Loss of Dispersion Energy Changes the Stability and Folding/Unfolding Equilibrium of the Trp-Cage Protein

Jiří Černý,[†] Jiří Vondrášek,[†,‡] and Pavel Hobza*[,‡,§]

Institute of Biotechnology, Academy of Sciences of the Czech Republic, 142 00 Prague 4, Czech Republic, Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Systems, 166 10 Prague 6, Czech Republic, and Department of Physical Chemistry, Palacký University, Olomouc, 771 46 Olomouc, Czech Republic

Received: January 16, 2009

The structure of proteins as well as their folding/unfolding equilibrium are commonly attributed to H-bonding and hydrophobic interactions. We have used the molecular dynamic simulations in an explicit water environment based on the standard empirical potential as well as more accurately (and thus also more reliably) on the QM/MM potential. The simulations where the dispersion term was suppressed have led to a substantial change of the tryptophan-cage protein structure (unfolded structure). This structure cannot fold without the dispersion energy term, whereas, if it is covered fully, the system finds its native structure relatively quickly. This implies that after such physical factors as temperature and pH, the dispersion energy is an important factor in protein structure determination as well as in the protein folding/unfolding equilibrium. The loss of dispersion also affected the α-helical structure. On the other hand, weakening the electrostatic interactions (and thus H-bonding) affected the α-helical structure only to a minor extent.

Introduction

The structure of bio-macromolecules such as DNA, RNA, or proteins is determined by the noncovalent interactions between the bio-macromolecular building blocks as well as by hydrophobic interactions.[1] Whereas the stabilization caused by the former interactions originates in energy (enthalpy) changes, the latter is due to entropy changes.

Among the various noncovalent interaction motifs, hydrogen bonding and stacking interactions play the most important role, but their origin is fundamentally different. Hydrogen bonding is caused by dipole−dipole interactions and charge transfer (hyperconjugation), whereas stacking is almost exclusively caused by London dispersion energy. There is no doubt about the role of specific H-bonding in both DNA and protein structures. On the other hand, the role of stacking interactions, which are nonspecific, is less clear and should be explained in more detail. Highly accurate quantum mechanical (QM) calculations of interaction energy performed for complexes of biomolecular building blocks convincingly demonstrated the important role of stacking, or more generally dispersion, interactions.[2,3] Evaluation of the role of dispersion energy in the dynamical structure of DNA is, however, more complicated since it is necessary to go beyond the 0 K, beyond an isolated state, and also beyond the static description. Recently, we have performed molecular dynamics (MD) simulations of the B-DNA dodecamer in a water environment at 300 K and demonstrated that a loss of dispersion energy induces a structural transition from the three-dimensional double-helical structure to the quasi two-dimensional ladder-like structure.[4] This structure is clearly unable to fulfill the basic function of DNA, i.e., to store and transfer genetic information.

The same can be performed in silico experiment for a protein. How important are particular interactions or forces for protein stability and the folding/unfolding equilibrium? Such a question is fully justified due to the common objection of the biochemical community that dispersion interaction between amino acids of the hydrophobic core or between protein and nonpolar ligands is roughly of the same magnitude as interactions of these molecules with water and thus cannot be a driving force of the processes. It has been shown previously[3,5] that the role of dispersion in the stabilization of the hydrophobic core or protein−ligand complexes is significant. But this description (in the case of a theoretical model) lacks the important features of the process—its dynamics and the presence of the environment.

The role of dispersion and H-bonding (electrostatic) interactions on protein stability and the folding/unfolding equilibrium in the water environment is investigated in the present paper by two fundamentally different methods. First, MD simulations were performed with standard energy functions and, subsequently, with empirical energy functions where the dispersion and electrostatic energies were modified. Second, MD simulations were performed with a more reliable QM/MM potential where the QM part was described by the approximate density functional theory (DFT) procedure with or without empirical dispersion energy, and the MM part was described by a standard empirical potential.

