction of these contacts can be considered to be H-bonds; the NH$-\pi$ interactions are in fact outnumbered by the aromatic−amide stacked structures[12,14] with the sp^2-hybridized N atom. This finding is in accordance with the higher number of conventional H bonds the NH group can form in the stacked orientation.[15] The significance of the aromatic−amide interaction for stabilization of secondary structure and local motifs has been proved by PDB mining.[16] It includes stabilization of α-helix termini and β-sheet edges as well as regular turns.[11] The aromatic ring−peptide bond interaction has also been shown to be important in certain biological recognition processes.[17]

The aromatic−amide interaction has previously been studied with a benzene−formamide model.[18] The authors performed a restricted optimization and reported counterpoise-corrected interaction energy values of 4.0 and 2.0 kcal/mol for the T-shaped and the stacked complex, respectively, at the MP2/6-311G(2d,2p) gas-phase level. Another study explored the T-shaped arrangement of the *N*-methylformamide (NMF)−benzene complex[19] at the MP2/6-31G** level with counterpoise correction. The maximum gas-phase stabilization energy found

* Corresponding authors: fax (+420) 220-410-320; e-mails: jiri.vondrasek@uochb.cas.cz, pavel.hobza@uochb.cas.cz.

[†] Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and Center for Biomolecules and Complex Molecular Systems.

[‡] Department of Physical Chemistry and Center for Biomolecules and Complex Molecular Systems, Palacký University.

10.1021/jp072859+ CCC: $37.00
Published on Web 08/02/2007

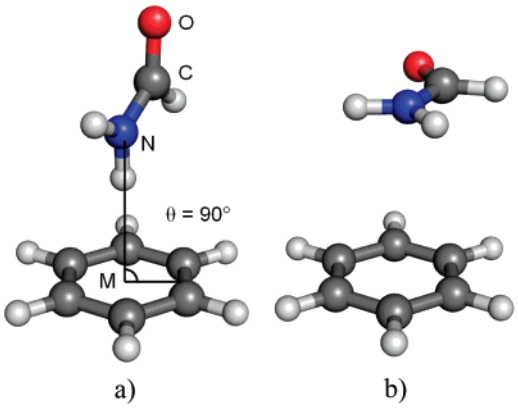


Figure 1. Initial geometries of FMA−benzene complex: (a) T-shaped arrangement and (b) stacked arrangement. M, center of benzene ring; θ, angle between the plane of benzene ring and the NCO plane of the peptide bond moiety.

was 4.4 kcal/mol. In our previous study[20] we examined the interaction between a phenylalanine residue and a peptide bond (modeled as *N*-methylformamide) in the crystal structure of the small protein rubredoxin with high-level ab initio calculations. The interaction energy determined by use of coupled-cluster single and double excitation model [CCSD(T)] at the complete basis set limit (CBS) amounted to −8.2 kcal/mol. Such an unexpectedly strong interaction motivated us to investigate the aromatic−amide interaction in more detail. Another reason is that the above-mentioned studies employed a standard computational level, which might have affected the quality of the results.

In this work we investigate spatial preferences of aromatic ring−peptide bond interaction. We aim at investigating complexes representative of the arrangements found in protein structure. We optimize geometries and calculate interaction energies for both parallel and T-shaped arrangements. Further, we uncover the underlying forces governing the directional preferences of the interaction using interaction energy decomposition.

Computational Methods

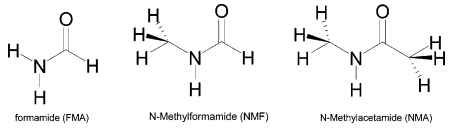
RI-MP2 Gradient Optimization. We use the RI-MP2/cc-pVTZ method for optimization of all subsystems [benzene, formamide (FMA), *N*-methylformamide (NMF), and *N*-methylacetamide (NMA)] as well as a reference method for optimization of the complexes. The resolution of identity (RI) procedure is used to make the calculations more economic.

CCSD(T)/CBS Interaction Energy. In the optimized complexes, the CCSD(T)/CBS interaction energy is approximated as

$$\Delta E_{\mathrm{CCSD(T)}}{}^{\mathrm{CBS}} = \Delta E_{\mathrm{MP2}}{}^{\mathrm{CBS}} + (\Delta E_{\mathrm{CCSD(T)}} - \Delta E_{\mathrm{MP2}})^{\text{small basis set}} \quad (1)$$

The former term is determined by use of the two-point Helgaker extrapolation scheme.[21] The Hartree−Fock and correlation MP2 energies necessary for the extrapolation are determined at the RI-MP2/aug-cc-pVXZ (X = D, T) level. The CCSD(T) term is calculated with a smaller basis set, 6-31G*(0.25) (the exponent of d-functions changed from standard value of 0.8 to a more diffuse one of 0.25). The difference between the MP2 and CCSD(T) interaction energies does not strongly depend on basis set size (unlike the interaction energies themselves); thus the 6-31G*(0.25) basis set already gives satisfactory values for this difference.[22] The deformation energy E^{def} is determined as the difference between the energy of isolated subsystems in the complex geometry and the energy of the supersystem is also calculated at the RI-MP2/cc-pVTZ level to correct the interaction energy. The total interaction energy is calculated as

CHART 1: Systems Studied in Interaction with Benzene



$$\Delta E^{\mathrm{total}} = \Delta E_{\mathrm{CCSD(T)}}{}^{\mathrm{CBS}} + \Delta E^{\mathrm{def}}$$

All interaction energies are counterpoise-corrected for basis-set superposition error (BSSE).

Interaction Energy by Use of DFT Functionals. Both the B3LYP functional,[23] which is commonly used for calculating biopolymers, and the PWB6K functional designed by Zhao and Truhlar[24] are tested for their ability to correctly describe the interaction in the studied complexes. These two functionals are used with the 6-31G** basis set on the geometries obtained by MP2/cc-pVTZ optimization.

RI-DFT-D Optimization. The RI-DFT-D (DFT with an empirical dispersion term[25]) optimization is performed with the TPSS functional and extended Pople-type basis set [6-311++G-(3df,2pd), denoted as LP]. The interaction energy of the optimized complexes is calculated at the same level. The BSSE at this level is negligible due to the large basis set employed, and it is thus not necessary to correct the interaction energy for its effect.

Symmetry-Adapted Perturbation Treatment. Components of the total interaction energies are determined by DFT-symmetry-adapted perturbation treatment (SAPT).[26,27] Calculations are performed for T-shaped and tilted T-shaped structures of the NMA−benzene complex optimized by RI-MP2/cc-pVTZ as well as for a set of stacked structures varying in the NM distance. PBE0AC exchange−correlation functional with density fitting and the aug-cc-pVDZ basis set are used. The aug-cc-pVDZ set is large enough to provide a reliable estimate of the electrostatic, induction, and exchange components (note that the SAPT calculations are not burdened with the BSSE). Although this basis set underestimates the dispersion component by about 10−20%,[27] it should nevertheless serve well for the purpose of comparing relative strengths of this term in different arrangements.

In SAPT, total interaction energy E_{int} is calculated as a sum of electrostatic, exchange, induction, and dispersion components, with the dispersion and induction components also having their exchange counterparts:

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

The exponents in eq 2 refer to the intermolecular perturbation order. E^1 is defined as the sum of $E^1{}_{\mathrm{pol}}$ and $E^1{}_{\mathrm{ex}}$, and E^2 is the sum of $E^2{}_{\mathrm{ind}}$, $E^2{}_{\mathrm{ex-ind}}$, $E^2{}_{\mathrm{disp}}$, and $E^2{}_{\mathrm{ex-disp}}$ energies, respectively. δHF denotes the estimate for higher-order contributions. In the present study, we do not use the standard SAPT method (which is prohibitively expensive for the complexes investigated) but employ its DFT version,[26,27] in which the intramolecular correlation is treated fully by the DFT, whereas the intermolecular interaction is calculated by perturbational treatment. Since perturbation theory exploits orbital energies, the inherently

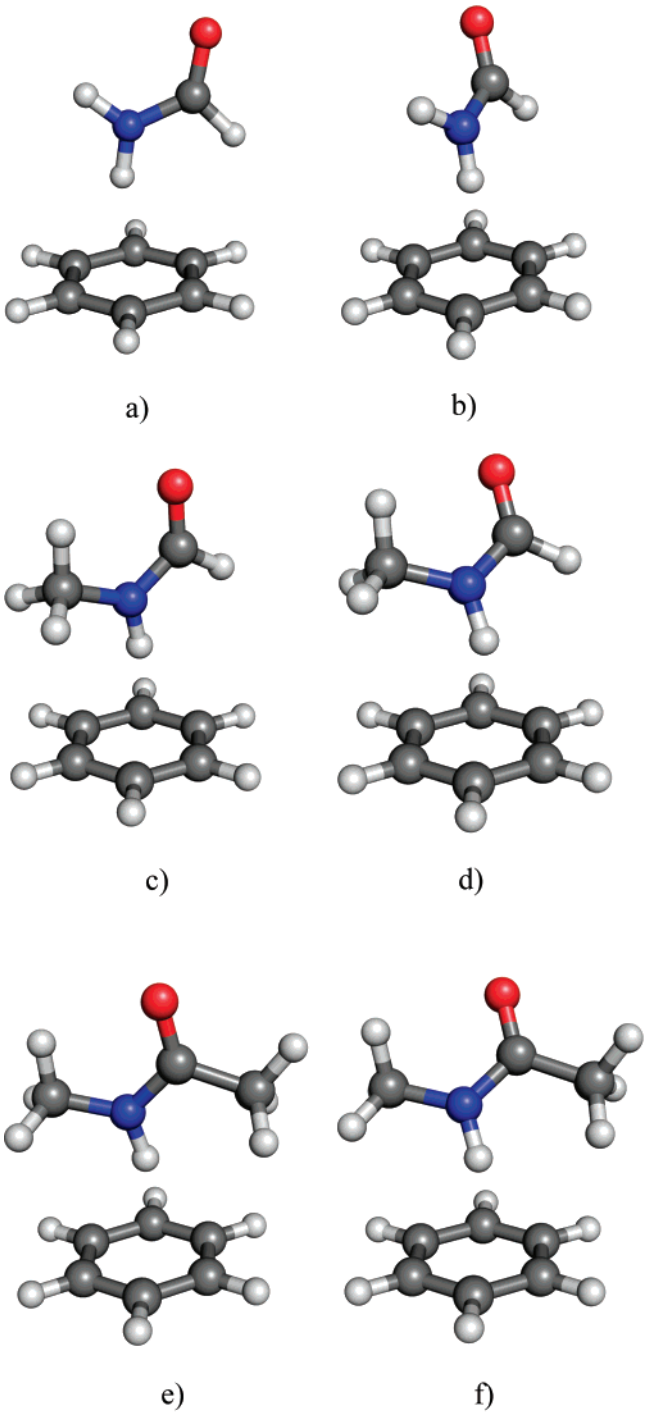


Figure 2. Geometries of optimized complexes: (a) FMA−benzene starting from T-shaped arrangement; (b) FMA−benzene starting from stacked arrangement; (c) NMF−benzene starting from T-shaped arrangement; (d) NMF−benzene starting from stacked arrangement; (e) NMA−benzene starting from T-shaped arrangement; (f) NMA−benzene starting from stacked arrangement.

incorrect DFT orbitals are corrected by gradient-controlled shift procedure,[26,27] with the difference (shift) between the vertical ionization potential (IP) and highest occupied molecular orbital (HOMO) energy of the DFT method employed as an input. Herein the IPs are calculated at the PBE0/TZVP level while the HOMO values are taken from the aug-cc-pVDZ calculation. Our shift values are 0.0715 and 0.0855 Eh for benzene and NMA, respectively.

We use Gaussian 03,[28] Turbomole,[29] and MOLPRO[30] codes as well as our own code for RI-DFT-D calculations.[25]

TABLE 1: Geometrical Parameters of Optimized X−Benzene Complexes[a]

	MP2/cc-pVTZ		RI-DFT-D	
system	NM,[b] Å	θ,[c] deg	NM,[b] Å	θ,[c] deg
NMA_T	3.221	89	3.569	88
NMA_S	3.188	52	3.422	58
NMF_T	3.210	74	3.420	83
NMF_S	3.144	55	3.374	56
FMA_T	3.274	83	3.524	86
FMA_S	3.259	68	3.342	31

[a] X = NMA, NMF, and FMA. [b] Distance between the amide nitrogen of X and the benzene ring center (M). [c] Interplanar angle.

Results and Discussion

We use three models for the peptide bond moiety: formamide (FMA), *N*-methylformamide (NMF), and *N*-methylacetamide (NMA) (see Chart 1). Two main geometrical arrangements are possible for aromatic ring−peptide bond interactions: parallel and T-shaped, shown in Figure 1 for FMA and benzene. For the T-shaped arrangement, the angle between the plane of the benzene ring and the NCO plane of the peptide bond moiety $\theta = 90°$ (the planes are perpendicular) and the NH bond is directed toward the center of the benzene ring. For the parallel arrangement, the angle $\theta = 0°$ (the planes are mutually parallel) and the N atom lies in the C6 axis of benzene. We use the annotation _T and _S for T-shaped and stacked **starting** geometries, respectively.

Geometries of optimized complexes are shown in Figure 2, and the results of the geometry optimization are summarized in Table 1. The interacting complexes are characterized by the NM distance between the amide nitrogen N and the center of the benzene ring M. In geometry optimizations starting from the T-shape arrangement we observe only moderate differences between initial and final geometries in all three models. On the other hand, when starting from the parallel arrangement, the geometries undergo profound changes. The θ angle changes from the initial value of 0 degrees to values of about 50−70 degrees. (Hereafter, we term the structures characterized by θ angle values in the range of 20−70 degrees as tilted-T-shaped ones, discerning them from both strictly parallel with θ angle 0−20 degrees and strictly T-shaped with θ angle of 70−90 degrees.) The changes are similar for NMA and NMF complexes while being less profound for FMA. Although all three model systems pass from parallel arrangements to tilted-T-shaped ones, methylation evidently disfavors T-shape arrangement and favors the tilted-T-shape. Given the shallowness of the PES, the RI-DFT−D and MP2/cc-pVTZ results agree reasonably well, although the distances are slightly larger for DFT−D optimized complexes.

Table 2 shows energy characteristics for the optimized complexes. Columns 2−4 show MP2 interaction energies determined with the aug-cc-pVDZ and aug-cc-pVTZ basis sets and at the CBS limit, respectively. Passing to a larger basis set yields a substantial stabilization energy increase (~0.6 kcal/mol), and extrapolation to CBS limit yields even larger stabilization energies (the highest being 6.4 kcal/mol for NMA−benzene). However, it is known[22] that the MP2/CBS stabilization energies are overestimated and that the CCSD(T) correction term should be included. This term is positive, that is, of a repulsive character for all systems, being similar for both structures of NMA and NMF (~1 kcal/mol) and slightly smaller for both structures of FMA (~0.6 kcal/mol).

The final stabilization energies, E^{total}, including the deformation energy, are shown in column 8. Evidently, these energies

TABLE 2: Interaction Energies[a] for MP2/cc-pVTZ and RI-DFT−D Optimized Structures

system	aDZ	aTZ	CBS	ΔCCSD(T)	CCSD(T)/CBS	E^{def}	E^{total}	PWB6K	B3LYP	RI-DFT-D
NMA_T	−5.3	−6.0	−6.3	1.0	−5.3	0.3	−5.0	−4.2	0.4	−5.1
NMA_S	−5.4	−6.1	−6.4	1.1	−5.3	0.1	−5.2	−4.1	0.8	−5.3
NMF_T	−4.9	−5.6	−5.8	0.9	−5.0	0.0	−5.0	−4.2	−0.4	−5.0
NMF_S	−5.0	−5.7	−6.0	0.9	−5.1	0.1	−5.0	−4.2	0.0	−5.0
FMA_T	−4.3	−4.9	−5.2	0.6	−4.6	0.0	−4.6	−4.4	−1.0	−4.4
FMA_S	−4.3	−4.9	−5.2	0.6	−4.6	0.0	−4.6	−4.3	−0.9	−4.6

[a] Interaction energies are given in kilocalories per mole and were determined at various theoretical levels. aDZ, aTZ, and CBS denote aug-cc-pVDZ, aug-cc-pVTZ, and complete basis set limit, respectively. E^{def} is the deformation energy.

