How the Stabilization of INK4 Tumor Suppressor 3D Structure Evaluated by Quantum Chemical and Molecular Mechanics Calculations Corresponds Well with Experimental Results: Interplay of Association Enthalpy, Entropy, and Solvation Effects

Michal Otyepka,[†] Petr Sklenovský,[†] Dominik Horinek,[‡] Tomáš Kubař,[‡] and Pavel Hobza*[,†,‡]

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

Received: November 28, 2005; In Final Form: January 12, 2006

The folding free energy of the INK4c tumor suppressor core, consisting of 10 helices, was determined as the sum of gas-phase interaction enthalpy, gas-phase interaction entropy, and dehydration and hydration free energy. The interaction energy and the hydration free energy were determined using the nonempirical density functional theory (DFT) method, augmented by a dispersion-energy correction term, the semiempirical density-functional tight-binding method covering the dispersion energy, and the density functional theory/conductor-like screening model (DFT/COSMO) procedure, whereas the interaction entropy was calculated with the empirical Cornell et al. force field. Alternatively, all contributions were evaluated consistently using empirical methods. All the values of the interaction energy of helix pairs are stabilizing, and the dominant stabilizing terms stem from the London dispersion energy and, in the case of charged systems, the electrostatic energy. The stabilization energy of the core, determined as the difference of the energy of the core and 10 separate helices, amounts to ∼450 kcal/mol. Systematically, the difference in the hydration free energy of a helix pair and its separate components is smaller in magnitude than the interaction energy, and it is negative for some pairs while positive for others. The average total free energy of a core formation amounts to −29.6 kcal/mol (yielded by scaled quantum-chemical methods) and +13.9 kcal/mol (resulting from empirical methods). These values are considerably smaller than their single components, which are dominated by the interaction energy. The computationally predicted interval encloses the experimental value of the folding free energy (−2.8 kcal/mol).

Introduction

The INK4 family of proteins consists of four known members: $p16^{INK4a}$, $p15^{INK4b}$, $p18^{INK4c}$, and $p19^{INK4d}$. The biological function of INK4 is to control the G1/S phase of a cell cycle by a specific inhibition of CDK4/Cyclin D and CDK6/Cyclin D complexes. Despite their being ubiquitous and mediating a diverse range of biological functions, very little is still known about their folding and assembly properties. The mutations or loss of INK4 inhibitors are associated with several types of cancer.

The INK4 proteins contain four or five ankyrin (ANK) repeat motifs; such a motif contains ∼33 amino acids and forms a β-hairpin−helix−turn−helix fold (see Figure 1).

The crystal structures and NMR experiments exhibit an L-shaped structural scaffold of the ANK repeat. The helices of the helix−turn−helix motif pack against each other in an antiparallel fashion to form a structure that resembles a horseshoe. These horseshoe motifs are connected by long loops folding back, forming β-turns (Figure 2).

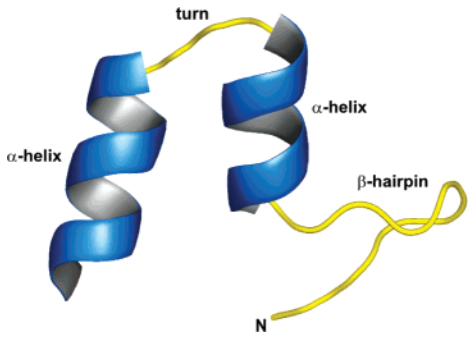


Figure 1. Depiction of the ankyrin repeat motif, which contains ∼33 amino acid residues; it consists of β-hairpin, helix, turn, and helix secondary structure motifs.

The ANK repeats in the INK4 proteins stack in a linear fashion, giving an elongated, nonglobular structure. $p16^{INK4a}$ and $p15^{INK4b}$ contain four ANK repeats, whereas $p18^{INK4c}$ and $p19^{INK4d}$ have five ANK repeats. Most of the interactions occur between the helical segments of neighboring repeats, and the lack of longer-range interactions could lead to the destabilization of the molecule.

The fundamentally different nature of these structures, compared to globular proteins, makes them a novel and important target of protein-folding studies. Furthermore, these protein types, which are formed by several structural motifs,

* Author to whom correspondence should be addressed. E-mail address: pavel.hobza@uochb.cas.cz.