We selected the tryptophan-cage (Trp-cage) protein (Figure 1) for the purpose of this study. The Trp-cage is an artificial, small protein that has been designed by the utilization of the known protein scaffold, whose role for theoretical as well as experimental purposes is enormous. This system is frequently used for investigating stability and folding processes. The stability of the Trp-cage (as well as of any other protein) can be affected by various factors, with the temperature, solvent, and pH playing the most important roles. As the temperature

* To whom correspondence should be addressed. Tel.: (+420) 220 410311. E-mail: pavel.hobza@uochb.cas.cz.
[†] Institute of Biotechnology, Academy of Sciences of the Czech Republic.
[‡] Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Systems.
[§] Palacký University.

10.1021/jp9004746 CCC: $40.75
Published on Web 03/26/2009

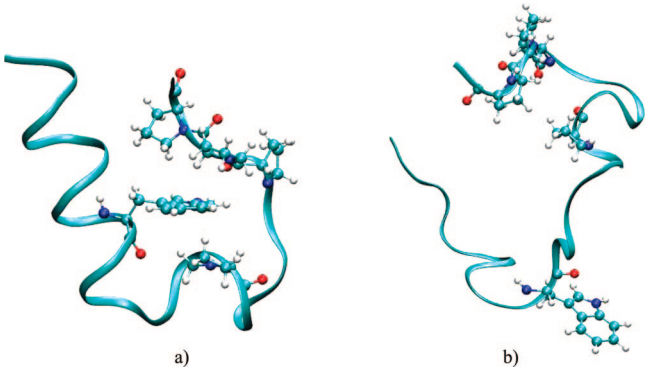

Figure 1. Initial structure (a) of the protein and the structure obtained by performing the 100 ns MD simulation with the ε scaled by 0.01 (b).

increased, the protein unfolded and the theoretical temperature (from the MD/Amber simulations) was found to be higher than the experimental temperature.[6] The same process can occur by changing the solvent, and here hydrophobic interactions play a key role.[7]

In the present study, we have investigated another factor affecting the structure and the folding/unfolding equilibrium, namely, the presence and/or accurate description of the dispersion and electrostatic energies. It was well-justified, because we had studied the Trp-cage protein recently using advanced methods of quantum chemistry and had revealed[8] the substantial stabilization between tryptophan and prolines as well as the key role played by dispersion energy. It should be mentioned here that an accurate description of the dispersion energy is difficult, and several advanced QM procedures, such as Hartree−Fock or standard DFT, do not cover it at all, whereas others, such as MP2, generally overestimate it dramatically. The lack of a dispersion energy term is also one of the reasons preventing the wider use of semiempirical QM methods in the realm of biodisciplines.

Methods

The molecular dynamics simulations were performed with an AMBER parm99 empirical force field,[9] and the following modifications were introduced in the nonbonded part of the empirical force field, which describes the potential energy of the system (see eq 1) and is divided into electrostatic and Lennard-Jones terms.

$$V(r) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \qquad (1)$$

The modifications of the dispersion energy were introduced by scaling the ε parameter, and the following values of the scaling factor were considered: 0.01, 0.5, 1.0, and 2.0. We are aware that scaling of the ε value affects both the dispersion and repulsion parts of the potential. We found,[4] however, that smaller repulsion does not introduce any problem during the simulations. The simulations with the modified electrostatic term were performed with atomic point charges scaled by 0.1 and also with full (unscaled) point charges. For all the MD simulations, the modifications were applied on all of the protein atoms, maintaining the original parameters for the TIP3P water molecules. The protein was placed into a cubic water box with the shortest distance of 10 Å from the molecule to the edges of the box. All the simulations were performed for 100 ns at a constant temperature of 300 K and under constant pressure conditions of 1 bar, starting from the parameters obtained after a 1 ns equilibration run. The Gromacs 3.3 MD package[10] was used for all the simulations.