TABLE 3: Components of Interaction Energy for Structures of NMA−Benzene Complex[a]

	R (Å)	E^1_{pol}	E^1_{ex}	E^{2*}_{ind}[b]	E^{2*}_{disp}[c]	δHF	E^1	E^2	E_{int}	E^{2*}_{disp}/E^1_{pol}
T	3.2	−7.3	13.4	−1.8	−10.6	−1.3	6.1	−12.4	−7.1	1.45
tT	3.2	−8.0	15.3	−1.9	−11.6	−1.5	7.3	−13.5	−7.0	1.45
S	3.2	−9.3	24.9	−0.9	−15.3	−1.9	15.7	−16.3	−2.5	1.65
S	3.4[d]	−4.5	11.0	−0.6	−10.6	−0.8	6.5	−11.2	−5.5	2.36
S	3.5[d]	−3.4	8.1	−0.5	−9.2	−0.6	4.7	−9.7	−5.6	2.71
S	3.6[d]	−2.6	6.0	−0.5	−8.0	−0.4	3.4	−8.4	−5.5	3.08

[a] Components of interaction energy (T, T-shaped; tT, tilted T-shaped; S, stacked) are given in kilocalories per mole and were determined by DFT-SAPT. [b] $E^{2*}_{ind} = E^2_{ind} + E^{2*}_{ex-ind}$. [c] $E^{2*}_{disp} = E^2_{disp} + E^{2*}_{ex-disp}$. [d] Hydrogens of methyl group(s) optimized.

are similar to those obtained at the MP2/aug-cc-pVDZ level. This is due to a partial compensation of the stabilization energy increase between the aug-cc-pVDZ and the CBS level and the decrease brought by the CCSD(T) correction term. However, this compensation cannot be relied upon, as for some complexes the MP2/aug-cc-pVDZ method may yield substantially incorrect stabilization energies. We may conclude that the total stabilization energies are relatively high, varying from 4.6 kcal/mol for FMA−benzene to 5.2 kcal/mol for NMA−benzene. The magnitude of these interactions is comparable to classical H-bonding.

Columns 9 and 10 in Table 2 show DFT interaction energies determined with the recently introduced PWB6K functional and with the widely used B3LYP functional, respectively. The limited performance of the DFT procedure for stacked complexes caused by a lack of London dispersion energy in DFT methods is well-known. Evidently, the B3LYP functional completely fails and stabilization energies are strongly underestimated. The performance of the PWB6K functional is more promising. However, the PWB6K stabilization energies are systematically underestimated and do not reflect the effect of system size. The last column of Table 2 shows the RI-DFT-D stabilization energies for structures optimized at the same level. These results are very similar to the CCSD(T)/CBS ones; the effect of methylation is described correctly. This is in accord with the fact that the RI-DFT-D method is parametrized toward CCSD(T)/CBS values and therefore it is supposed to yield very accurate energies as well as geometries.[25] An interesting fact is that, for all models, the final interaction energies for T-shaped and tilted T-shaped arrangements do not differ too much. This suggests that the interaction hypersurface of these two systems is quite shallow and that no significant barrier exists between different local minima.

To elucidate the nature of these interactions in more detail, we determined the energy components using the perturbation DFT-SAPT technique. Table 3 shows the interaction energy components for NMA−benzene complexes in different arrangements. The interaction energies of T-shaped and tilted T-shaped structures are very similar and their components differ only moderately. The electrostatic energy (E^1_{pol}) is more attractive for the tilted T-shaped arrangement, but this growth is compensated by an increase of the exchange-repulsion term. The effective induction energy (E^{2*}_{ind}) is small for both arrangements. The effective dispersion term (E^{2*}_{disp}), which constitutes a major contribution to the overall stabilization, is larger for the tilted T-shaped structure, but again the difference is not significant.

Although no stacked arrangement minimum has been found by geometry optimizations, this arrangement is found in protein structures.[11] Therefore it would be interesting to see whether the described trends remain valid for this arrangement. Upon changing from the T-shaped to a parallel arrangement while the NM distance is kept constant, the interaction energy becomes much smaller (−2.5 vs −7.1 kcal/ mol). This is due to the enhanced repulsion between the methyl group(s) and benzene, reflected by a large exchange-repulsion term. To reduce this repulsion we optimize the position of methyl hydrogens and perform a scan along the NM axis. Table 3 shows that the optimal distance in the stacked arrangement is larger, with a value of 3.5 Å. The electrostatic energy is considerably less attractive in the parallel arrangement than in both the T-shaped and tilted T-shaped arrangements; also, the exchange-repulsion term is notably smaller. Effective induction energy is rather small, with the majority of the stabilization originating from effective dispersion energy, which, despite the larger distance in the stacked complex, is comparable for all three arrangements. The ratio E^{2*}_{disp}/E^1_{pol} is, however, significantly higher in the stacked arrangement than in both the T-shaped and tilted T-shaped ones (2.71 comparing to 1.45).

We have to conclude that similar interaction energies in the T-shaped and parallel arrangements arise from the complementarity of a repulsive E^1 term (the sum of E^1_{pol} and E^1_{ex}) and an attractive E^2 term (the sum of E^2_{ind}, E^2_{ex-ind}, E^2_{disp}, and $E^2_{ex-disp}$) energies. The E^1 term is more favorable (less repulsive) for the parallel arrangement, while the attractive E^2 energy is more favorable for the T-shaped arrangement.

Conclusions

The most stable arrangements found in optimizations of all three studied model systems are either T-shaped or tilted T-shaped; the energetic difference between these two arrangements is small, and no substantial barrier exists between these two minima. The interaction energies in both arrangements are large [up to −5.3 kcal/mol in NMA−benzene at the CCSD(T)/CBS level] and comparable to a classic H-bond. Such arrangements exist in proteins and, consequently, their contribu-

tion to protein stabilization should be quite significant. The size of the system brings a large increase of interaction, which is in accordance with the result of our previous work, where the calculated interaction energy in phenylalanine−NMF complex was −8.2 kcal/ mol.

The impressive performance of RI-DFT-D is notable. RI-DFT-D/LP stabilization energies agree well with the CCSD-(T)/CBS values, even though they require several orders of magnitude less CPU computational time. The MP2/aug-cc-pVDZ interaction energies are also very close to the CCSD-(T)/CBS ones, probably due to cancellation of errors. It is, however, not advisable to rely on this cancellation.

Decomposition analysis of interaction energy in T-shaped, tilted T-shaped, and parallel arrangements that exist in proteins was performed. All the arrangements exhibit comparable interaction energies in DFT-SAPT, which is an important observation in light of the fact that H-bonding (expected to be a dominant stabilization feature) exists only in the former two arrangements. The interaction energy decomposition indicated different natures of their stabilization due to the complementarity between repulsive E^1 and attractive E^2 energies. Dispersion energy is a major stabilizing term in stacked as well as T-shaped structures; however, they differ significantly in the $E^{2*}_{\mathrm{disp}}/E^{1}_{\mathrm{pol}}$ ratio.

Acknowledgment. This work was supported by Grants 203/05/0009 and 203/06/1727 from the Grant Agency of the Czech Republic and Grant LC512 from the Ministry of Education (MSMT) of the Czech Republic. It was also part of research projects Z40550506 and MSM6198959216. We are grateful to Kevin E. Riley for his comments and help with language corrections.

References and Notes

(1) Frank, B. S.; Vardar, D.; Buckley, D. A.; McKnight, C. J. The role of aromatic residues in the hydrophobic core of the villin headpiece subdomain. *Protein Sci.* **2002**, *11* (3), 680−687.

(2) Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem., Int. Ed.* **2003**, *42* (11), 1210−1250.

(3) Burley, S. K.; Petsko, G. A. Aromatic-Aromatic Interaction: a Mechanism of Protein Structure Stabilization. *Science* **1985**, *229* (4708), 23−28.

(4) Burley, S. K.; Petsko, G. A. Dimerization Energetics of Benzene and Aromatic Amino Acid Side Chains. *J. Am. Chem. Soc.* **1986**, *108* (25), 7995−8001.

(5) Chipot, C.; Jaffe, R.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. Benzene dimer: A good model for π-π interactions in proteins? A comparison between the benzene and the toluene dimers in the gas phase and in an aqueous solution. *J. Am. Chem. Soc.* **1996**, *118* (45), 11217−11224.

(6) Hobza, P.; Selzle, H. L.; Schlag, E. W. Potential energy surface for the benzene dimer. Results of ab initio CCSD(T) calculations show two nearly isoenergetic structures: T-shaped and parallel-displaced. *J. Phys. Chem.* **1996**, *100* (48), 18790−18794.

(7) Muller-Dethlefs, K.; Hobza, P. Noncovalent interactions: A challenge for experiment and theory. *Chem. Rev.* **2000**, *100* (1), 143−167.

(8) Hobza, P.; Zahradnik, R.; Muller-Dethlefs, K. The world of non-covalent interactions: 2006. *Collect. Czech. Chem. Commun.* **2006**, *71* (4), 443−531.

(9) Zhao, Y.; Tishchenko, O.; Truhlar, D. G. How well can density functional methods describe hydrogen bonds to π acceptors? *J. Phys. Chem. B* **2005**, *109* (41), 19046−19051.

(10) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. Origin of the attraction and directionality of the NH/π interaction: Comparison with OH/π and CH/π interactions. *J. Am. Chem. Soc.* **2000**, *122* (46), 11450−11458.

(11) Steiner, T.; Koellner, G. Hydrogen bonds with π-acceptors in proteins: Frequencies and role in stabilizing local 3D structures. *J. Mol. Biol.* **2001**, *305* (3), 535−557.

(12) Mitchell, J. B. O.; Nandi, C. L.; Mcdonald, I. K.; Thornton, J. M.; Price, S. L. Amino/Aromatic Interactions in Proteins: Is the Evidence Stacked Against Hydrogen Bonding? *J. Mol. Biol.* **1994**, *239* (2), 315−331.

(13) Flocco, M. M.; Mowbray, S. L. Planar Stacking Interactions of Arginine and Aromatic Side-Chains in Proteins. *J. Mol. Biol.* **1994**, *235* (2), 709−717.

(14) Singh, J.; Thornton, J. M. Sirius: An Automated-Method for the Analysis of the Preferred Packing Arrangements Between Protein Groups. *J. Mol. Biol.* **1990**, *211* (3), 595−615.

(15) Nishio, M.; Hirota, M.; Umezawa, Y. The CH/pi Interaction; Wiley−VCH: New York, 1998.

(16) Toth, G.; Watts, C. R.; Murphy, R. F.; Lovas, S. Significance of aromatic−backbone amide interactions in protein structure. *Proteins: Struct., Funct., Genet.* **2001**, *43* (4), 373−381.

(17) Perutz, M. F. The Role of Aromatic Rings As Hydrogen-Bond Acceptors in Molecular Recognition. *Philos. Trans. R. Soc. London, Ser. A* **1993**, *345* (1674), 105−112.

(18) Duan, G. L.; Smith, V. H.; Weaver, D. F. An ab initio and data mining study on aromatic−amide interactions. *Chem. Phys. Lett.* **1999**, *310* (3−4), 323−332.

(19) Cheng, J. G.; Kang, C. M.; Zhu, W. L.; Luo, X. M.; Puah, C. M.; Chen, K. X.; Shen, J. H.; Jiang, H. L. *N*-Methylformamide-benzene complex as a prototypical peptide N-H$\cdots\pi$ hydrogen-bonded system: Density functional theory and MP2 studies. *J. Org. Chem.* **2003**, *68* (19), 7490−7495.

(20) Vondrasek, J.; Bendova, L.; Klusak, V.; Hobza, P. Unexpectedly strong energy stabilization inside the hydrophobic core of small protein rubredoxin mediated by aromatic residues: Correlated ab initio quantum chemical calculations. *J. Am. Chem. Soc.* **2005**, *127* (8), 2615−2619.

(21) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N-2, and H_2O. *Chem. Phys. Lett.* **1998**, *286* (3−4), 243−252.

(22) Jurecka, P.; Hobza, P. On the convergence of the ($\Delta E^{\mathrm{CCSD(T)}}-\Delta E^{\mathrm{MP2}}$) term for complexes with multiple H-bonds. *Chem. Phys. Lett.* **2002**, *365* (1−2), 89−94.

(23) Becke, A. D. Density-Functional Thermochemistry. 3. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648−5652.

(24) Zhao, Y.; Truhlar, D. G. Design of density functionals that are broadly accurate for thermochemistry, thermochemical kinetics, and nonbonded interactions. *J. Phys. Chem. A* **2005**, *109* (25), 5656−5667.

(25) Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R. Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. *J. Comput. Chem.* **2007**, *28* (2), 555−569.

(26) Hesselmann, A.; Jansen, G.; Schutz, M. Density-functional theory−symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *J. Chem. Phys.* **2005**, *122* (1).

(27) Hesselmann, A.; Jansen, G.; Schutz, M. Interaction energy contributions of H-bonded and stacked structures of the AT and GC DNA base pairs from the combined density functional theory and intermolecular perturbation theory approach. *J. Am. Chem. Soc.* **2006**, *128* (36), 11730−11731.

(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(29) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. Electronic-Structure Calculations on Workstation Computers the Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162* (3), 165−169.

(30) Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R. Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; and Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. MOLPRO, version 2006.1: a package of ab initio programs.

Proteins: Structure, Function, and Bioinformatics

Copy of e-mail Notification

Your article (21938) from "Proteins: Structure, Function, and Bioinformatics" is available for download

Proteins: Structure, Function, and Bioinformatics Published by John Wiley & Sons, Inc.

Dear Sir or Madam,

PDF page proofs for your article are ready for review.

Please refer to this URL address

http://kwglobal.co.in/jw/retrieval.aspx

Login: your e-mail address
Password: ----

The site contains 1 file. You will need to have Adobe Acrobat Reader software to read these files. This is free software and is available for user downloading at http://www.adobe.com/products/acrobat/readstep.html.

This file contains:

Author Instructions Checklist
Adobe Acrobat Users - NOTES tool sheet
Reprint Order form
A copy of your page proofs for your article

After printing the PDF file, please read the page proofs. Since this article has been reviewed and accepted based on its original content, now is not the time to make substanive changes to text, tables, or figures. Next:

1) indicate changes or corrections in the margin of the page proofs;
2) answer all queries on the last page of the PDF proof;
3) proofread any tables and equations carefully;
4) check that any Greek, especially "mu", has translated correctly.

Within 48 hours, please return the following to the address given below:

1) original PDF set of page proofs,
2) Reprint Order form

Return to:

Production Editor, PROT
Cadmus Professional Communications
300 West Chestnut Street

Proteins: Structure, Function, and Bioinformatics

Copy of e-mail Notification

Ephrata, PA 17522
U.S.A.
E-mail: jrnlprodprot@cadmus.com
Fax: 717-738-9478 or 717-738-9479

Your article will be published online via our EarlyView service soon after correction receipt. Your prompt attention to and return of page proofs is crucial to faster publication of your work.
If you experience technical problems, please contact Birender/Sundeep (wileysupport@kwglobal.com, or +91(44) 4205-8888 (ext. 310).

If you have any questions regarding your article, please contact me. PLEASE ALWAYS INCLUDE YOUR ARTICLE NO. (21938) WITH ALL CORRESPONDENCE.