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

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

10.1021/jp056890s CCC: $33.50
Published on Web 02/11/2006

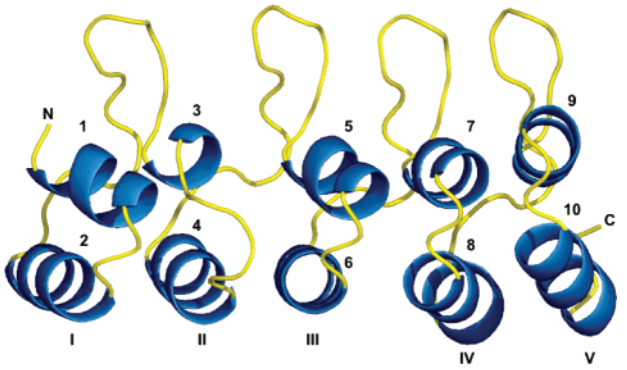


Figure 2. Crystal structure of the $p18^{INK4c}$ chain B (PDB code 1IHB) showing five (I−V) ankyrin motifs connected in a parallel fashion. Helix charges are equal to −1e, −1e, 0e, 1e, 0e, −2e, 1e, 1e, 0e, and −1e for helices 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10, respectively.

may serve for the testing of experimental and theoretical hypotheses, which have emerged from the analyses of globular protein folding.[1−8]

The proper protein fold, DNA replication, translation and transcription, receptor ligand recognition—simply every biomolecular recognition process is driven or determined by a gentle free-energy change. The free-energy changes associated with biologically important processes give values of 5−10 kcal/mol. Such mild changes allow reliable recognition and sufficiently strong binding but, conversely, do not rule out dissociation. It has been a long time since the first discussions about the importance of free-energy contributions arose. As it is written in textbooks, free energy consists of enthalpic and entropic terms, and the water environment must be taken into consideration by means of the solvation free energy. However, relatively small values of total free energy come from a partial compensation of large values of enthalpy, entropy, solvation, and desolvation contributions. Understanding the nature of the aforementioned processes requires a detailed knowledge of each single process, and it is impossible to rely on their compensation. Evidently, each term should be determined as accurately as possible. It may happen, however, that some terms are insensitive to the level of calculation, and, thus, the reliable values of these terms are already obtained at a rather low theoretical level. Recently, we have performed an accurate quantum chemical and statistical thermodynamic determination of binding free-energy components for several types of complexes. Specifically, we determined the stabilization energy of DNA base pairs,[9−12] amino acid pairs,[13] and also for the intercalation of ethidium to DNA.[14] In the latter case, we discovered[14] that a rather small interaction free energy (approximately −5 kcal/mol) is due to a partial compensation of the large contributions of interaction energy, entropy, and solvation free energy, which are clearly dominated by the first-mentioned one.

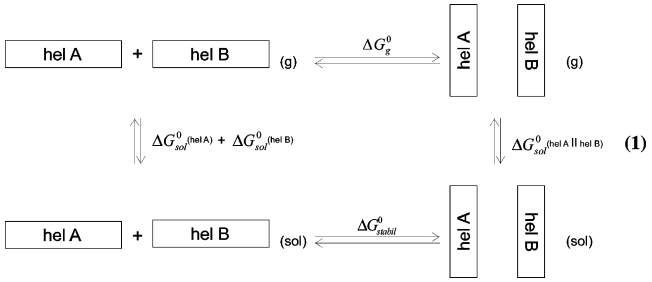
The aim of this study is to determine the change of free energy upon the formation of a hydrophobic core consisting of 10 helices; the entire process occurs in the water phase. The formation of a hydrophobic core is widely believed to originate in external entropic forces, whereas the interaction energy (enthalpy) contributions are repulsive or negligible. In our recent papers ,[13] we showed that the stabilization energy of the amino acid complex in the rubredoxin hydrophobic core is surprisingly large and we concluded that it must have a role in the folding process. In this paper, we extend the treatment and we intend to evaluate accurately not only the interaction energy but also the other contributions. In addition, a major step toward the biological reality is that, instead of single amino acids, we consider much larger helix motifs consisting of several amino acid residues as components of the hydrophobic core. The total binding free energy is determined as the sum of various contributions, the most important ones of which being (i) the interaction enthalpy of 10 helices forming the supersystem (the core) in the gas phase; (ii) the entropy term corresponding to the formation of the supersystem from isolated helices; (iii) the dehydration free energy of the isolated helices; and (iv) the hydration free energy of the supersystem. Despite the extensive size of the system, we evaluated contributions (i), (iii) and (iv) using reliable quantum-chemical techniques, and this allowed us to discuss the accuracy of the total binding free energy obtained.

Strategy of Calculation

The size of our system (more than 1000 atoms) dictates the choice of computational methods, and a consistent description of all processes is possible only at the empirical level. Our recent studies have shown that molecular mechanics methods are sometimes surprisingly successful, and the results obtained are very similar to the ab initio quantum chemical methods. Sometimes, however, the empirical potentials fail. The only way to determine the performance of a molecular mechanics force field is to compare the force-field results with the quantum-chemical results. Such a comparison can scarcely be done for a real system, and it is necessary to adopt a smaller model system that retains the characteristic features of the real system. In this study, we first determined the empirical interaction energy of helix pairs and compared it with the density functional theory (DFT) one. The ratio of these values was used in the second step to scale the molecular mechanics interaction energy of the entire complex, determined as the difference between the supersystem energy and the sum of the energy of 10 isolated helices. Needless to say, such a calculation at a nonempirical level is impractical. Subsequently, the same procedure was performed for the hydration free energy—for the helix pairs, we calculated the difference between the hydration free energy of the pair and its components, using both the empirical method and the conductor-like screening model (COSMO) model. The ratio of the quantum-chemical hydration free energy and the molecular mechanics hydration free energy was then used as the scaling factor in the calculation of the hydration free energy of the supersystem, which is too large to be examined using the COSMO model. The entropy contribution was not scaled at all, because entropy is known to be quite independent of the theoretical level used, and the molecular mechanics values are considered to be sufficiently accurate.