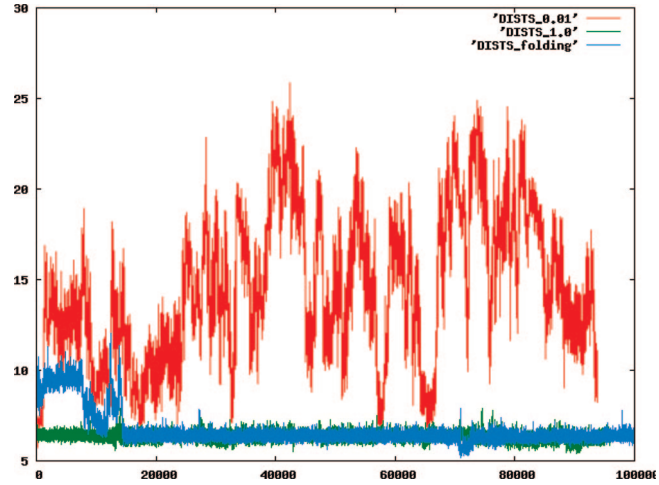

Figure 2. Description of the time (x-axis, ps) and mean distance (y-axis, Å) of the centers of mass of the tryptophan and proline residues. The green line corresponds to the simulations with the standard (unmodified) potential, with the mean distance being about 6 Å. A reduction of the dispersion energy almost to zero with the ε scaled by 0.01 (the red curve) leads to a dramatic change of the protein structure characterized by considerably larger mean distances between Trp and Pro (see also Figure 1b). When, however, the dispersion energy is fully recovered, the unfolded structure (found after 100 ns simulation without dispersion energy) is reversibly folded (the blue line).

The more reliable QM/MM MD simulations were performed employing the self-consistent charge density-functional tight-binding (SCC-DFTB) method empirically augmented with dispersion energy[11] as the QM part. The MM part was described by the Amber force field parm99.[9] The SCC-DFTB method provides reliable estimates of the noncovalent interactions in molecular complexes as well as in complex molecular systems (e.g., proteins) only if the dispersion energy is negligible. In cases where dispersion is important (e.g., with the stacking interactions of nucleic acid bases or aromatic amino acids), the SCC-DFTB-D procedure provides surprisingly accurate energy characteristics.[11] The application of the QM/MM method on all the residues of the studied protein as the QM part is impractical, and thus a smaller model should be adopted. We selected only the atoms of the hydrophobic core residues (four Pro and one Trp residue), which were described quantum mechanically by the SCC-DFTB-D procedure, whereas their backbone atoms and all the water molecules, as well as the rest of the protein, were treated by an unmodified empirical force field. The simulation was performed for 3 ns at a constant temperature of 300 K and under constant pressure conditions of 1 bar. For the QM/MM calculations, the AMBER10 package was employed.[12]

Results and Discussion

MD/MM Simulations. By varying the parameters which influence the dispersion energy term, the structure of the Trp-cage protein was modified; the situation is described in Figure 2. It shows the mean distance of the centers of mass for the tryptophan (Trp) and proline (Pro) residues in the structure of the Trp cage. The green line corresponds to the simulations with a standard (unmodified) potential and a mean distance of about 6 Å, which remains practically unchanged during the whole simulations. A nearly identical curve (not shown) was obtained

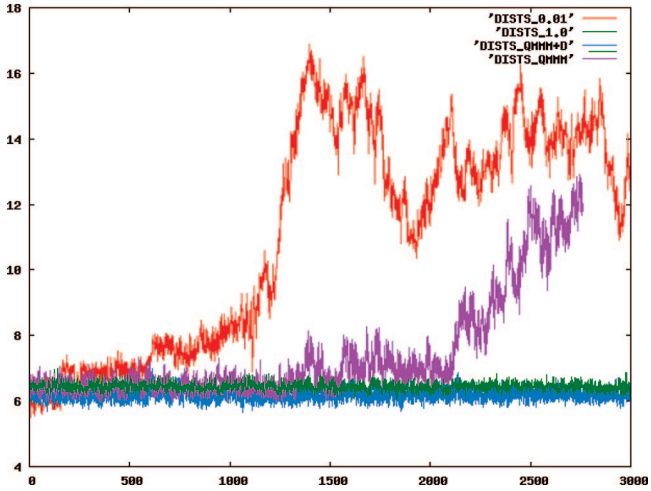

Figure 3. Comparison of the mean Pro−Trp distance for the unmodified MD simulation (the green line) and for the MD simulation with the ε scaled by 0.01 (the red curve) to the results of the QM/MM MD simulations. The blue line represents the simulation where the SCC-DFTB-D method was used as the QM part, whereas the violet curve corresponds to the SCC-DFTB method (without the empirical dispersion correction term). The *x*- and *y*-axes describe the time (ps) and mean distance (Å) of the centers of mass of the tryptophan and proline residues.