This e-proof is to be used only for the purpose of returning corrections to the publisher.

Sincerely,

Production Editor, PROT
Cadmus Professional Communications
300 West Chestnut Street
Ephrata, PA 17522
U.S.A.

Tel: 717-738-9334
Fax: 717-738-9478 or 717-738-9479
E-mail: jrnlprodprot@cadmus.com

Author Satisfaction Survey - Proteins: Structure, Function, and Bioinformatics

Please follow the below link to an online survey concerning your experience in publishing in this journal. Your answers will assist Wiley in refining the publishing process to best suit your needs and those of your fellow authors. Thanks for your time and consideration.

http://interscience.wiley.com/survey/prot

WILEY
Publishers Since 1807

111 RIVER STREET, HOBOKEN, NJ 07030

*****IMMEDIATE RESPONSE REQUIRED*****

Your article will be published online via Wiley's EarlyView® service (www.interscience.wiley.com) shortly after receipt of corrections. EarlyView® is Wiley's online publication of individual articles in full text HTML and/or pdf format before release of the compiled print issue of the journal. Articles posted online in EarlyView® are peer-reviewed, copyedited, author corrected, and fully citable via the article DOI (for further information, visit www.doi.org). EarlyView® means you benefit from the best of two worlds--fast online availability as well as traditional, issue-based archiving.

Please follow these instructions to avoid delay of publication.

☐ **READ PROOFS CAREFULLY**

- This will be your only chance to review these proofs**. Please note that once your corrected article is posted online, it is considered legally published, and cannot be removed from the Web site for further corrections.**
- Also note that the volume and page numbers shown on the proofs are for position only.

☐ **ANSWER ALL QUERIES ON PROOFS** (Queries for you to answer are attached as the last page of your proof.)

- Mark all corrections directly on the proofs. Note that excessive author alterations may ultimately result in delay of publication and extra costs may be charged to you.

☐ **CHECK FIGURES AND TABLES CAREFULLY**

- Check size, numbering, and orientation of figures.
- All images in the PDF are downsampled (reduced to lower resolution and file size) to facilitate Internet delivery. ----- These images will appear at higher resolution and sharpness in the printed article.
- Review figure legends to ensure that they are complete.
- Check all tables. Review layout, title, and footnotes.

☐ **COMPLETE REPRINT ORDER FORM**

- Fill out the attached reprint order form. It is important to return the form even if you are not ordering reprints. You may, if you wish, pay for the reprints with a credit card. Reprints will be mailed only after your article appears in print. This is the most opportune time to order reprints. If you wait until after your article comes off press, the reprints will be considerably more expensive.

RETURN ☐**PROOFS**
☐**REPRINT ORDER FORM**
☐**CTA (If you have not already signed one)**

RETURN WITHIN 48 HOURS OF RECEIPT VIA FAX TO DEBRA KEENER **717-738-9478 or 717-738-9479.**

QUESTIONS ?

Production Editor, PROT
Phone: 717-738-9334 fax: 717-738-9360
E-mail: jrnlprodprot@cadmus.com
Please refer to journal acronym and article production number

Softproofing for advanced Adobe Acrobat Users - NOTES tool

NOTE: ACROBAT READER FROM THE INTERNET DOES NOT CONTAIN THE NOTES TOOL USED IN THIS PROCEDURE.

Acrobat annotation tools can be very useful for indicating changes to the PDF proof of your article. By using Acrobat annotation tools, a full digital pathway can be maintained for your page proofs.

The NOTES annotation tool can be used with either Adobe Acrobat 6.0 or Adobe Acrobat 7.0. Other annotation tools are also available in Acrobat 6.0, but this instruction sheet will concentrate on how to use the NOTES tool. Acrobat Reader, the free Internet download software from Adobe, DOES NOT contain the NOTES tool. In order to softproof using the NOTES tool you must have the full software suite Adobe Acrobat Exchange 6.0 or Adobe Acrobat 7.0 installed on your computer.

Steps for Softproofing using Adobe Acrobat NOTES tool:

1. Open the PDF page proof of your article using either Adobe Acrobat Exchange 6.0 or Adobe Acrobat 7.0. Proof your article on-screen or print a copy for markup of changes.

2. Go to Edit/Preferences/Commenting (in Acrobat 6.0) or Edit/Preferences/Commenting (in Acrobat 7.0) check “Always use login name for author name” option. Also, set the font size at 9 or 10 point.

3. When you have decided on the corrections to your article, select the NOTES tool from the Acrobat toolbox (Acrobat 6.0) and click to display note text to be changed, or Comments/Add Note (in Acrobat 7.0).

4. Enter your corrections into the NOTES text box window. Be sure to clearly indicate where the correction is to be placed and what text it will effect. If necessary to avoid confusion, you can use your TEXT SELECTION tool to copy the text to be corrected and paste it into the NOTES text box window. At this point, you can type the corrections directly into the NOTES text box window. **DO NOT correct the text by typing directly on the PDF page.**

5. Go through your entire article using the NOTES tool as described in Step 4.

6. When you have completed the corrections to your article, go to Document/Export Comments (in Acrobat 6.0) or Comments/Export Comments (in Acrobat 7.0). Save your NOTES file to a place on your harddrive where you can easily locate it. **Name your NOTES file with the article number assigned to your article in the original softproofing e-mail message.**

7. **When closing your article PDF be sure NOT to save changes to original file.**

8. To make changes to a NOTES file you have exported, simply re-open the original PDF proof file, go to Document/Import Comments and import the NOTES file you saved. Make changes and reexport NOTES file keeping the same file name.

9. When complete, attach your NOTES file to a reply e-mail message. Be sure to include your name, the date, and the title of the journal your article will be printed in.

John Wiley & Sons, Inc.

Publishers Since 1807

REPRINT BILLING DEPARTMENT • 111 RIVER STREET • HOBOKEN, NJ 07030
PHONE: (201) 748-8789; FAX: (201) 748-6326
E-MAIL: reprints @ wiley.com

PREPUBLICATION REPRINT ORDER FORM

Please complete this form even if you are not ordering reprints. This form **MUST** be returned with your corrected proofs and original manuscript. Your reprints will be shipped approximately 4 weeks after publication. Reprints ordered after printing are substantially more expensive.

JOURNAL: ***PROTEINS: Structure, Function, and Bioinformatics*** VOLUME_____ ISSUE_____

TITLE OF MANUSCRIPT_____

MS. NO._____ NO. OF PAGES_____ AUTHOR(S)_____

REPRINTS 8 1/4 X 11

No. of Pages	100 Reprints	200 Reprints	300 Reprints	400 Reprints	500 Reprints
	$	$	$	$	$
1-4	336	501	694	890	1,052
5-8	469	703	987	1,251	1,477
9-12	594	923	1,234	1,565	1,850
13-16	714	1,156	1,527	1,901	2,273
17-20	794	1,340	1,775	2,212	2,648
21-24	911	1,529	2,031	2,536	3,037
25-28	1,004	1,707	2,267	2,828	3,388
29-32	1,108	1,894	2,515	3,135	3,755
33-36	1,219	2,092	2,773	3,456	4,143
37-40	1,329	2,290	3,033	3,776	4,528

****** REPRINTS ARE ONLY AVAILABLE IN LOTS OF 100. IF YOU WISH TO ORDER MORE THAN 500 REPRINTS, PLEASE CONTACT OUR REPRINTS DEPARTMENT AT (201)748-8789 FOR A PRICE QUOTE.

Please send me _____ reprints of the above article at..... $_____

Please add appropriate State and Local Tax {Tax Exempt No._____} $_____
Please add 5% Postage and Handling..... $_____
TOTAL AMOUNT OF ORDER** ..... $_____

***International orders must be paid in U.S. currency and drawn on a U.S. bank*

Please check one: Check enclosed Bill me Credit Card
If credit card order, charge to: American Express Visa MasterCard Discover
Credit Card No._____ Signature_____ Exp. Date_____

Bill To:

Name_____
Address/Institution_____
_____
_____
_____

Ship To:

Name_____
Address/Institution_____
_____
_____
_____

Purchase Order No. _____ Phone _____ Fax _____
E-mail: _____

WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Identifying stabilizing key residues in proteins using interresidue interaction energy matrix

Lada Bendová, Pavel Hobza, and Jiří Vondrášek*

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, Flemingovo nám. 2, Prague, Czech Republic

ABSTRACT

We are proposing an interresidue interaction energy map (IEM)—a new tool for protein structure analysis and protein bioinformatics. This approach employs the sum of pair-wise interaction energies of a particular residue as a measure of its structural importance. We will show that the IEM can serve as a means for identifying key residues responsible for the stability of a protein. Our method can be compared with the interresidue contact map but has the advantage of weighting the contacts by the stabilization energy content which they bring to the protein structure. For the theoretical adjustment of the proposed method, we chose the Trp-cage mini protein as a model system to compare a spectrum of computational methods ranging from the ab initio MP2 level through the DFT method to empirical force-field methods. The IEM method correctly identifies Tryptophane 6 as the key residue in the Trp-cage. The other residues with the highest stabilizing contributions correspond to the structurally important positions in the protein. We have further tested our method on the Trp2Cage miniprotein—a P12W mutant of the Trp-cage and on two proteins from the rubredoxin family that differ in their thermostability. Our method correctly identified the thermodynamically more stable variants in both cases and therefore can also be used as a tool for the relative measurement of protein stability. Finally, we will point out the important role played by dispersion energy, which contributes significantly to the total stabilization energy and whose role in aromatic pairs is clearly dominant. Surprisingly, the dispersion energy plays an even more important role in the interaction of prolines with aromatic systems.

Proteins 2008; 00:000–000.

Key words: noncovalent interactions; protein stability; ab initio quantum chemical calculations; hydrophobic core; bioinformatics.

INTRODUCTION

The specific spatial arrangement of a folded protein is the result of the restraints brought by the covalent bonds in the protein backbone and the various noncovalent interactions. The Anfinsen's principle[1] states that spontaneously folding proteins contain all the information necessary for unambiguous folding in their sequence. Therefore, what determines the folding of a particular sequence into a single most stable native conformation are the residues' mutual noncovalent interactions and their interactions with solvent.

It is known from mutational studies that some residues are crucial for the stability of the folded protein while others can be mutated freely. Not only do the mutational studies provide information about the native state, they can also be used to derive information on the residue interactions in the transition state (TS). Interpreting the quantity $<\phi>_i^{exp}$ (the ratio of the change in stability of the transition state ensemble (TSE) to that of the native state caused by the mutation of residue i) as a fraction of the native-like contacts of the residue in the TS and using it as the restraints in Monte Carlo simulations, Vendruscolo *et al.*[2] were able to show that a small number of key residues are sufficient for the definition of the native-like topology of a protein in the TS via their tertiary contacts.[2] All these residues are a part of the hydrophobic core and form a large number of contacts with other residues.

The experimental determination of the key residues by means of mutational studies is a robust and definitive method, but its tedious nature poses a large limitation due to the necessity of preparing and experimentally probing a number of protein mutants. Several other approaches have been proposed for the estimation of the relative importance of a specific residue in a structure of a protein.[3] These are usually based on sequence conservation, contact order or pertinence to a secondary structure element. Up to now, however, none of them have been shown to be able to identify the key residues in a protein with reliability.

Grant sponsor: Czech Science Foundation; Grant numbers: 203/05/0009, 203/06/1727, 203/05/H001; Grant sponsor: Ministry of Education, Youth and Sports (MSMT) of the Czech Republic; Grant number: LC512; Grant sponsors: Z40550506, MSM6198959216.
*Correspondence to: Jiří Vondrášek, Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic. E-mail: jiri.vondrasek@uochb.cas.cz.
Received 30 August 2007; Revised 31 October 2007; Accepted 29 November 2007
Published online 00 Month 2008 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21938

AQ4

It is well known that the structurally important residues are highly conserved.[4] Hence, in proteins with a sufficient number of known homologs or analogs, sequential or structural alignments might be used to identify the important residues and their positions based on their conservancy. This is, however, usually not available for *de novo* designed proteins or proteins with no known homologs. Moreover, the accuracy of the alignment methods is limited, because there might be discrepancy between sequential conservancy and the importance of a residue in its structural context. It is a known fact that it is difficult to distinguish between structurally and functionally conserved residues.[4]

Several approaches have been proposed for the theoretical prediction of the key residues. One such method is the Gaussian Network Model[3,5,6]—a simplified normal mode analysis where the protein chain is described as beads connected by harmonic springs. In this model, the slowest (global) fluctuations correspond with protein functions; the fastest modes are dependent on the local environment. However, none of these modes can distinguish the key residues. First, the model only takes into account the C_α atoms, therefore the interresidue interactions cannot be described reasonably well. More importantly, the key residues do not necessarily correspond to those with the lowest mean-square fluctuation, because their role is to preserve the overall topology rather than rigidity of the whole structure; they can thus be simultaneously rigid and flexible.[2,3]

Another theoretical method is the distance-based contact theory. As the key residues can maintain the 3D structure of the protein by linking it together by a number of nonlocal interactions, they should also have the highest number of contacts, which are usually defined by distance criteria. However, it has been shown that the number of contacts varies greatly with the distance threshold.[3] Moreover, the interactions strongly depend on the accurate geometry arrangement. For example, a salt bridge between two oppositely charged residues, traditionally seen as an important stabilizing element, may in some cases be destabilizing.[7,8] This, along with the fact that different types of interactions between residues have different geometry parameters, means that the distance-based contact theory (as well as any other simple geometric criteria) is not capable of identifying the key residues.

It can be seen that the prediction of residue importance from a structural point of view remains a complex and challenging task. The method able to localize successfully the key residues in a protein structure has to take into account the physically sound description of interresidue interactions. One method that fulfills this criterion is the analysis of the interaction energy potential. The total interresidue interaction energy defined using a CHARMM27 force-field potential with the Generalized Born/Surface Area implicit solvent model has been used for identifying the key residues in a Gbeta domain of a multidomain protein transducin,[3] a propeller-like protein composed of seven blades. In this investigation, the most strongly interacting residues, therefore identified as the key ones, were mainly the charged residues linking together the beta sheet blades.

This would not necessarily be true in a small globular single domain protein. Among the various interactions stabilizing the structure of a globular protein, a significant role has to be assigned to the interactions between the residues composing the protein core. Not only do these residues bring a substantial energetic contribution to the overall stabilization of the structure,[9] but their aggregation during the hydrophobic collapse event can also be considered as the driving force in protein folding process.[10] The small number of residues that are necessary to define the native-like topology of the folding TS via their tertiary contacts typically correspond to the hydrophobic core.[2,11] The core residues are usually also more conserved across protein families or folding motives than other residues.[12–14]

Therefore, it seems plausible to assume that the key residues in a small single domain protein will be mainly those constituting the hydrophobic core. We can conclude that the key residues should be characterized by a large number of stabilizing nonlocal interactions and consequently also by the highest total stabilization energy.