Computational Model

Regarding the structure of $p18^{INK4c}$ formed by five ANK repeats (Figure 2), a model consisting of 10 helical peptide units was chosen for the estimation of stabilization free energy. Such a model enables the use of quantum chemical methods for the calculations of pairwise helix interactions and a comparison with empirical results. The loop−loop interactions are considered to contribute negligibly to the stabilization free energy, as discussed below. Equation 1 defines the thermodynamic cycle used for the calculation of the pairwise helix free energy of stabilization, both in the gas phase and in the water environment.



The free energy of helix-pair motif stabilization in the gas phase (ΔG_g^0) at the standard state (1 atm, 298.15 K), according to this thermodynamic cycle, is given by eq 2:

$$\Delta G_g^0 = \Delta H_g^0 - T\Delta S_g^0 \quad (2a)$$

$$\Delta H_g^0 = \Delta E_g + \Delta E_g^{ZPV} \quad (2b)$$

$$\Delta S_g^0 = f(q_{tr} q_{rot} q_{vib} q_{el.deg}) \quad (2c)$$

where ΔH_g^0, ΔE_g, ΔE_g^{ZPV}, and ΔS_g^0 represent the stabilization enthalpy, stabilization electronic energy, the zero-point vibration energy (ZPVE) change, and the interaction entropy (always negative), respectively. All terms are related to the formation of a helix pair. While the energy is determined by quantum chemical and molecular mechanics procedures, the entropy is calculated using the statistical thermodynamics approach with the ideal gas, rigid rotor and harmonic oscillator approximation.

The solvation free energy (ΔG_{sol}^0(X)), at the standard state (1 mol dm^{-3}, 298.15 K) is evaluated using a continuum solvent model. The free energy of helix pair formation in the solvent (ΔG_{stabil}^0) consists of two terms: the gas-phase free energy of helix-pair formation (ΔG_g^0) and the solvation free-energy difference between the pair and its separate components ($\Delta\Delta G_{sol}^0$; see eq 3).

$$\Delta G_{stabil}^0 = \Delta G_g^0 + \Delta\Delta G_{sol}^0 \quad (3a)$$

$$\Delta\Delta G_{sol}^0 = \Delta G_{sol}^0(\text{helA||helB}) - (\Delta G_{sol}^0(\text{hel}A) + \Delta G_{sol}^0(\text{helB})) \quad (3b)$$

We are aware that the interaction of extended (and, moreover, charged) molecular systems and its distance dependence may constitute a more complicated issue to address.[15] In our thermodynamic analysis, however, it is only the initial and final states of the system that are to be described—i.e., the perfectly separated helices and the completely folded protein. Therefore, our simple treatment is correct.

Methods

The coordinates of the 10-helical complex p18^{INK4C} (p18) were taken from the PDB database (PDB code: 1IHB, chain B). First, the protonation states of all histidine residues were checked manually, to create an optimal hydrogen-bond network. The inhibitor structure, which contained heavy atoms and crystal water molecules, was neutralized by two Na^+ counterions. All H atoms were added by the LEaP module of the AMBER package,[16] and the entire system was immersed in a rectangular water box with a minimum distance between the molecule and the box wall of 9 Å. The position of H atoms added was optimized in the periodic water box using the SANDER module of AMBER. Water molecules and counterions were removed for further calculations. The Cornell et al. force field (*parm99*)[17] was used for all molecular mechanics simulations.

Quantum Chemical Calculations. *Density Functional Theory (DFT).* Two levels of the density functional theory (DFT) were used for the calculation of the interaction energy of the selected helix pairs: the B-LYP[18,19] functional with the medium-sized SV(P) (3s2p1d/2s1p) basis set, and the TPSS[20] functional with the larger TZVP (5s3p1d/2s1p) basis set. To keep the computational expense of the DFT calculations tractable, with the largest helix pair having 376 atoms and 7640 basis functions at the TZVP level, we used the resolution of identity approximation.[21] A correct description of dispersion energy, which is absent within the DFT framework, was achieved by the inclusion of empirical dispersion terms (DFT-D).[22]

The London dispersion interaction originates in the electron correlation and, classically, it is described as the instantaneous-dipole-induced−dipole interaction. Let us emphasize that both the Hartree−Fock (HF) method and a majority of DFT calculations lack a correct description of dispersion interaction. On the other hand, both HF and DFT describe the classical energy components such as induction (polarization) and exchange-repulsion correctly. In this work, we are not going to evaluate the components of the interaction energy separately.