An important question in the determination of the energy contribution of a particular residue to the protein stability is the accuracy of the employed method. Since the hydrophobic residues, which play an important role in the protein fold stability, interact mainly by forces of dispersive nature, it is crucial to select methods able to describe the dispersion correctly. Most of the calculations for biologically relevant systems are performed by molecular mechanics utilizing empirical potential force fields. Although this method can be quite reliable in many cases,[15] one has to be also aware of its shortcomings.[16,17] It is therefore necessary to perform a benchmark study of the performance of a selected force field to justify the method prior to its routine usage. Nowadays, it is feasible to perform nonempirical, *ab initio* quantum chemical calculations for systems of biological relevance, but it is a known fact that some of the methods, like Hartree-Fock (HF) or Density functional theory (DFT) do not cover dispersion at all. This failure is important as it prevents these methods from being used in any biological and biophysical applications. One option is to utilize either other high-level *ab initio* calculations covering dispersion energy, such as the popular second-order Møller-Plesset perturbation treatment (MP2) covering a substantial part of correlation energy. Another possibility is to use DFT methods corrected and parameterized to cover the dispersion term.[18–20]

We selected the MP2 method, which is known to describe interaction energy of amino-acid residues in proteins[9] well, as the reference method. We are certainly

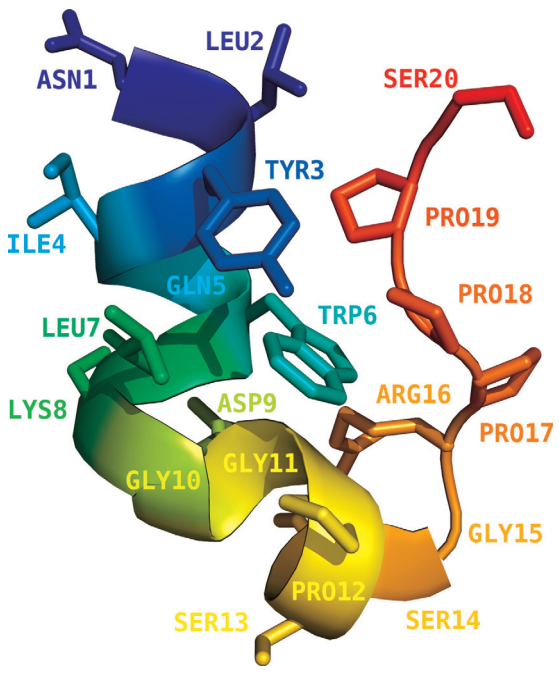
 DOI 10.1002/prot

J_ID: Z7E Customer A_ID: 21938 Cadmus Art: PROT21938 Date: 2-JANUARY-08 Stage: I Page: 3

aware of the fact that MP2 method is not accurate for noncovalent interactions. This is due to the fact that MP2 when combined with extended basis sets (or even using the complete basis set limit) overestimates the dispersion energy.[21–24] This overestimation is eliminated when passing to CCSD(T) procedure. The other possibility is to perform the MP2 calculation with medium basis set and thus rely on compensation of errors. In the present article we selected this strategy and MP2 calculations were performed with aug-cc-pVDZ basis set (see also Ref. 9) The other possibility is to use DFT procedure augmented by empirical dispersion energy which, due to the parametrization of dumping function connected with dispersion energy, provides accurate interaction energies comparable to the accurate CCSD(T) values. Finally, the standard force field parm94[25] implemented in Amber package, commonly used for molecular modeling of proteins was also applied.

The DFT-D method was parameterized toward a set of stabilization energies of various molecular clusters (H-bonded, stacked, mixed) determined at a highly accurate CCSD(T) level.[19] The method, yielding very accurate stabilization energies, structures and other properties (e.g. vibrational frequencies), is computationally not demanding, which allows it to be used even for extended clusters. It must be mentioned here that the CCSD(T) method itself is extremely time consuming and thus applicable only to small systems.

The model system used in this study is the Trp-cage de novo miniprotein (see Fig. 1), where the role of the key residue can be clearly ascribed to the Trp6 residue. The protein was designed by mutating a sequence of a common Trp-cage fold XFXXWXXXXGPXXXXPPPX. The size of the model system (only 20 amino-acid residues) facilitates the usage of the high-level *ab inito* reference method. To test the IEM method and its reliability initially, we applied it to the Trp2Cage miniprotein, a P12W mutant of the Trp-cage. Further, we applied the method to capture the difference in stabilization energy which can be assigned to higher/lower thermostability of rubredoxin variants.



COLOR

Figure 1

The Trp-cage miniprotein (PDB code 1L2Y) and its sequence.

METHODS

Strategy of calculations

In each studied protein, we evaluated the interaction energy between pairs of amino-acid residues. For this purpose, we split the polypeptide chain into fragments according to the procedure described below. This enabled us to calculate the interaction energy between all pairs of nonneighboring residues, or in all pairs of *n*th and *m*th ($m > n + 2$) residue during empirical calculation in the Amber force field (FF).

Interaction energies (in kcal/mol) were then collected in a matrix, and hereinafter we use the abbreviation interaction energy matrix (IEM). An important entry is not only the interaction energy between two (nonneighboring) amino-acid residues but also the sum of interaction energies of a particular amino acid (with the exception of its neighbors) in the system investigated. To facilitate orientation in the IEM, the following color code is adopted based on the value of the interaction energy: the dark blue ($(-\infty) - (-5.0)$ kcal/mol), the light blue ($(-5.0) - (-0.5)$ kcal/mol), the gray ($(-0.5) - 0.5$ kcal/mol) and the yellow ($0.5 - \infty$ kcal/mol).

Trp-cage as a model system

The model system is the Trp-cage miniprotein (PDB code 1L2Y), a 20-residue construct and the smallest protein-like cooperatively folded tertiary structure known to date. Its sequence NLYIQWLKDGGPSSGRPPPS was designed by a mutation of a common Trp-cage fold sequence XFXXWXXXXGPXXXXPPPX. In the native structure where Tryptophan 6 is "encaged" between an alpha helix and a β-hairpin (see Fig. 1), the tryptophan residue side-chain is surrounded by three proline residue side-chains and other hydrophobic residues. The proline residues contribute to the stability of the helix.[26] The tryptophane residue plays a role of a "connecting hub" in the structure, which glues together the alpha helix and the β-hairpin. Such an important structural feature of

ID: srinivasanv Date: 2/1/08 Time: 19:51 Path: J:/Production/PROT/Vol00000/070671/3B2/C2PROT070671

the tryptophan residue should also be demonstrated by unique physical–chemical properties of the side-chain and the number of its non-covalent interactions and enormous contribution to the overall stabilization.

To verify the range of sensitivity of the suggested concept, we selected two systems to demonstrate the advantages and drawbacks of the IEM method. The first system was naturally closest to the model system—a more stable variant of the Trp-cage; the second system was mesophilic and hyperthermophilic variants of the small protein rubredoxin. The hypothesis that stabilization energy inside a protein has a direct connection with the overall stability (temperature stability) should be then proved or disproved by the IEM method straightforwardly.

Trp2Cage

The Trp2Cage is a P12W mutant of the Trp-cage miniprotein, designed by Saven and coworkers[27] using the Monte Carlo (MC) search. This mutation is known to contribute to the stability of the structure enormously, raising its T_m by 15°C.

We adopted two strategies to model the system and verify the sensitivity of our method. First, we fitted the designed rotamer of Tryptophane at the position 12 of the original Pro residue in the Trp-cage structure taken from PDB. Second, we took the MC model structure of the Trp2Cage (with Tryptophane at position 12) obtained by Saven *et al.* and mutated the Trp12 into Pro residue. All geometries were optimized using parm94 FF in Amber prior to the calculation of IEM.

Rubredoxin variants

We selected two variants of a small globular protein from the rubredoxin family—hyperthermophilic (PDB code—1BRF) from *Pyrococcus furiosus* and mesophilic (PDB—code1RB9) from *Desulfovibrio vulgaris*. The hyperthermohilic variant is known to be stable above 100°C; whereas for the mesophilic one, the temperature stability decreases substantially over 40°C. Despite the fact that the two systems are structurally almost indistinguishable from one another, there are a few very important differences in the structure. The RMSD difference for 7 residues in the hydrophobic core is quite high (~1.5 Å), reflecting the fact that in the hyperthermophilic variant of rubredoxin the amino-acid side-chains are more tightly packed. Another difference is the presence of Tryptophane residue in the hyperthermophilic variant composing the hydrophobic core as against the presence of Tyrosine in mesophilic rubredoxin. Both structures were taken from PDB, hydrogen atoms were added and optimized prior to the calculation of the IEM by the DFT method.

Fragmentation

So that the interaction energy could be calculated, the peptide chain has to be split into fragments. We used two different fragmentation procedures:

a. The C_α—C(O) bond (the least polar bond along the chain) was cut and the fragments were capped with H atoms. This scheme was used with RI-MP2 and RI-DFT-D methods.
b. The peptide bond was cut and the fragments were capped with acetyl group at the N terminus (H_3C—C=O—) and with N-methyl at the C terminus (—NH—CH_3). This is standard method of amino acids capping by AMBER program and it was used for empirical force-field calculations.

Calculation of interaction energy

In both RI-MP2 (Resolution of Identity Møller-Plesset method) and DFT-D[19] calculations, we first created all possible pairs of residues and optimized hydrogens at the HF/6-31G(d) level in Gaussian03 (while keeping the heavy atoms fixed at the positions from the PDB structure). Then we calculated the interaction energies for all residue pairs at the **MP2**/aug-cc-pVDZ (counterpoise (CP) corrected) or **DFT-D**|TPSS/TZVP level in Turbomole. Both types of calculation were performed either in the gas phase or using the COSMO model of solvent. We used the COSMO method in Turbomole Program package with the default values of the parameters.[28] In the empirical potential calculations, we optimized hydrogens in separate residues and then calculated the interaction energy for all pairs using the parm94 FF.

Sometimes we used the term "stabilization energy" instead of "interaction energy." Stabilization represents the negatively conceived value of interaction energy.

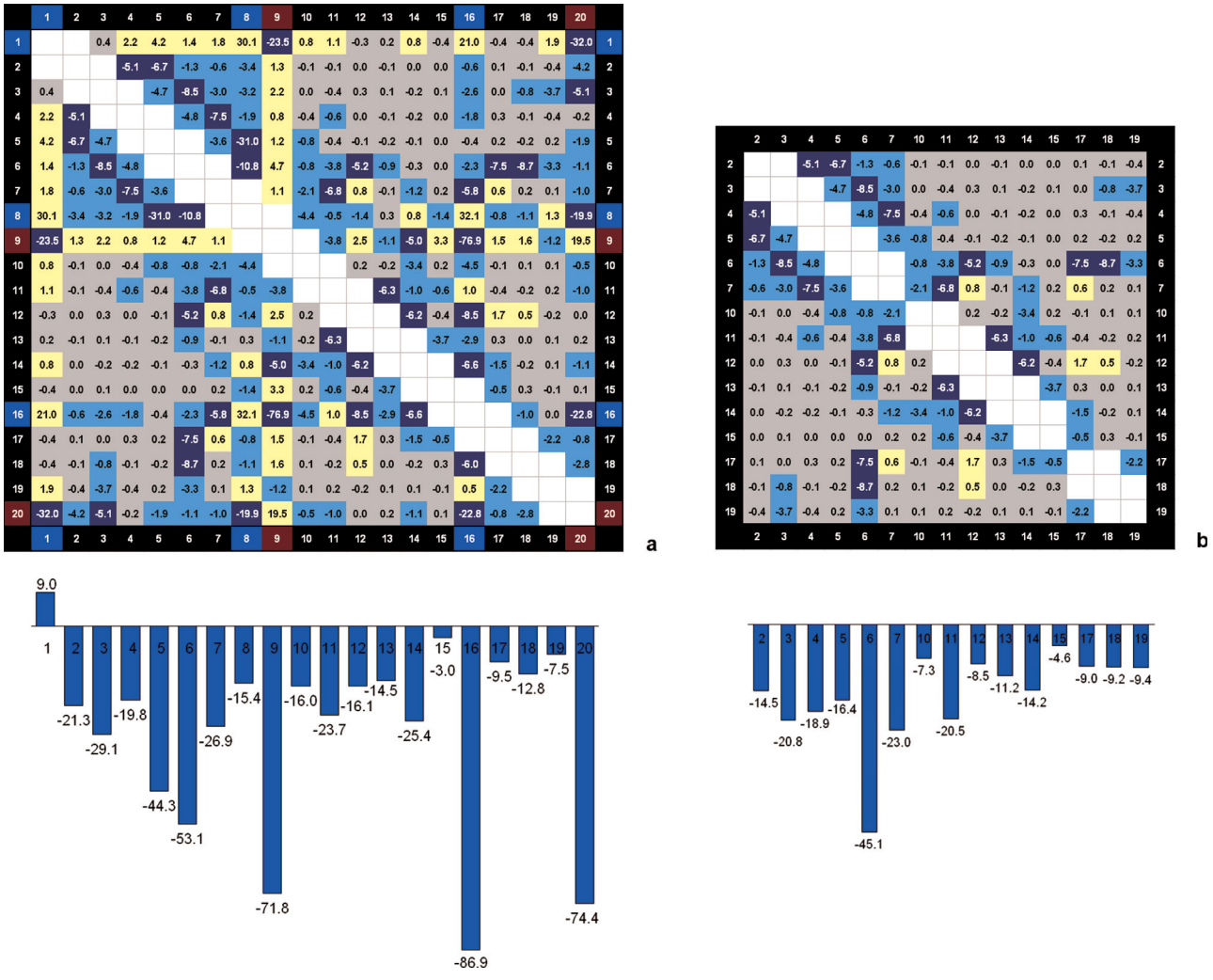
RESULTS AND DISCUSSION

Gas phase calculations on the Trp-cage system—comparison of MP2, DFT, and FF

The gas phase calculations were performed at the MP2 level to set up a benchmark for further evaluation of DFT and empirical potential methods. Nevertheless, it is known that the interactions between charged residues, as well as between neutral and charged residues, are strongly overestimated in the gas phase in comparison with the protein-like or water environment due to unshielded Coulomb interactions.[29] Figure 2a shows that our MP2 gas phase calculations also manifest this behavior. The largest stabilization contributions are brought by residues Asp9, Arg16 and Ser20 (negatively charged —COO^- terminus). F2

DOI 10.1002/prot

J_ID: Z7E Customer A_ID: 21938 Cadmus Art: PROT21938 Date: 2-JANUARY-08 Stage: I Page: 5



COLOR

Figure 2

*MP2 gas phase IEM (**a**) for the whole set of residues and total interaction energy (in kcal/mol) for individual residues and (**b**) for the reduced set of residues. The color code is based on the value of the interaction energy: the dark blue ((−∞) − (−5.0) kcal/mol), the light blue ((−5.0) − (−0.5) kcal/mol), the gray ((−0.5) − 0.5 kcal/mol) and the yellow (0.5 − ∞ kcal/mol).*

Since most of the stability-important residues in the Trp-cage protein are neutral (the consensus sequence of synthesized Trp-cage variants is X**F**XXWXXXXGPXX XXPPPX, the conserved residues in the exendin-4 fragment used to design a Trp-cage according to the ConSeq[4] sequence alignment are F3, V/I4, W6, L7, G10, and P12), we decided first to evaluate interactions within a reduced set which contained only the uncharged residues (The charged residues Lys8, Asp9, and Arg16 as well as the terminal charged residues Asn1 and Ser20 were not considered). The results are shown in Figure 2b, depicting the MP2 gas phase interaction matrix for such a set. Most of the interactions within this matrix have an attractive character (i.e. contribute to protein stability), with no pair exhibiting strong repulsion. Only 2 pairs (Leu7-Pro12, Leu7-Pro17) possess higher repulsion than 0.5 kcal/mol. Some of the interactions of the aliphatic and aromatic residues can be comparable in strength to classical H-bonding (e.g. the interaction Tyr3-Pro19, amounting to −3.7 kcal/mol). Moreover, residues with larger side-chains (e.g. aromatic) can simultaneously participate in several strong interactions; the highest number of attractive interactions stronger than 1 kcal/mol possesses Trp6. Its interaction energy with Pro18 and Tyr3 is larger than 8 kcal/mol, and only slightly weaker stabilization energy was found for its interaction with Pro17. Again, these stabilization energies are very large; a strong H-bond in the water dimer is about 5 kcal/mol. Two complexes (Tyr3-Trp6, Trp6-Pro17) possess N—H···O main-chain–main-chain and main-chain–side-chain

ID: srinivasanv Date: 2/1/08 Time: 19:51 Path: J:/Production/PROT/Vol00000/070671/3B2/C2PROT070671

H-bonds, respectively, but the complex 6–18 does not have any H-bond; the enormous stabilization (8.7 kcal/mol) originates solely in stacking! It must be mentioned here that only one subsystem (Trp) contains aromatic electrons, the other one (Pro) is a heterocyclic aliphatic system. So as to answer the question whether the stabilization in the other two pairs comes mainly from H-bond, we have studied two model systems: the interaction of Tyr3-Trp6 side-chains, modeled from the C_β atom, for example without the H-bond, amounts to −2.0 kcal/mol; in the Trp6-Pro17 complex, where the H-bond has been removed by replacing the —CHO group in proline by $—CH_3$ group, the interaction energy amounts to −6.3 kcal/mol.) Besides these three very attractive interactions, Trp6 exhibits further four additional interactions with stabilization higher than 3 kcal/mol. The total interaction energy provided by one residue in the reduced set is always negative. The highest value was obtained for Trp6 and exceeded 45 kcal/mol, which is an extreme stabilization coming from one residue (in the gas phase). The other most stabilizing residues are: Leu7 (23.0 kcal/mol), Tyr3 (20.8 kcal/mol), and Gly11 (20.5 kcal/mol). Evidently, aromatic Trp6 forms a "glue," which dominantly contributes to the stability of the Trp-cage. Other aromatic but also aliphatic residues also contribute, but their role is less important. The C-terminal proline residues—structurally conserved elements of the Trp-cage fold—contribute each by ∼9 kcal/mol to the overall stability. This confirms that gas phase calculations of the stabilization energy provided by individual residues are able to identify the key stabilizing residues in the protein structure of the Trp-cage.

Having calculated the IEM using the reference MP2 method, we used the results to evaluate the performance of the DFT and empirical potential methods. The comparison of MP2 gas phase results with the DFT and empirical potential is shown in Figure 3. The values are sums of the pair-wise interaction energies for each neutral residue in the structure. The DFT results agree very well with the MP2 values. The results of the empirical force-field calculations were overestimated approximately by a scaling factor of 2. Such large overestimation of the stabilization energy is exclusively the result of different fragmentation procedure used for this method, which yielded larger fragments and thus also larger stabilization energies of the residue pairs. Having been corrected by this scaling factor, the results qualitatively agree with both the MP2 and the DFT-D values (see Fig. 3). An important finding is that all the three methods identified the same residues (e.g. Trp6, Tyr3, Leu7, and Gly11) as the key ones in the reduced set of residues.

The DFT-D interaction energies should be by definition (and parameterization) close to the CCSD(T)/CBS ones and hence more accurate than the MP2 ones (We are, of course, aware that the DFT-D procedure contains empirical constants and is thus not a pure *ab initio* method). The DFT-D interaction energy is constructed as a sum of DFT interaction energy and dispersion energy. The dispersion energy determined by the C_6/R^6 expression (see the original Ref. [19]) agrees surprisingly well with the dispersion energy determined by the Symmetry Adopted Perturbation Treatment. Hence, the DFT-D interaction energy can be easily decomposed into the DFT part (covering roughly electrostatic, induction and exchange-repulsion energies) and the dispersion part.

Total DFT-D interaction energy and its dispersion part are presented in Figure 4. Evidently, dispersion energy forms an important and with some residues even dominant part of total stabilization energy. This is even more evident from the histogram shown in Figure 4a, and the following conclusions can be made: (i) stabilization of the most stable (neutral) residue Trp6 is dominantly (76%) caused by dispersion energy. If the partner is an aromatic system (i.e. interaction 6-3), the role of dispersion energy is, following expectations, even higher (79%). Very surprising is, however, the dominant role of dispersion for the interaction of Trp6 with a cyclic, aliphatic proline. In the case of Pro18, the dispersion forms 94% of the total stabilization energy. This value is even higher (125%) for Pro19, where DFT interaction energy (the difference between total interaction energy and dispersion energy) is repulsive, and total stabilization is entirely realized by dispersion energy; (ii) stabilization of the next three most stable residues (Leu 7, Gly 11, and Tyr 3) is also due to dispersion energy (56, 46, and 72%, respectively); (iii) dispersion energy contributes, again following expectation, almost negligibly to the stabilization of ionic residues (1, 8, 9, 16, 20); (iv) having summed total stabilization energies of each residue, we obtain a total stabilization energy of 628 kcal/mol. The dispersion energy contributes to this energy by 33%. When only nonionic residues are considered, this portion increases to 53%; (v) dispersion energy of prolines (18 and 19) with aromatic Tyr3 and Trp6 is very large. In the case of the 19-3 and 19-6 pairs, it is even larger than total stabilization energy, and in the case of 18-6 pair, it forms a dominant part of total stabilization energy.

The role of solvent—comparison of DFT and empirical potential methods

To describe the interactions involving polar or charged residues correctly, we need to include the effect of the environment in the calculations. Last but not least, the aliphatic parts of charged side-chains are also involved in van der Waals stabilization, and their contribution is therefore not purely electrostatic. The environment was included in DFT calculations by employing the implicit solvent model (COSMO).[28] We used the value of dielectric constant $\varepsilon = 80.0$. We are aware of the fact that the dielectric constant of protein is much lower, mostly between 4 and 10. The present values thus represent the

ID: srinivasanv Date: 2/1/08 Time: 19:51 Path: J:/Production/PROT/Vol00000/070671/3B2/C2PROT070671

J_ID: Z7E Customer A_ID: 21938 Cadmus Art: PROT21938 Date: 2-JANUARY-08 Stage: I Page: 7

COLOR

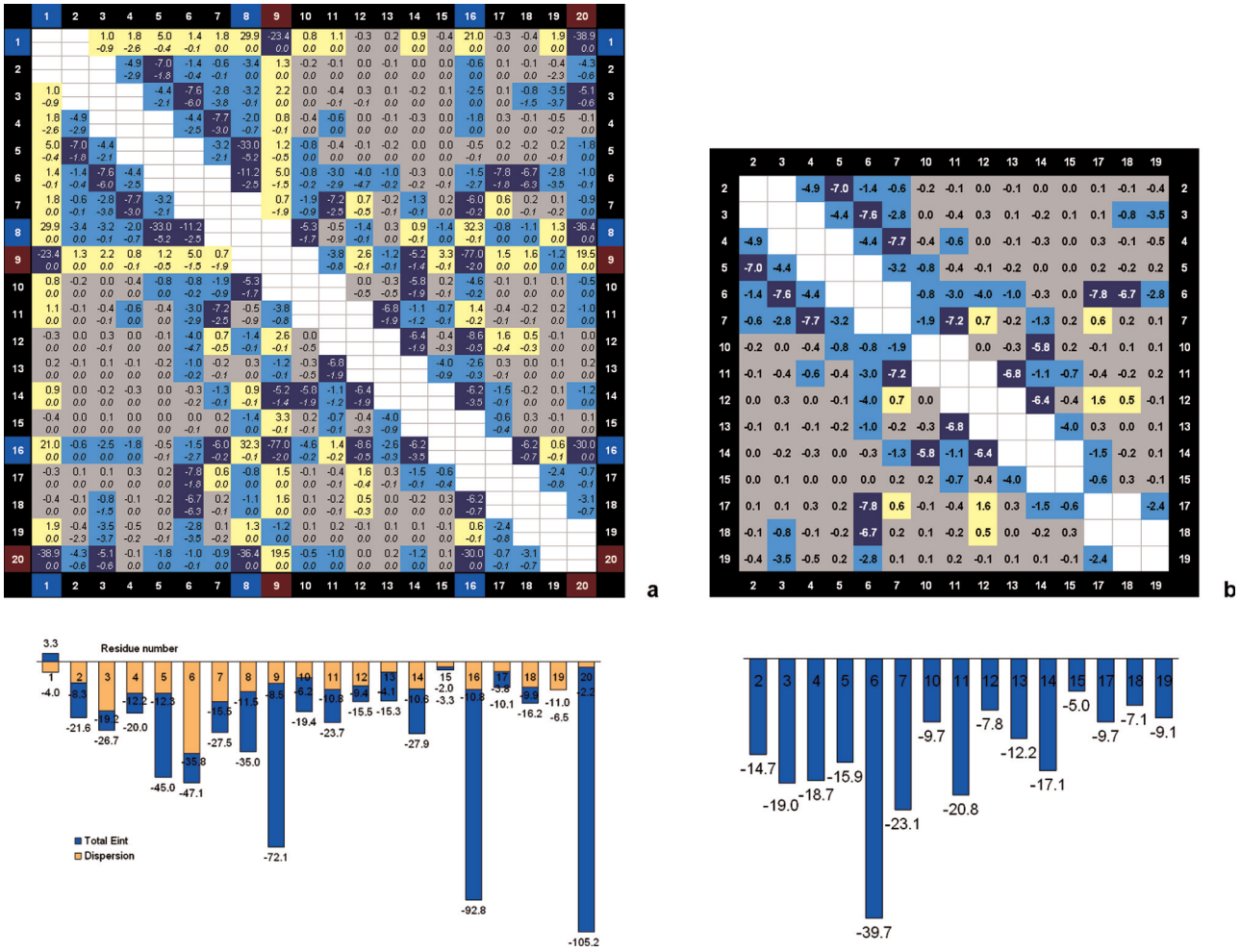


Figure 3

*DFT gas phase IEM (**a**) for the whole set of residues and total interaction energy for individual residues and (**b**) for the reduced set of residues. The color code is based on the value of the interaction energy: the dark blue ((−∞) − (−5.0) kcal/mol), the light blue ((−5.0) − (−0.5) kcal/mol), the gray ((−0.5) − 0.5 kcal/mol) and the yellow (0.5 − ∞ kcal/mol).*

lower limit of the actual interaction strength; the actual values should thus be closer to the "gas-phase" values.

F5 The results are shown in Figure 5. The contributions of some of the charged residues provided by the DFT/COSMO calculations remained important (especially Lys8), but, typically due to a single strong interaction (e.g. salt bridge Asp9-Arg16, H-bonded interaction Gln5-Lys8), their absolute values decreased substantially when compared with the gas phase values. However, this decrease was not uniform for all the charged residues, for example the stabilization contribution of Lys8 decreased from its gas phase value of 35.0 to 9.3 kcal/mol, the Asp9 contribution decreased from the gas phase value of 72.1 kcal/mol even to a negative value of −0.2 kcal/mol. None of the charged residues possessed higher total stabilization energy than the Trp6 residue, which also maintained the highest number of stabilizing interactions stronger than 1.0 kcal/mol ($n = 6$). An important fact is that all residues except for 1, 9, and 15 bring stabilizing contribution.

To evaluate the performance of the empirical potential and its reliability for the description of the residue–residue interactions in the solvent environment, we utilized the generalized born (GB) solvent model with the standard value of dielectric constant of 78.5 for water. The empirical potential values were again higher, than the DFT/COSMO results; the proportionality constant was 1.436. On the other hand, the overall trend was similar to DFT/COSMO results. We can therefore use the empirical potential with the GB model to assess qualitatively the contributions of individual residues to the protein stability. The comparison of DFT/COSMO results with the scaled empirical potential GB results is shown in Figure 5.

ID: srinivasanv Date: 2/1/08 Time: 19:51 Path: J:/Production/PROT/Vol00000/070671/3B2/C2PROT070671

COLOR

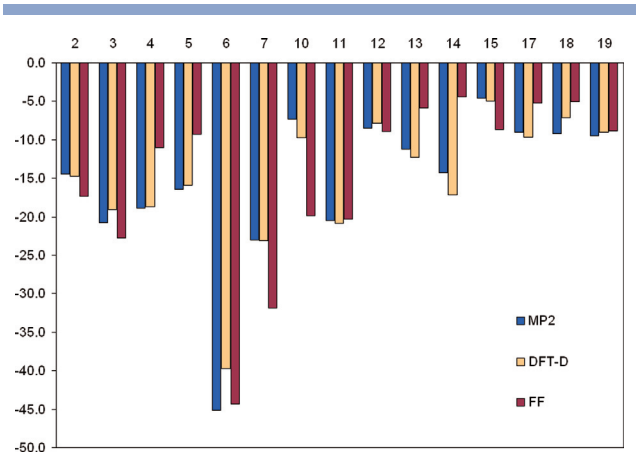


Figure 4

Trp-cage—a comparison of total interaction energy for individual residues calculated by MP2, DFT, and empirical potential in the gas phase.

Application test case no. 1—Trp2Cage

To show the usefulness of our method, we tested its ability to reproduce the stabilization of the Trp2Cage,[27] a recently reported P12W mutant of the Trp-cage, designed by Monte Carlo (MC) search of side-chain rotamers and experimentally shown to have a 15°C higher melting temperature (T_m) as compared with the original Trp-cage structure. However, its geometry has not been determined up to now either by NMR or by X-ray crystallography.

So as to overcome this problem, we generated the Trp2Cage geometry by using the Trp-cage geometry and replacing the P12 residue manually by Trypthophane (for computational details, see the Methods section.). We than compared the IEMs calculated for the original Trp-cage and its P12W mutant using empirical potential with the GB solvent model. The comparison of the total stabilization energies provided by individual residues in both studied structures is shown in Figure 6a, the differences between them are shown in Figure 6b. To verify the results, we also compared the IEM for the Trp2Cage MC model geometry and its W12P mutant (the Trp2Cage MC model geometry where we manually replaced the W12 residue by the original Proline residue). The results are also shown in Figure 6a.

F6

The structures originating in the Trp2Cage MC model geometry have a much smaller sum of stabilization energy than the two structures with the Trp-cage geometry (~13 kcal/mol). It seems to be apparent that the MC model of the Trp2Cage is not at its energy minimum. The sum of stabilization energy is by 0.8 kcal/mol larger in the Trp-cage P12W mutant as compared with the original Trp-cage. In the case of the Trp2Cage W12P mutant, the sum of stabilization energy is by 0.2 kcal/mol lower than in the Trp2Cage. This is a very good agreement between the two approaches, when we consider the high number of contributing pair-wise interaction energies in each of the compared sums and the propagation of only one mutation.

Application test case no. 2—Rubredoxin variants

Our aim was to apply the IEM method in a practical way when physico-chemical differences are known and experimental data provide a solid basis for correlation with calculations. In this example, we show the ability of our method to distinguish the key residues in the structure of natural (nonartificial) proteins and also its ability to show the differences between mesophile and thermophile proteins. There is already a study dealing with this problem computationally on a comparable level of complexity.[30] The results of our calculations for two rubredoxin protein variants are shown in Figure 7a. The hyperthermophilic rubredoxin (1BRF) provided total stabilization energy by 10 kcal/mol higher than its mesophilic variant (with the exception of the terminal residues 1 and 51). Another important feature of the stabilizing contributions is that the core residues (e.g. Y/W3, Y12, F29, W36, and F48—numbering according to 1BRF) add the most to this stabilization (by 6 kca/mol). The differences for the individual residues are shown in Figure 7b. The residues conserved in rubredoxin sequences according to Jenney and Adams[31] are the following: C5, C8, Y12, W36, C38, P39, C41, K45, and F48. We have made an independent analysis of residue conservancy according to ConSeq[4]—web-based bioinformatical tool for sequence alignment. This analysis provided the following amino acids supposed to be conserved in the sequences of rubredoxins—C5, C8, Y12, D13, E14, G17, T27, F29, L32, P33, W36, C38, P39, C41, K45, and F48. It is known that the four cysteines are necessary to hold one Fe2+ cation in tetrahedral arrangement and do not contribute to the stability of the fold and structure.[32] A careful look at the graph in Figure 7a shows that the IEM method provided the following residues which contribute the most to the overall stability of the protein—Y/W3; Y12; F29; L32; W36; K45; and F48. This is a very promising result in terms of combining sequential and structural/energy data into one framework. It is also apparent that the most important residues are part of the hydrophobic core, which makes sense when we take into account the stabilizing function of the protein core.