The interaction energy was obtained by subtracting the sum of the energy of the helices from the energy of the complex. It was not necessary to make any corrections for the basis-set superposition error (BSSE), the reason being that the DFT-D framework works with the BSSE-uncorrected interaction energy, and the dispersion-energy correction is fitted to high-level ab initio quantum chemical data corrected for the BSSE. Therefore, this dispersion-energy correction effectively removes the BSSE.

The self-consistent-charges density-functional tight-binding scheme with an empirical dispersion-energy term (SCC-DFTB-D)[23] was used for the study of additivity. The interaction energy of the complete hydrophobic core consisting of 10 helices was calculated both directly and as the sum of pairwise helix−helix interactions.

The Role of Environment. Hydration effects were introduced through the conductor-like screening model (COSMO)[24] within DFT. In this method, the solute molecule is placed into a cavity surrounded by a conducting medium, and the screening energy is scaled to account for a finite dielectric constant. The COSMO calculations were performed at the B-LYP/SV(P) level; to estimate the error, the higher-level TPSS/TZVP was used for two pairs.

All DFT calculations were made using the TURBOMOLE[25] (calculation of interaction energy) and GAUSSIAN[26] (optimization and calculation of hydration free energy) program packages.

Molecular Mechanics Methods. The interaction energy in the gas phase (ΔE_g) was calculated using the Cornell et al. force field (*parm99*)[17] without cutoff, and the dielectric constant ϵ_r was set to 1. The calculations of association entropy (ΔS_g^0) and the ZPVE change (ΔE_g^{ZPV}) were performed by the normal-mode analysis[27] (the NMODE module of AMBER). Prior to every calculation of vibrational frequency values, the respective system was energy-minimized in the gas phase, using the Newton−Raphson algorithm implemented in NMODE. The limit of gradient norm for the energy minimization (*drms*) was set to 0.00001 kcal mol^{-1} Å^{-1}. The *cutoff* was set to 99.0 Å in all entropy/ZPVE normal-mode calculations. All these calculations were performed under standard conditions (at a pressure of 1 atm and a temperature of 298.15 K).

The free energy of solvation ($\Delta\Delta G_{sol}^0$) was evaluated using the generalized Born (GB) solvation model,[28] as implemented

TABLE 1: Components of Binding Free Energy of 21 Helix Pairs Calculated Using the Quantum Chemical Methods SCC-DFTB-D, DFT-D (TPSS/TZVP) and the COSMO Model for Solvation Contributions, as Well as Using the Molecular Mechanics Method with the Cornell et al. Force Field and the GB Method for Solvation Contributions

	Binding Free Energy Component (kcal/mol)								
	Quantum Chemical Methods				Molecular Mechanics Methods				
helix pair	DFT-D ΔE_g	SCC-DFTB-D ΔE_g	COSMO $\Delta\Delta G^0_{sol}$	DFT-D + COSMO $\Delta E_{sol} = \Delta E_g + \Delta\Delta G^0_{sol}$	ΔE_g	$\Delta\Delta G^0_{sol}$	ΔE_{sol}	ΔE	$-T\Delta S^0_g$
hel 1−2	−19.9	−23.2	−6.6	−26.5	−19.1	3.3	−15.8	3.1	18.4
hel 1−3	−6.4	−13.2	−4.9	−11.3	−7.5	−3.1	−10.7	1.1	12.9
hel 1−4	−37.1	−41.6	28.1	−9.0	−30.9	26.4	−4.5	3.0	17.2
hel 2−3	−2.4	−2.7	−1.6	−4.0	−2.0	1.7	−0.3	2.1	16.8
hel 2−4	−64.8	−68.4	28.2	−36.6	−58.5	34.3	−24.2	3.8	20.5
hel 3−4	−11.3	−13.8	2.5	−8.8	−10.9	2.9	−8.0	1.3	13.8
hel 3−5	−0.5	−2.3	−2.2	−2.7	1.2	−3.6	−2.5	2.7	15.2
hel 3−6	−7.0	−7.2	1.8	−5.2	−2.8	2.1	−0.8	1.0	12.2
hel 4−5	−4.4	−4.7	4.1	−0.3	−4.4	4.2	−0.1	3.5	18.7
hel 4−6	−95.6	−107.8	74.9	−20.7	−84.5	66.2	−18.3	5.3	20.8
hel 5−6	−32.0	−42.3	18.8	−13.2	−14.5	1.7	−12.8	2.3	19.5
hel 5−7	−5.0	−10.2	−10.2	−15.3	2.5	−16.2	−13.7	2.3	15.7
hel 5−8	−21.9	−19.6	17.1	−4.8	−13.5	10.8	−2.7	2.6	17.3
hel 6−7	−60.3	−40.8	64.9	4.6	−28.9	27.6	−1.2	1.9	14.8
hel 6−8	−87.2	−97.8	72.1	−15.1	−61.9	46.4	−15.5	1.2	14.2
hel 7−8	−7.2	−9.8	−9.4	−16.6	−5.2	−8.2	−13.5	3.8	19.3
hel 7−9	−20.4	−25.9	5.5	−14.9	−20.2	8.3	−11.9	3.5	19.2
hel 7−10	−30.5	−23.7	30.4	−0.1	−17.5	13.9	−3.6	2.1	16.7
hel 8−9	−10.1	−3.9	9.7	−0.4	−1.9	1.7	−0.2	3.9	17.5
hel 8−10	−64.9	−66.0	42.5	−22.4	−58.2	41.4	−16.8	4.3	20.4
hel 9−10	−20.3	−19.6	6.5	−13.8	−16.7	4.0	−12.6	1.9	14.2
sum	−609.2	−644.7	372.1	−237.1	−455.4	265.8	−189.6	56.4	355.2