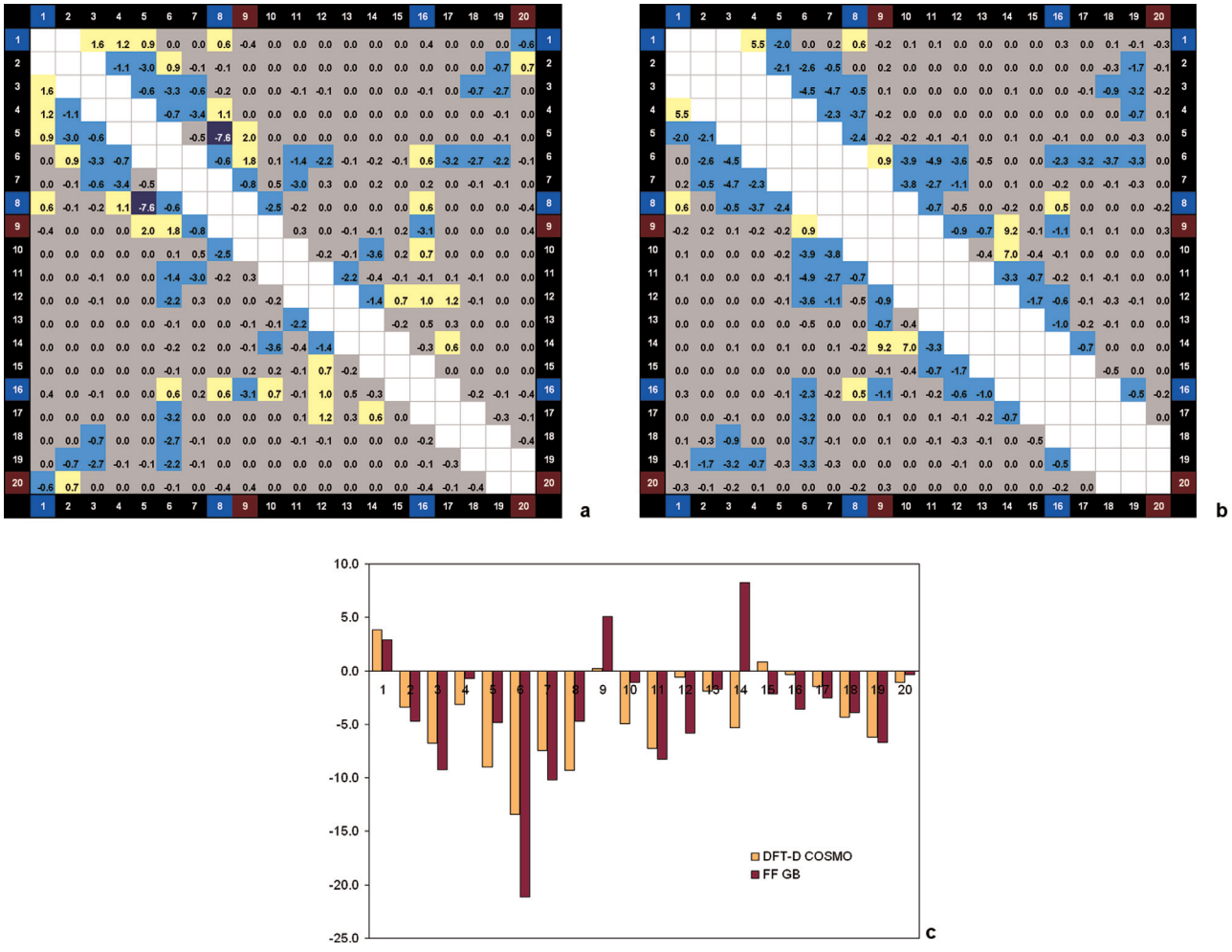
F7

CONCLUSIONS

The results presented here demonstrate how capable the IEM method is of adequately capturing the stabilizing contributions of the key residues to the overall stability of our model system—the Trp-cage miniprotein. The

DOI 10.1002/prot

J_ID: Z7E Customer A_ID: 21938 Cadmus Art: PROT21938 Date: 2-JANUARY-08 Stage: I Page: 9



COLOR

Figure 5

*Trp-cage—a comparison of (**a**) the DFT-D with COSMO IEM and (**b**) the empirical potential with the GB solvent model IEM and (**c**) histograms showing sums of interaction energies for individual residues in kcal/mol). The color code is based on the value of the interaction energy: the dark blue ((−∞) − (−5.0) kcal/mol), the light blue ((−5.0) − (−0.5) kcal/mol), the gray ((−0.5) − 0.5 kcal/mol) and the yellow (0.5 − ∞ kcal/mol).*

identified residues are mostly buried in the protein interior, composing its hydrophobic core.

We have also shown that gas phase calculations are sufficient to capture the stability contributions of individual residues at the roughest level of approximation and that they provide a useful first guess determination of important residues from a stability point of view. This approximation neglects contributions from charged amino acids and evaluates the stability for the reduced set of amino acids containing only neutral residues. Both DFT and rescaled empirical potential values reproduce the gas phase MP2 values at a reasonable level of accuracy. It is worth to mention here that standard treatment of isolated amino acids in AMBER is not ideal and provides systematically larger values of interaction energy for complexes where an amino acid is one of the interacting partners.

Dispersion energy contributes significantly to total stabilization energy, and its role with aromatic pairs is dominant. Surprisingly, an even more important role is played by dispersion energy for the interaction of prolines with an aromatic system.

To justify and adequately describe a system not far from real conditions, we had to take the role of solvent into account. The calculated values of pair-wise interaction energies for the whole set of residues including the charged ones provide a better estimation of the proportion between stabilizing contributions coming from different residues in an environment. We ascertained that the empirical potential with the GB model of solvent is able to reproduce the DFT/COSMO results—our reference method while being at the same time fast enough to be used for large biological systems. Empirical as well as

ID: srinivasanv Date: 2/1/08 Time: 19:52 Path: J:/Production/PROT/Vol00000/070671/3B2/C2PROT070671

COLOR

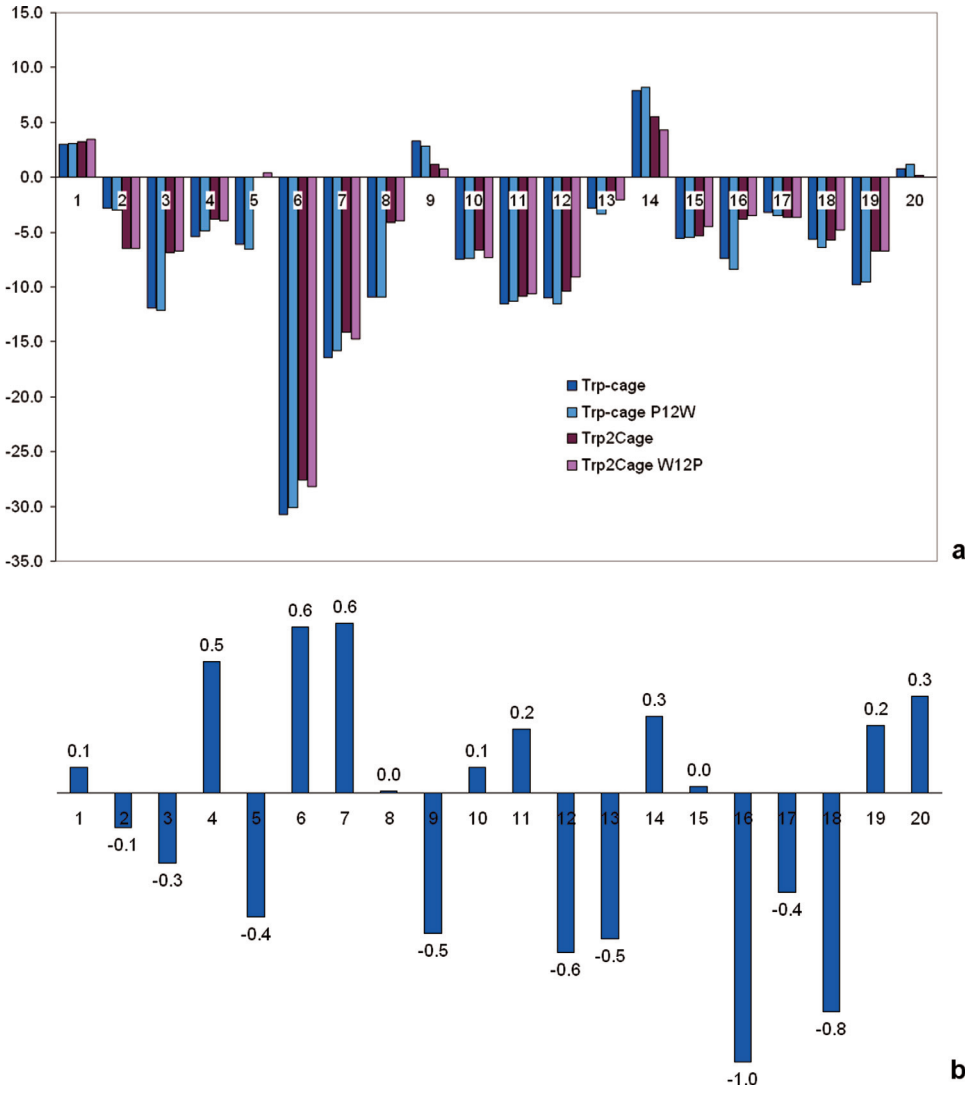


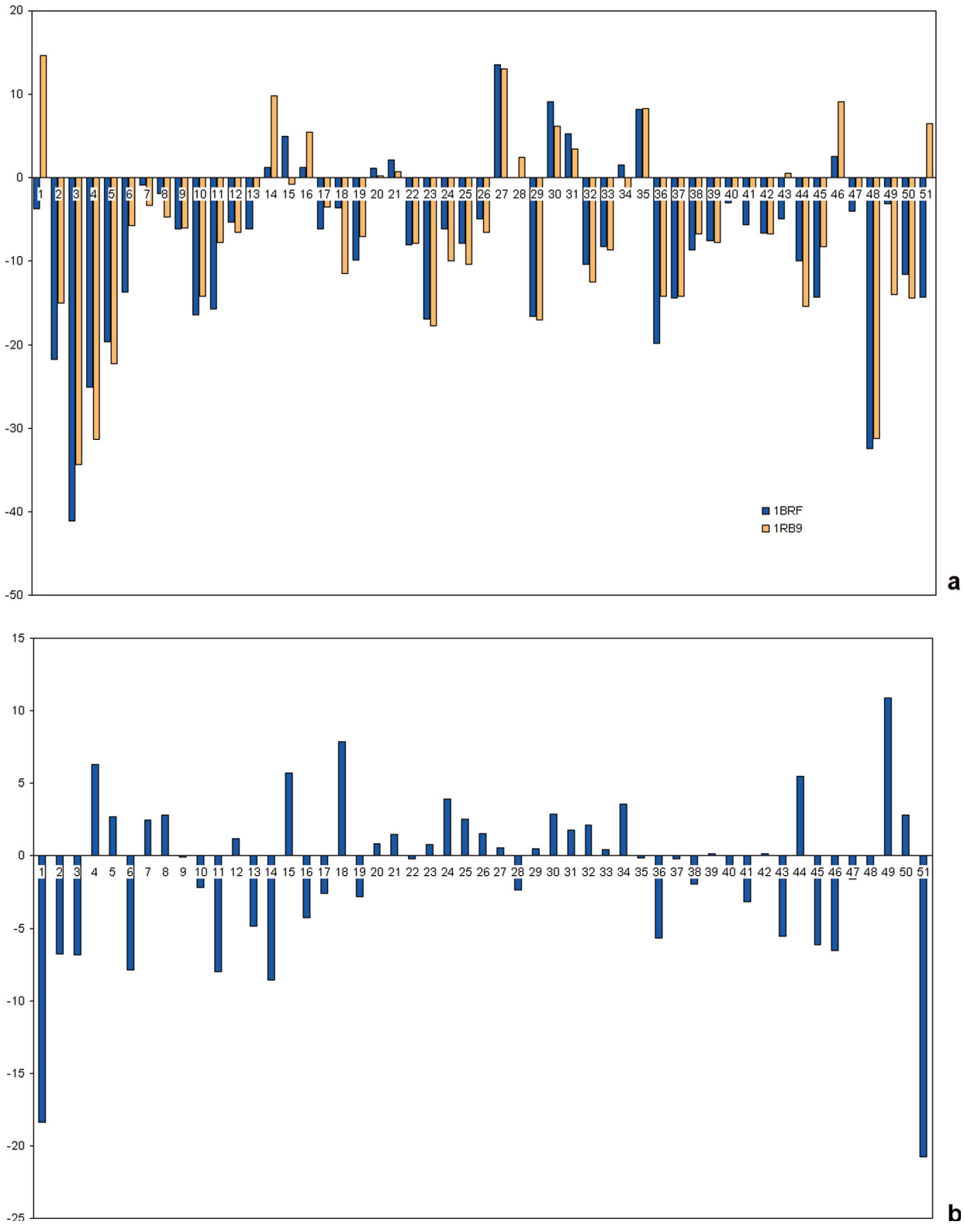
Figure 6

*Results of the Trp2Cage study calculated using the empirical potential with the GB solvent model: (**a**) Comparison of the sums of interaction energies (in kcal/mol) for individual residues and (**b**) Differences in the sums of interaction energy (in kcal/mol) for individual residues between the Trp-cage WT and the P12W mutant on the Trp2Cage X-ray structure.*

ab initio methods succeeded in identifying Tryptophane 6 as the key residue among all others and evaluate its contribution to the overall stability in its structural and sequential context. Applying the strategy of IEM on the Trp2Cage, we obtained a sound difference between the wild-type variant and the engineered one, in which the original Pro residue was replaced by a Tryptophane. The results show that the enthalpy difference is an indicator of stability change and can be used as a descriptor in mutagenesis experiments.

An interesting picture of the difference between two variants of the same protein was obtained by applying the IEM method on the hyperthermophilic and mesophilic variants of rubredoxin. It has been previously shown that mutation of an aromatic amino acid in the hydrophobic core[33] leads to a decrease of the thermostability and is also accompanied by a large drop of stabilization enthalpy. IEM confirmed that not only a total sum of pair-wise energy stabilization contributions differs significantly between the two variants but, which is more interesting, this trend is present even when only the hydrophobic core residues are taken into account. We have concluded that the method is reasonably sensitive to be applied in protein engineering routinely and is a valuable descriptor of a protein in a bioinformatical context.

DOI 10.1002/prot



COLOR

Figure 7

*Comparison of 1RB9 and 1BRF results calculated with empirical force field (numbering of residues according to 1BRF): (**a**) Comparison of the sums of interaction energies (in kcal/mol) for individual residues and (**b**) Differences in the sums of interaction energy (in kcal/mol) for individual residues between 1RB9 and 1BRF.*

ACKNOWLEDGMENTS

We would like to thank Tomáš Kubař and Jiří Černý for their script for IEM calculations in Amber package. We are also grateful to Prof. Jeffery G. Saven, Department of Chemistry University of Pennsylvania for providing us the TrpP12W coordinates for modeling the higher temperature stable version of the Trp cage.