in SANDER. The dielectric constant of the protein exterior ϵ_r was set to 78.5, whereas the cutoff value was set to 999.0 Å and the value of ϵ_r inside the protein interior was set to 1 consistently with the GB model parametrization.[28] The nonelectrostatic contributions ($\Delta G^{nonelst}_{sol}$)—namely, cavitation, dispersion, and repulsion terms—were estimated linearly from the solvent-accessible surface area (SA).[29] All results of the empirical calculations were critically correlated with the nonempirical results, as discussed later.

Results and Discussion

Stabilization Energy of Helix Pairs. Table 1 shows the interaction energy of 21 helix pairs, determined with the DFT-D method.

First, all the interaction energy values are negative, i.e., all pairs investigated exhibit stabilization. The largest values of stabilization energy, which were found for the helix pairs 4/6 and 6/8, exceed 80 kcal/mol. For three helix pairs (8/10, 2/4, and 6/7), it is larger than 60 kcal/mol, whereas for six other pairs, it is larger or close to 20 kcal/mol. The largest values of stabilization energy were found for oppositely charged helices in the pairs 4/6 (+1e/−2e), 6/8 (−2e/+1e), 8/10 (+1e/−1e), 2/4 (−1e/+1e), and 6/7 (−2e/+1e), and it is evidently the electrostatic term that is responsible for such a large stabilization. However, even in the case of uncharged species, the stabilization energy is large (e.g., the helix pairs 5/6, 9/10 and 7/9). In the three examples mentioned, the entire stabilization stems exclusively from the dispersion energy; this contribution to the interaction energy gives values of −21, −17, and −23 kcal/mol for the helix pairs 5/6, 9/10, and 7/9, respectively. In the charged helix pairs, the contribution of the dispersion energy is also significant. We can thus conclude that, besides the electrostatic term, the dispersion energy contributes dominantly to the overall stabilization within the hydrophobic core. Another conclusion is that a method not covering the dispersion energy cannot be used for the study of a hydrophobic core, even if its components are charged.

Let us further investigate the applicability of computationally much more feasible procedures, such as the SCC-DFTB-D method and the empirical Cornell et al. force field. Comparing the second and third columns in Table 1, we have discovered a very good agreement between these values, with the SCC-DFTB-D stabilization energy values being ∼6% larger; the 25% smaller stabilization energy yielded by the Cornell et al. potential (the sixth column in Table 1) can still be considered acceptable. The total stabilization within the core is enormous—ca. 600 kcal/mol yielded by quantum-chemical methods and ca. 450 kcal/mol by the empirical potential—and this finding supports our previous conclusions about the importance of the stabilization energy in the hydrophobic core of rubredoxin.[13] The assumption that the interaction energy of hydrophobic core components is small or even repulsive is evidently wrong, and this textbook opinion should be corrected.

For the final estimate of the free-energy change accompanying the formation of the core, it is necessary to scale the values of interaction energy yielded by the semiempirical SCC-DFTB-D method with respect to the results of an accurate DFT-D methodology. The reason is that the interaction energy of a complex formed by 10 helices (determined as the difference of the cluster energy and the sum of 10 values of single-helix energy) can be evaluated only at the latter level. As mentioned above, the SCC-DFTB-D stabilization energy is slightly overestimated. As a result, the scaling factor, which is determined as the ratio of sums of pair stabilization energy values yielded by DFT-D and SCC-DFTB-D (cf. Table 1), amounts to 0.95. Another way to determine the scaling factor is to compare all single-pair stabilization energy values on their own (rather than their sum) by a linear regression method. The correlation of semiempirical and DFT gas-phase data is described by eq 4,

$$\Delta E_g^{DFT-D} = (0.92 \pm 0.10)\Delta E_g^{SCC-DFTB-D} - (0.93 \pm 4.27) \quad (4)$$

where n is the number of data pairs, R^2 is the coefficient of determination, and s is the standard deviation. (For eq 4, $n = 21$, $R^2 = 0.951$, and $s = 6.52$.) The resulting value of the scaling factor (0.92) is slightly smaller than that determined using the sum of helix-pair stabilization energy values.

The pair stabilization energy values calculated by the Cornell et al. force field are underestimated, with respect to the DFT-D results; however, the overall agreement of both methods is surprisingly good (see the second and sixth columns in Table 1). The scaling factor determined from the sum of helix-pair stabilization energy values amounts to 1.34, whereas that from a linear regression of all pair energies (eq 5) is slightly smaller (1.15).

$$\Delta E_g^{\mathrm{DFT-D}} = (1.15 \pm 0.15)\Delta E_g^{\mathrm{MM}} - (4.14 \pm 4.94) \quad (5)$$

(For eq 5, $n = 21$, $R^2 = 0.927$, and $s = 7.96$.)

Interaction Entropy. All values of the interaction entropy are negative (cf. the last column in Table 1 for the values of $-T\Delta S^{\mathrm{int}}$, which are all positive). This is easily understandable, because the formation of the complex is associated with the loss of configuration freedom and, thus, the loss of entropy. Furthermore, all $-T\Delta S^{\mathrm{int}}$ values are similar and they all lie in a rather narrow interval between 12.2 kcal/mol and 20.8 kcal/mol. On the other hand, the situation with stabilization energy is entirely different, and the smallest and the largest terms differ by more than an order of magnitude. What is more important, however, is that the entropy terms are sensitive neither to the size of system nor to the level of calculation. Note that the magnitude of entropy accompanying the formation of various DNA-base pairs (both hydrogen-bonded and stacked) determined from ab initio characteristics (9.9−12.2 kcal/mol)[30] are only slightly smaller than the present values which relate to much more extended systems.

Hydration Free Energy. The change of COSMO hydration free energy accompanying the formation of a helix pair is shown in the fourth column of Table 1. The resulting values are mostly positive, indicating that a pair is hydrated worse than both helices. It is not surprising that the largest values of the hydration free energy were observed for charged helices. The negative values of the hydration free-energy change, indicating a better hydration of the pair when compared to the two helices, were found in six cases; in four cases, both helices are uncharged, and in two cases, one of them bears a positive (+1) charge. Although a positive change of hydration free energy compensates the negative value of interaction energy, in the case of negative changes of hydration free energy, both values add up. The fifth column of Table 1 shows the sum of the DFT-D interaction energy and the COSMO hydration free energy. The difference of the most negative and the most positive value gives 41 kcal/mol, which is considerably smaller than the difference found for the interaction energy (~95 kcal/mol). Evidently, the change of hydration free energy has a tendency to compensate the large difference of bare stabilization energy.

The COSMO hydration free-energy values are similar to the empirical values (see the seventh column in Table 1), and the sum of all hydration free-energy values is larger in the case of the COSMO method. The scaling factor, which is determined as the ratio of both values, is 1.40. Similar to the previous cases, we determined the scaling factor also by a linear regression of the hydration free-energy values for all pairs. The correlation of empirical and quantum-chemical data is described by eq 6, and the resulting value of 1.18 is again slightly smaller than that determined from the sum of total hydration free-energy values.

TABLE 2: Evaluation of the Total Stabilization Free Energy of the Complex Formed by 10 Helices Calculated by the Molecular Mechanics

		Total Stabilization Free Energy (kcal/mol)	
	term calculated	molecular mechanics	corrected value[a]
$\Delta H_g^0 = \Delta E_g + \Delta E_g^{\mathrm{ZPV}}$	ΔE_g	−374.6	−476.6[b]
$\Delta H_g^0 = \Delta E_g + \Delta E_g^{\mathrm{ZPV}}$	$\Delta E_g^{\mathrm{ZPV}}$	27.4	27.4
$\Delta H_g^0 = \Delta E_g + \Delta E_g^{\mathrm{ZPV}}$	$-T\Delta S_g^0$	177.1	177.1
$\Delta G_g^0 = \Delta H_g^0 - T\Delta S_g^0$	ΔG_g^0	−170.1	−272.1
$\Delta G_g^0 = \Delta H_g^0 - T\Delta S_g^0$	$\Delta\Delta G_{sol}^0$	184.0	237.4[c]
$\Delta G_{stabil}^0 = \Delta G_g^0 + \Delta\Delta G_{sol}^0$	ΔG_{stabil}^0	+13.9	−34.7
	$-T\Delta S_{sol}^0$	−103.2[d]	−103.2[d]

[a] This column contains data scaled to the quantum chemical results. [b] SCC-DFBT-D linearly scaled to DFT-D. [c] GB/SA linearly scaled to DFT-D/COSMO. [d] Estimated hydrophobic effect, as described by Raschke et al.[31]

$$\Delta\Delta G_{sol}^{\mathrm{COSMO}} = (1.18 \pm 0.25)\Delta\Delta G_{sol}^{\mathrm{GB}} + (2.74 \pm 5.80) \quad (6)$$

(For eq 6, $n = 21$, $R^2 = 0.842$, and $s = 10.72$.)