REFERENCES

1. Anfinsen CB. Principles that govern folding of protein chains. Science 1973;181:223–230.
2. Vendruscolo M, Paci E, Dobson CM, Karplus M. Three key residues form a critical contact network in a protein folding transition state. Nature 2001;409:641–645.
3. Chen CJ, Li L, Xiao Y. Identification of key residues in proteins by using their physical characters. Phys Rev E 2006;73.
4. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, Casadio R, Ben-Tal N. ConSeq: the identification of functionally and structurally important residues in protein sequences. Bioinformatics 2004;20:1322–1324.
5. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 1997;2:173–181.
6. Haliloglu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. Phys Rev Lett 1997;79:3090–3093.
7. Kumar S, Nussinov R. Salt bridge stability in monomeric proteins. J Mol Biol 1999;293:1241–1255.
8. Kumar S, Nussinov R. Close-range electrostatic interactions in proteins. Chembiochem 2002;3:604–617.
9. Vondrasek J, Bendova L, Klusak V, Hobza P. Unexpectedly strong energy stabilization inside the hydrophobic core of small protein rubredoxin mediated by aromatic residues: correlated ab initio quantum chemical calculations. J Am Chem Soc 2005;127:2615–2619.
10. Tsai J, Gerstein M, Levitt M. Simulating the minimum core for hydrophobic collapse in globular proteins. Protein Sci 1997;6:2606–2616.
11. Lindorff-Larsen K, Vendruscolo M, Paci E, Dobson CM. Transition states for protein folding have native topologies despite high structural variability. Nat Struct Mol Biol 2004;11:443–449.
12. Finucane MD, Tuna M, Lees JH, Woolfson DN. Core-directed protein design. I. An experimental method for selecting stable proteins from combinatorial libraries. Biochemistry 1999;38:11604–11612.
13. Koshi JM, Goldstein RA. Mutation matrices and physical-chemical properties: correlations and implications. Proteins 1997;27:336–344.
14. Mirny LA, Shakhnovich EI. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. J Mol Biol 1999;291:177–196.
15. Daggett V. Long timescale simulations. Curr Opin Struct Biol 2000;10:160–164.
16. Chessari G, Hunter CA, Low CMR, Packer MJ, Vinter JG, Zonta C. An evaluation of force-field treatments of aromatic interactions. Chemistry 2002;8:2860–2867.
17. Duan GL, Smith VH, Weaver DF. Validation of the applicability of force fields to reproduce ab initio noncovalent interactions involving aromatic groups. Adv Quant Chem 2004;92.
18. Grimme S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. J Comput Chem 2004;25:1463–1473.
19. Jurecka P, Cerny J, Hobza P, Salahub DR. Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. J Comput Chem 2007;28:555–569.
20. Zhao Y, Truhlar DG. Design of density functionals that are broadly accurate for thermochemistry, thermochemical kinetics, and nonbonded interactions. J Phys Chem A 2005;109:5656–5667.
21. Zhao Y, Truhlar DG. Density functionals for noncovalent interaction energies of biological importance. J Chem Theory Comput 2007;3:289–300.
22. Zhao Y, Truhlar DG. Assessment of model chemistries for noncovalent interactions. J Chem Theory Comput 2006;2:1009–1018.
23. Jurecka P, Sponer J, Cerny J, Hobza P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes. DNA base pairs, and amino acid pairs. Phys Chem Chem Phys 2006;8:1985–1993.
24. Jurecka P, Hobza P. True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine center dot center dot center dot cytosine, adenine center dot center dot center dot thymine, and their 9-and 1-methyl derivatives: complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment. J Am Chem Soc 2003;125:15608–15613.
25. Weiner PK, Kollman PA. Amber—assisted model-building with energy refinement—a general program for modeling molecules and their interactions. J Comput Chem 1981;2:287–303.
26. Barua B, Andersen NH. Determinants of miniprotein stability: can anything replace a buried H-bonded Trp sidechain? Lett Pept Sci 2001;8:221–226.
27. Bunagan MR, Yang X, Saven JG, Gai F. Ultrafast folding of a computationally designed Trp-cage mutant: Trp(2)-cage. J Phys Chem B 2006;110:3759–3763.
28. Schafer A, Klamt A, Sattel D, Lohrenz JCW, Eckert F. COSMO implementation in TURBOMOLE: extension of an efficient quantum chemical code towards liquid systems. Phys Chem Chem Phys 2000;2:2187–2193.
29. Lund M, Jönsson B, Woodward CE. Implications of a high dielectric constant in proteins. J Chem Phys 2007;126:225103.
30. Chen CJ, Li L, Xiao Y. All-atom contact potential approach to protein thermostability analysis. Biopolymers 2007;85:28–37.
31. Jenney FE, Adams MWW. Rubredoxin from *Pyrococcus furiosus*. Methods Enzymol 2001;55.
32. Strop P, Mayo SL. Rubredoxin variant folds without iron. J Am Chem Soc 1999;121:2341–2345.
33. Vondrášek J, Kubaø T, Jenney FJ, Jr, Adams MWW, Kožíšek M, Èerný J, Sklenáø V, Hobza P. Dispersion interactions govern the strong thermal stability of a protein. Chemistry 2007. Published online on http://www3.interscience.wiley.com.

 DOI 10.1002/prot

AQ1: Kindly provide the volume or page numbers as appropriate for Ref. 3.

AQ2: Kindly provide the volume nos. for Refs. 17 and 31.

AQ3: Kindly note that the references are renumbered so as to make them sequential.

AQ4: Kindly check whether the corresponding author information is OK as typeset.

J|A|C|S
ARTICLES
Published on Web 02/05/2005

Unexpectedly Strong Energy Stabilization Inside the Hydrophobic Core of Small Protein Rubredoxin Mediated by Aromatic Residues: Correlated Ab Initio Quantum Chemical Calculations

Jiří Vondrášek, Lada Bendová, Vojtěch Klusák, and Pavel Hobza*

Contribution from the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic

Received September 6, 2004; E-mail: pavel.hobza@uochb.cas.cz

Abstract: The formation of a hydrophobic core of globular proteins is believed to be the consequence of exterior hydrophobic forces of entropic nature. This, together with the low occurrence of hydrogen bonds in the protein core, leads to the opinion that the energy contribution of core formation to protein folding and stability is negligible. We show that stabilization inside the hydrophobic core of a small protein, rubredoxin, determined by means of high-level correlated ab initio calculations (complete basis set limit of MP2 stabilization energy + CCSD(T) correction term), amounted to ~50 kcal/mol. These results clearly demonstrate strong attraction inside a hydrophobic core. This finding may lead to substantial changes in the current view of protein folding. We also point out the inability of the DFT/B3LYP method to describe a strong attraction between studied amino acids.

Introduction

Protein folding involves two critical elements, stability and specificity. The native structure of a typical protein is only 5−15 kcal/mol more stable than the unfolded state.[1] Hence, small differences in energy between multitudes of possible noncovalent interactions are summed up to provide the properly folded structure. To gain control of protein secondary and tertiary structure requires an understanding of how these noncovalent interactions provide both stabilization and specificity.[2]

Every globular and water-soluble protein has a hydrophobic core. The core is an arrangement of hydrophobic residues buried in the protein interior. The formation of a hydrophobic core, which is the driving process of protein folding in terms of energy, is connected with the existence of a folding nucleus,[3,4] a conserved region of protein which initiates the folding.[5,6] Evidence for a nucleation condensation mechanism can be found in the work of Itzhaki et al.,[7] which can be taken as one of the most important works in the field. Some forces, such as packing forces and H-bonding during protein folding in the context of the hydrophobically driven folding, are discussed in great detail by Zhu et al.[8] and Honig et al.[9]

Core formation is believed to be the consequence of exterior hydrophobic forces of entropic nature,[10,11] an example of classical hydrophobic effect[12] characterized by small contribution (repulsive or attractive) of complexation enthalpy. This, together with low occurrence of hydrogen bonds in the protein core, leads to the assumption that the energy (enthalpy) contribution of the core formation to protein folding is small or negligible.

Recent theoretical and experimental investigations of various types of noncovalent interactions have shown[13] that a rather large attraction could be gained not only from hydrogen bonding but also from other types of noncovalent interactions. Thus, the question arises of how strong are the stabilizing contributions of amino acids in a hydrophobic core. This question is of key importance for understanding the mechanism of protein folding as well as understanding protein secondary and tertiary structure.

The aim of the present work was to evaluate the stabilization energy of a model hydrophobic core based on a high-resolution X-ray structure of rubredoxin, a small soluble FeS protein (PDB code 1RB9). Stabilization energy was determined using high-level correlated ab initio calculations, specifically, as a sum of the complete basis set limit of the MP2 stabilization energy and CCSD(T) correction term.

Computational Model and Methods

Structure. Rubredoxin is a typical globular one-domain protein and contains a densely packed cluster of interacting residues centered around

(1) Branden, C.; Tooze, J. *Introduction to Protein Structure*, 2nd ed.; Garland Publishing: New York, 1999.
(2) Tatko, C. D.; Waters, M. L. *J. Am. Chem. Soc.* **2004**, *126*, 2028.
(3) Abkevich, V. I.; Gutin, A. M.; Shaknovich, E. I. *Biochemistry* **1994**, *33*, 10026.
(4) Fehrst, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1525.
(5) Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *J. Mol. Biol.* **2000**, *296*, 1183.
(6) Shakhnovich, E. I.; Abkevich, V.; Ptitsyn, O. *Nature* **1996**, *379*, 96.
(7) Itzhaki, L. S.; Otzen, D. E.; Fehrst, A. *J. Mol. Biol.* **1995**, *254*, 260.
(8) Zhu, B.-Y.; Zhou, N. E.; Kay, C. M.; Hodges, R. S. *Protein Sci.* **1995**, *2*, 383.
(9) Honig, B.; Yang, A. S. *Adv. Protein Chem.* **1995**, *46*, 27.
(10) Tanford, C. *Science* **1978**, *200*, 1012.
(11) Rose, G. D.; Geselowitz, A. R.; Lesser, G. J.; Lee, R. H.; Zehfus, M. H. *Science* **1985**, *229*, 834.
(12) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210.
(13) Müller-Dethlefs, K.; Hobza, P. *Chem. Rev.* **2000**, *100*, 143.

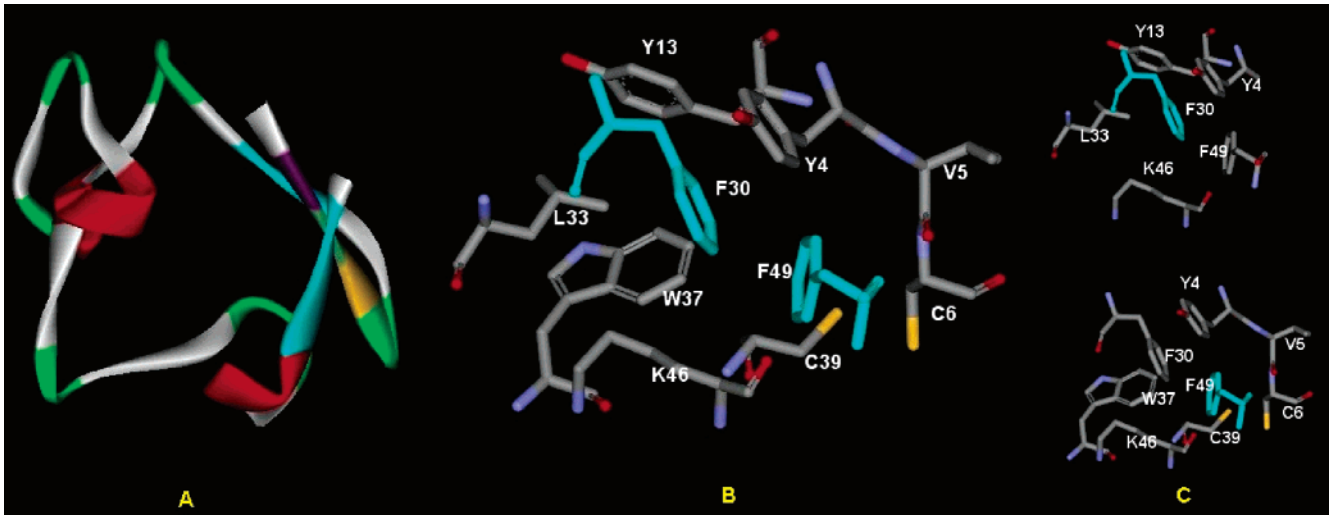


Figure 1. Rubredoxin. (A) Schematic view of the protein, (B) supercluster of F30 and F49, and (C) both subclusters individually.

two phenylalanines (F30 and F49) in the interior of the protein (Figure 1A,B). Localization of the cluster was determined by STING, a Web-based suite of programs (http://www.cbi.cnptia.embrapa.br), with distance criteria of 4.0 Å for the hydrophobic contacts of two amino acids. Water molecules are not present in the core, and thus, water is not directly participating in the core stabilization. The whole cluster was partitioned into two distinct clusters (named after the central residues, F30 and F49) and was further fragmented into well-defined, chemically distinct pairs of neutral amino acids (modeled as methylated aminoacyl residues). The central F30 and F49 phenylalanines thus interact with five (F49, K46, L33, Y13, and Y4) and seven (C39, C6, F30, K46, V5, W37, and Y4) amino acids, respectively. There is one H-bond ascribed to the F30 cluster (a classical CO···HN H-bond in the F30···L33 pair) and another two H-bonds are ascribed to the F49 cluster (a classical CO···HN H-bond in the F49···K46 pair, as well as an unusual CH···π interaction between the methyl group of the capped O terminus of V5 and the π system of the phenylalanine in the F49···V5 pair; cf. Figure 1B,C).

Interaction Energy. The heavy atom coordinates in all amino acid pairs were kept fixed at the positions from the X-ray structure (1RB9). Positions of the hydrogens were then optimized at the DFT/B3LYP/6-31G** level. The stabilization energies for all pairs of amino acids in clusters F30 and F49 were determined at the RIMP2 (resolution of identity Møller−Plesset method) level using a complete atomic orbital basis set (CBS) limit and, for a few selected pairs, also at the CCSD-(T) level. It should be mentioned here that the stabilization energies of the H-bonded model and stacked clusters evaluated at the CCSD(T) and CCSDT levels were practically identical,[14] which gave full confidence in using CCSD(T) calculations.

The CCSD(T)/CBS interaction energy was approximated as

$$\Delta E_{\mathrm{CBS}}^{\mathrm{CCSD(T)}} = \Delta E_{\mathrm{CBS}}^{\mathrm{MP2}} + (\Delta E^{\mathrm{CCSD(T)}} - \Delta E^{\mathrm{MP2}})|_{\mathrm{small\ basis\ set}} \quad (1)$$

The former term was determined using the Helgaker extrapolation scheme.[15] The Hartree−Fock and correlation MP2 energies necessary for the extrapolation were determined with aug-cc-pVXZ (X = D, T) basis sets. The CCSD(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.[16] All interaction energies were corrected for the basis set superposition error. The DFT calculations were performed using Gaussian03;[17] RIMP2 calculations were done by Turbomole,[18] and CCSD(T) calculations were performed using MOLPRO.[19]

Molecular Mechanics Calculations. All molecular mechanics calculations were performed using MOE (The Molecular Operating Environment), version 2004.03, software available from Chemical Computing Group Inc. (1010 Sherbrooke Street West, Suite 910, Montreal, Canada H3A 2R7; http://www.chemcomp.com).

Discussion of the Ab Initio Methods. The most common method of choice for the systems with sizes similar to the model intermolecular complexes studied in this work is either the DFT technique or the second-order Møller−Plesset (MP2) perturbation method.[20] The DFT methods provide reliable characteristics of isolated systems as well as H-bonded complexes, and due to their favorable size/CPU time ratio, they are extensively used also for biomolecular systems. The MP2 methods, covering a rather large portion of correlation energy, can be utilized not only for isolated systems and H-bonded complexes but also for stacked complexes. When used together with the resolution of identity technique (RIMP2 method[21]), a very favorable accuracy/CPU time ratio can be achieved.[22,23] The CBS limit extrapolated from the aug-cc-pVDZ and aug-cc-pVTZ energies is slightly underestimated with

(14) Pittner, J.; Hobza, P. *Chem. Phys. Lett.* **2004**, *390*, 496.
(15) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.
(16) Jurečka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89.
(17) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Pittsburgh, PA, 2004.
(18) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989** *162*, 165.
(19) *MOLPRO*, a package of ab initio programs designed by H.-J. Werner and P. J. Knowles, version 2002.1, R. D. Amos, A. Bernhardsson, A. Berning, P. Celani, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, P. J. Knowles, T. Korona, R. Lindh, A. W. Lloyd, S. J. McNicholas, F. R. Manby, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, R. Pitzer, G. Rauhut, M. Schütz, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and H.-J. Werner.
(20) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
(21) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359.