Formation Free Energy of the p18^{INK4c} Structure. The total stabilization energy cannot be determined as the sum of all pair interactions, because almost all helices are charged, and, thus, many-body contributions cannot be neglected. The same is true for all remaining contributions to the total free energy change of core formation. As mentioned previously, the size of the core (1429 atoms) prevents us from using accurate quantum chemical procedures. Instead, we use lower-quality methods and we scale the resulting values based on the comparison of pair interactions evaluated with both quantum chemical and empirical procedures (cf. Table 2).

We expect the stabilization energy of 10 helices to dominate all terms. The SCC-DFTB-D stabilization energy of 10 helices is large and exceeds 500 kcal/mol. However, this energy is overestimated and should be reduced by a suitable scaling factor. We found two values of this factor (0.95 and 0.92), and, in the following discussion, we use their arithmetic mean (i.e., 0.93). The resulting scaled stabilization energy is 476.6 kcal/mol. Adding the ZPVE energy change, we get the stabilization enthalpy of 449.2 kcal/mol. The interaction entropy is smaller, and the resulting free energy of core formation in the gas phase amounts to −272.1 kcal/mol. In the second step, the hydration free energy is added. The GB/SA results used for the entire core should be scaled by a factor of 1.29, which has been determined as the arithmetic mean of the values obtained previously (1.40 and 1.18). Finally, summing up the gas-phase free energies of core formation and the hydration free-energy change, we obtain the free energy of core formation in a water environment. This value (cf. Table 2) is much smaller than the values of enthalpy, entropy, and hydration free energy and amounts to −34.7 kcal/mol. The negative sign means that the formation of the core is favorable and the core is stabilized by 34.7 kcal/mol.

The aforementioned free-energy value is based on the SCC-DFTB-D stabilization energy. Another possibility is to use the empirical Cornell et al. stabilization energy, together with the corresponding scaling factors. The remaining terms do not change; therefore, the resulting value of the formation free energy in a water environment amounts to −24.5 kcal/mol.

We have obtained two values of free energy, differing by ~10 kcal/mol, and it is difficult to prefer one of them over the other. As the final result of our calculations, we therefore take the arithmetic mean of both values, which amounts to −29.6 kcal/mol. This value is based on semiempirical and empirical results scaled to quantum chemical data. As shown previously, our scaling was not systematic and some terms were not scaled at all. Although we collected evidence that these terms are not sensitive to the quality of calculations, this approach might be questioned. Another computational strategy, which avoids this problem, is to use empirical methods systematically and without any scaling. Table 2 shows that the resulting free energy of core formation amounts to +13.9 kcal/mol and it is repulsive (i.e., destabilizing) now. The difference between the final values of both procedures is ~40 kcal/mol, which seems large at first glance.

The scaled semiempirical/empirical treatment and the unscaled empirical treatment delimit the computationally obtained value of free energy to the interval from −29.6 kcal/mol to +13.9 kcal/mol. How does this estimate agree with the experimental value? Unfortunately, the comparison is not straightforward, because our results concern the formation of a core from 10 helices, supposing that every helix maintains its structure upon its release from the core. However, the experimentally determined free energy of folding concerns a slightly different situation, because the secondary structure of the unfolded state is partially disrupted and ill-defined. Consequently, the free energy of folding comes from two contributions: the formation of complete secondary structure motifs and the alignment of the secondary structure to the native fold. The former term corresponds to the model adopted in our computational treatment. The latter term (the secondary structure completion) corresponds predominantly to the completion of a hydrogen-bond network, and the free energy of this process would probably vanish. Therefore, the free energy of folding can be compared with the free energy of stabilization, taking into consideration that the contribution of the loop−loop interaction is negligible. By means of experimental techniques, the free energy of $p18^{INK4c}$ denaturation was determined as 2.70 kcal/mol (from ref 5) or 2.98 kcal/mol (from ref 7). This value should equal the negative of folding free energy, so $\Delta G^0_{fold} = -2.8$ kcal/mol.

When explaining the discrepancy between the results obtained with different methodologies (and the experimental value), several points should be discussed. It seems that the following statements generally hold for the calculations of association or folding free energy of extended molecular systems. The folding free energy is obtained as a small difference of several large values. In such cases, the errors add up and a large uncertainty is introduced, although state-of-the-art methods are used to calculate the components of free energy. However, the difference between the folding free energy obtained computationally and the experimental value is at least an order of magnitude smaller than the free-energy components calculated (interaction energy, entropy, and solvation free energy). Therefore, we may judge the relative magnitude and importance of these individual contributions and the results can be considered satisfactory for this purpose. On the other hand, it seems that it is extremely difficult to calculate the absolute value of the association free energy for an extended system. Thermodynamic calculations are one of the most challenging tasks of computational chemistry, and taking into account their importance for the description of processes involving extended biomolecules, we can conclude that further development in this area is necessary.

Conclusions

I. The interaction energy of every helix pair is stabilizing, and the largest values were found for the charged systems. The electrostatic energy represents the dominant stabilizing term, with the London dispersion energy being the second most important. For uncharged complexes, the stabilization energy values are smaller but still significant and they consist completely of the London dispersion energy.

II. The DFT-D and SCC-DFTB-D methods yield similar values of pair stabilization energies; the latter one slightly overestimates them. In the case of the empirical Cornell et al. force field, on the other hand, the stabilization energy is slightly underestimated.

III. The total stabilization energy of 21 helix pairs, determined as the sum of pair stabilization energy values, exceeds 600 kcal/mol.

IV. The values of the hydration free-energy change are generally smaller than the values of the pair stabilization energy, and they are stabilizing for some pairs while being destabilizing for others. The quantum chemical and empirical procedures yield similar values, but the latter procedure slightly underestimates the magnitude of the hydration free-energy change.

V. The total free energy of a core formation was estimated to be −34.7 kcal/mol (SCC-DFTB-D interaction energy scaled to DFT-D) and −24.5 kcal/mol (empirical interaction energy scaled to DFT-D), and the average value is −30.6 kcal/mol. When the empirical approach is applied systematically, the resulting value is repulsive (+12.0 kcal/mol). All these values are considerably smaller than their individual components; these are dominated by the stabilization energy.

VI. The computationally predicted interval encloses the experimental value of folding free energy (−2.8 kcal/mol).

VII. The folding free energy is calculated as a (small) difference of several large energy contributions. In this way, a relatively large error is introduced even if top-quality methods are used. However, the relative significance of these contributions to the overall free energy can be assessed.

Acknowledgment. We thank two anonymous reviewers whose comments and suggestions helped us to improve the comprehensibility of our article. This work was supported by Grant Nos. LC512 (M.O., P.H.) and MSM6198959216 (M.O.) from the Ministry of Education, Youth and Sports, Czech Republic.

References and Notes

(1) Sedgwick, S. G.; Smerdon, S. J. *Trends Biochem. Sci.* **1999**, *24*, 311.
(2) Tang, K. S.; Guralnick, B. J.; Wang, W. K.; Fersht, A. R.; Itzhaki, L. S. *J. Mol. Biol.* **1999**, *285*, 1869.
(3) Tang, K. S.; Fersht, A. R.; Itzhaki, L. S. *Structure* **2003**, *11*, 67.
(4) Venkataramani, R.; Swaminathan, K.; Marmorstein, R. *Nat. Struct. Biol.* **1998**, *5*, 74.
(5) Venkataramani, R. N.; MacLachlan, T. K.; Chai, X. M.; El-Deiry, W. S.; Marmorstein, R. *J. Biol. Chem.* **2002**, *277*, 48827.
(6) Zhang, B.; Peng, Z. Y. *J. Mol. Biol.* **2000**, *299*, 1121.
(7) Yuan, C. H.; Li, J. A.; Selby, T. E.; Byeon, I. J. L.; Tsai, M. D. *J. Mol. Biol.* **1999**, *294*, 201.
(8) Mosavi, L. K.; Cammett, T. J.; Desrosiers, D. C.; Peng, Z. Y. *Protein Sci.* **2004**, *13*, 1435.
(9) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247.
(10) Šponer, J.; Hobza, P. *Collect. Czech. Chem. Commun.* **2003**, *68*, 2231.
(11) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608.
(12) Šponer, J.; Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142.
(13) Vondrášek, J.; Bendová, L.; Klusák, V.; Hobza, P. *J. Am. Chem. Soc.* **2005**, *127*, 2615.

(14) Kubař, T.; Hanus, M.; Ryjáček, F.; Hobza, P. *Chem.—Eur. J.* **2006**, *12*, 280.
(15) Israelachvili, J.; Wennerström, H. *Nature* **1996**, *379*, 219.
(16) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.
(17) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
(18) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(19) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(20) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
(21) Eichkorn, K.; Treutler, O.; Ohm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283.
(22) Jurečka, P.; Černý, J.; Hobza, P. Unpublished results, 2005.
(23) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149.
(24) Klamt, A.; Schuurmann, G. *J. Chem. Soc. Perkins Trans. 2* **1993**, 799.
(25) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
(26) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; 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.
(27) Kottalam, J.; Case, D. A. *Biopolymers* **1990**, *29*, 1409.
(28) Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275.
(29) Weiser, J.; Shenkin, P. S.; Still, W. C. *J. Comput. Chem.* **1999**, *20*, 217.
(30) Hobza, P.; Šponer, J. *Chem. Phys. Lett.* **1996**, *261*, 379.
(31) Raschke, T. M.; Tsai, J.; Levitt, M. *Proc. Natl. Acad. Sci., U.S.A.* **2001**, *98*, 5965.