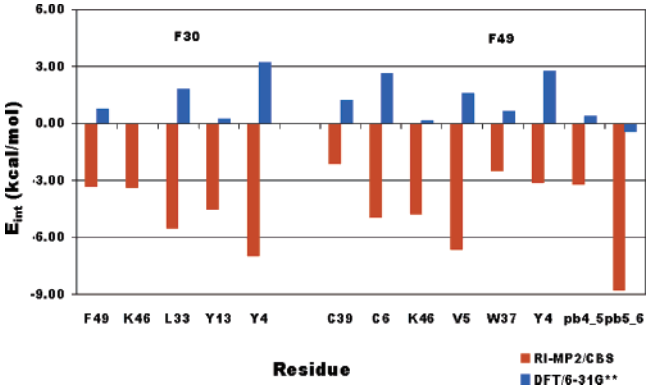


Figure 2. DFT and MP2/CBS interaction energies of F30 and F49 phenylalanines with selected amino acids from the rubredoxin core; DFT interaction energy of the F30···K46 pair is 0.

respect to the physically more justified CBS limit obtained from the aug-cc-pVTZ and aug-cc-pVQZ energies, but this underestimation is rather small (0.2−0.4 kcal/mol).[23] Higher-order correlation energy terms are important and should be included. In the CCSDT[24,25] calculations, all single, double, and triple and also part of quadruple and hextuple electron excitations are determined iteratively (i.e., up to the infinitive perturbation order). The CCSDT energies are very close to energies obtained from the full configuration interaction calculations and are thus approaching the true nonrelativistic energies. Though in the CCSD(T) calculations the triple excitations are determined only at the fourth perturbation order, the interaction energies, as mentioned in the text, are practically identical and the CCSD(T) method represents a method of choice even for larger biomolecular clusters.

Results and Discussion

The total stabilization energy of both clusters was determined as the sum of the pairwise stabilization energies of a central phenylalanine with the amino acids in its neighborhood. These energies were first determined at the frequently used DFT/B3LYP/6-31G** level. It should be mentioned here that DFT calculations are now widely used for biomolecular purposes and are even adopted in ab initio molecular dynamic simulations of the Car−Parinello-type. Figure 2 shows that 11 of 12 DFT pair interaction energies are repulsive, and the 12th one is only very slightly attractive. The DFT picture is thus consistent with the expected nature of interactions in a hydrophobic core with a low occurrence of hydrogen bonds. All pair interactions are either repulsive or negligible.

However, is this conclusion correct? It is evident in Figure 1 that the aromatic rings of the central phenylalanines are in contact with the aromatic and aliphatic side chains of the neighboring amino acids. These contacts should be stabilized by London dispersion energy. The theoretical description of London energy is difficult, and only highly accurate correlated ab initio calculations with extended basis sets are adequate in this case.[13]

Therefore, the calculations should be performed at the highest possible level, excluding the traditional problems of ab initio quantum chemical calculations, that is, the incompleteness of the AO basis set and insufficient amount of correlation energy covered.

Table 1. Pair of Interaction Energies (in kcal/mol) of the Selected Residues Clustered around F30[a]

	RIMP2			ΔCCSD(T)[b]	CCSD(T)/CBS
residue	aug-cc-pVDZ	aug-cc-pVTZ	CBS	6-31G*(0.25)	
F49	−3.1	−3.3	−3.3	−/0.6	
K46	−3.1	−3.3	−3.4	0.3/0.2	−3.10
L33	−4.9	−5.3	−5.5	0.5/0.2	−5.00
Y13	−4.2	−4.4	−4.5	0.6/0.4	−3.90
Y4	−6.5	−6.8	−7.0	−/1.7	
sum	−21.8	−23.2	−23.7		

[a] Compare Figure 1. [b] First number is the correction for whole modeled residue; second number is the correction for side chain only (side chain modeled from C_β atom).

Table 2. Pair of Interaction Energies (in kcal/mol) of the Selected Residues Clustered around F49[a]

	RIMP2		
residue	aug-cc-pVDZ	aug-cc-pVTZ	CBS
C39	−1.7	−2.0	−2.1
C6	−4.4	−4.8	−5.0
F30	−3.1	−3.3	−3.3
K46	−4.0	−4.6	−4.8
V5	−5.6	−6.4	−6.7
Y37	−2.3	−2.4	−2.5
Y4	−2.7	−3.0	−3.1
sum	−23.8	−26.5	−27.5

[a] Compare Figure 1.

Inspecting the RIMP2/CBS interaction energies (the lower part of Figure 2 and Tables 1 and 2), we found a very surprising picture. All 12 pairs of interaction energies were negative (i.e., stabilizing), and the stabilization energies were relatively high (for six pairs, even higher than 4.5 kcal/mol at the CBS limit). Especially important were the F30···Y4 and F49···V5 pairwise interactions with stabilization energies of about 7 kcal/mol. The first pair is stabilized by the interaction of the two aromatic rings, and the structure corresponds to a parallel-displaced structure of a benzene dimer. F49···V5 interaction is of a different nature. Due to the fragmentation procedure, the pair contains the CH···π contact instead of the π···π contact present in the real system (interaction of π electrons of phenylalanine and a peptide bond). As we will see later, the very large stabilization energy of the present system (F49···V5) only approaches the stabilization energy of an alternative model containing a phenylalanine ring and a peptide bond. The F30···Y4 and F49···V5 pairs clearly illustrate the stabilization role of the amino acid aromatic ring and show that strong stabilization (comparable or even higher than H-bonding) can originate from dispersion attraction without the presence of any classical H-bond.

Stabilization in the remaining 10 pairs is significant, as well, and supports the idea that the structural motifs of aromatic rings and aliphatic chains also contribute considerably to the stability of the system. It also supports a conclusion of Loladze et al.[26] that packing of nonpolar groups in the protein interior is favorable and is largely defined by a favorable enthalpy of van der Waals interactions. The CCSD(T) correction term, the calculation of which is extremely time demanding, is, in the

(22) Jurečka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578.
(23) (a) Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 112802. (b) Šponer, J.; Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142. (c) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608.
(24) Čížek, J. *J. Chem. Phys.* **1966**, *45*, 4256.
(25) Čížek, J. *Adv. Chem. Phys.* **1969**, *14*, 35−89.
(26) Loladze, V. V.; Ermolenko, D. N.; Makhatadze, G. I. *J. Mol. Biol.* **2002**, *320*, 343.

cases of F30···K46, F30···L33, and F30···Y13, rather small (contrary to the case of stacked DNA base pairs), and the CBS limit of the MP2 stabilization energy can thus serve as a measure of true stabilization. Using this conclusion, we can state that both F30 and F49 clusters are described sufficiently on the RIMP2/CBS level.

The total stabilization of F30 with the surrounding amino acids amounts to nearly 24 kcal/mol and, in the case of F49, nearly 28 kcal/mol (Tables 1 and 2). The average stabilization of phenylalanine with one partner is thus more than 4 kcal/mol. This is very high stabilization, equivalent to hydrogen bonding, and is definitely far from the expected negligible or even repulsive interaction energy. These energies are very comparable to the above-mentioned energy difference between the folded and unfolded states of protein (5−10 kcal/mol), further underlining the biological significance of this stabilization.

For residues in the F30 cluster, the interaction of side chains without the backbone atoms (e.g., starting with C_β atom of side chain) was also considered. The smaller size of these systems allowed us to evaluate the CCSD(T) correction (second term of eq 1) coming from the amino acid side chains themselves. As long as the inter-amino acid contact of the residues with F30 was mediated mainly by the side chains, it also helped us to get an approximate view of the effect of correlation energy covered by CCSD(T) in the case of the complexes exceeding our computational capabilities. The energies are also presented in Table 1.

There is just one case where the CCSD(T) correction for the side chain interaction exceeds 1 kcal/mol: the F30···Y4 interaction. It shows that the RIMP2/CBS interaction energy of −7 kcal/mol is overestimated by around 1.7 kcal/mol. Because the amino acids are in contact by their side chains only, it is plausible to consider the correction for the interaction of side chains as a correction for the interaction of whole amino acids.

In light of that, we did not evaluate the CCSD(T) correction term in the case of complex F49. The RIMP2/CBS values alone are, as written above, sufficient.

We have further used the RIMP2/CBS calculated stabilization energy for each pair of amino acids as a standard for the stabilization energies calculated by the empirical force fields frequently used in protein modeling (i.e., AMBER-parm94, parm99, ff02, ff03, CHARMm22, MMFF94, Engh-Huber, OPLS-AA, TAFF, and Rule). The situation with empirical potentials is more favorable than that for DFT (Figure 3). Three of the potentials (Charmm22, Engh-Huber, and TAFF) follow the trends of the reference method, but only qualitatively.

A more detailed examination into the structure of the F49 cluster revealed another way to treat interacting residues by keeping the existing peptide bonds (PB). In proximity to the central F49 aromatic side chain, two PBs (Y4···V5 and V5···C6) exist and can be alternatively considered as partners for the central phenylalanine (cf. Figure 4). One of these PBs (V5···C6) is parallel to the plane of the F49 side chain, while the other PB (Y4···V5) is perpendicularly oriented. The parallel arrangement is known to provide considerable stabilization energy.[27,28]

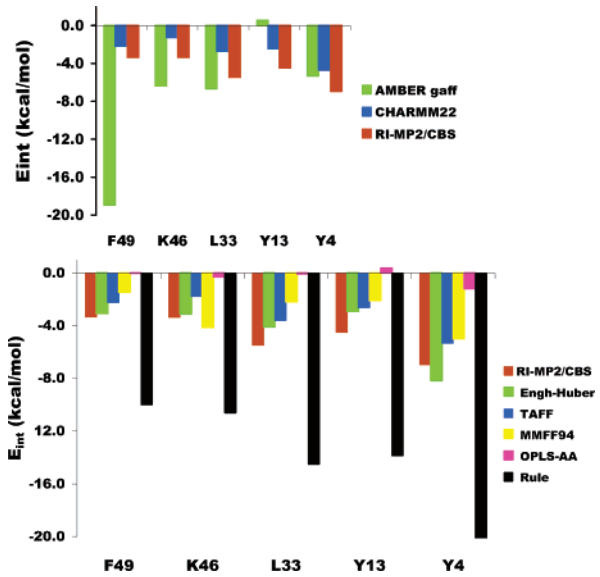


Figure 3. Comparison of the empirical interaction energies in cluster F30 evaluated by several force fields with the RIMP2/CBS results.

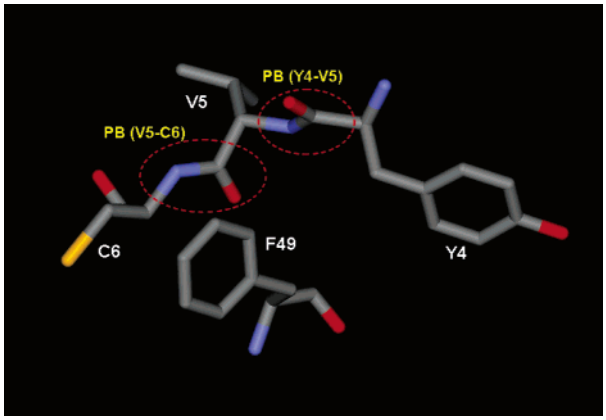


Figure 4. Two peptide bonds in proximity of the aromatic ring of the F49 residue. Dashed ellipses indicate both peptide bonds, PB(Y4−V5) and PB(V5−C6), for which a stabilization energy was evaluated.

Table 3. Pair of Interaction Energies (in kcal/mol) of the Two Peptide Bonds, PB(Y4−V5) and PB(V5−C6) with F30

	RIMP2			ΔCCSD(T)	CCSD(T)/CBS
complex	aug-cc-pVDZ	aug-cc-pVTZ	CBS	6-31G*(0.25)	
F49−PB (Y4−V5)	−3.0	−3.1	−3.2	0.4	−2.8
F49−PB (V5−C6)	−7.9	−8.5	−8.8	0.6	−8.2

We calculated the F49 interactions with these two PBs (each PB modeled as *N*-methylformamide), and the resulting CCSD(T)/CBS interaction energies (Table 3) amounted to −8.2 and −2.8 kcal/mol, respectively. The stabilization energy of the former motif (parallel orientation of the PB and an aromatic ring) is surprisingly high, even higher than that for hydrogen bonding, and sheds new light on the role of peptide bonds in the stabilization of protein structures. Let us only add here that the CCSD(T) correction term is repulsive (+0.6 kcal/mol), but the repulsion is only modest in comparison with that known in stacked structures of DNA base pairs.[29]

(27) Toth, G.; Watts, C. R.; Murphy, R. F.; Lovas, S. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 373.
(28) Duan, G.; Smith, W. H.; Weaver, D. F. *Int. J. Quantum Chem.* **2002**, *90*, 669.
(29) Šponer, J.; Hobza, P. *Collect. Czech. Chem. Commun.* **2003**, *68*, 2231.

To show that both fragmentation procedures are justified, we evaluated the size of the possible errors originating from various fragmentations. We calculated the interaction of F49 at the RIMP2/aug-cc-pVDZ level with the whole neutral tripeptide Y4−V5−C6 (i) and compared it to the results obtained for F49 interacting with three separated amino acids (Y4, V5, and C6) (ii) and for F49 interacting with two separated PBs (Y4···V5 and V5···C6) (iii). In the first case (i), we obtained a stabilization energy of 11.9 kcal/mol for the whole neutral Y4···V5···C6 complex. In the second case (ii), we obtained an energy of 12.7 kcal/mol (the sum for Y4, V5, and C6; see RIMP2/aug-cc-pVDZ results in Table 2). In the third case (iii), we obtained and energy of 10.8 kcal/mol (see Table 3). These results show that the difference between the three approaches is negligible, in a range of only 0.8 kcal/mol, and prove that both fragmentation methods are justified.

Conclusions

The present results show a complete failure of the DFT calculations, which are not even able to describe the attraction between central phenylalanines and neighboring amino acids. The results also fully support a known, but commonly ignored, fact that DFT methods cannot be recommended for simulating systems where London dispersion interactions play a major role.

The results presented here clearly demonstrate further the substantial attraction inside a hydrophobic core. This attraction, originating in London dispersion energy between aromatic rings or between an aromatic ring and an aliphatic chain, is comparable to classical H-bonding. Moreover, residues of aromatic nature can participate in several strong interactions at once, which may be crucial for the role of key residues in establishing small world networks inside a protein.[30]

Consequently, the current view on the nature of protein secondary and tertiary structure stabilization and, especially, the origin and nature of protein folding should thus be modified. Hydrophobic nature of a protein core implies that hydrophobic interactions can initiate the folding process. Present results indicate a decisive role of stabilization energy (enthalpy). Eventual consequences are that the energy (enthalpy) rather than hydrophobicity (entropy) can play a significant role during the early stage of protein folding.

Acknowledgment. We thank Pavel Jungwirth and Alex Wlodawer for critical reading and comments on the manuscript. This work is a part of the Research Project Z4055905 and was supported by a grant from the Grant Agency of the Academy of Sciences of the Czech Republic (Grant A400550510).

Supporting Information Available: Computational and additional data as well as figures and *xyz* coordinates of interacting pairs. This material is available free of charge via the Internet at http://pubs.acs.org.

JA044607H

(30) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. *Nature* **2001**, *409*, 641.