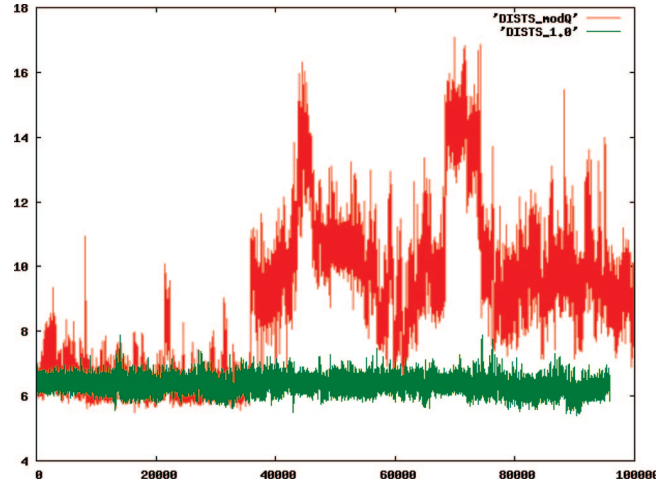

Figure 4. Mean Pro−Trp distance for the unmodified MD simulation (the green line) and for the simulation with the modified electrostatic term, where the atomic point charges of the protein atoms were scaled by 0.1 (the red curve). The *x*- and *y*-axes describe the time (ps) and mean distance (Å) of the centers of mass of the tryptophan and proline residues.

if the strength of dispersion was increased 2-fold. A reduction of the dispersion energy almost to zero (the dispersion energy was reduced to 1%; the red curve) leads to a dramatic change of the Trp-cage structure, characterized by considerably larger mean distances between the Trp and Pro. Evidently, the binding between Trp and Pro amino acids vanished, the Trp residue moved out from the protein interior and is fully exposed to the solvent, and the cage opens. Figure 1b shows the final structure. Besides the changed position of the Trp, it also demonstrates that a reduction of the dispersion energy causes an unwinding of the α helix. When the dispersion energy is decreased to 50% of its original value, the Trp-cage structure oscillates between the native (folded) and nonnative structures (not shown) during the 100 ns simulation.

We have shown that removing the dispersion energy leads to complete disruption of the structure: hence, another question arises, namely, what is going to happen when the dispersion energy term is fully recovered. The blue line in Figure 2 shows the trajectory for this case. We started from the nonnative structure obtained by the above-described procedure, and a 100 ns simulation was then performed. After a short time (≈ 10 ns), the structure has evidently fully recovered to the native state, and the blue line fully coincides with the green one.

Allow us to summarize the highlights of the process. Without dispersion energy, the protein unfolds and remains unfolded. When the dispersion energy is recovered, the protein reaches its native state in a relatively short time. This finding is important, because it tells us that the folding/unfolding equilibrium is affected not only by hydrophobic interactions[13,14] but substantially also by dispersion energy.

QM/MM MD Simulations. The QM/MM MD simulations fully confirmed the conclusions presented above. The blue line and violet curve in Figure 3 correspond to the simulations where the dispersion energy is fully covered (DFTB augmented by dispersion term) or is completely lacking. The results obtained by force field methods (the green line and red curve) are shown for comparison. Evidently, the unfolding in the case of QM/MM simulations occurred later than in the case of MM simulations (2 and 1 ns, respectively), which can be explained by a different model of dispersion energy scaling (see Methods). We did not perform more simulations, because it is evident that both the presence and absence of the dispersion energy term unambiguously affect the structure as well as the folding/unfolding equilibrium.

The fact that the QM/MM model fully agreed with the simpler MM one is significant. The reduction of the dispersion energy in the latter model is accompanied by a reduction of the repulsion energy, and the overall interaction energy is thus described only by the charge−charge electrostatic term (see eq 1). In the QM/MM model, this is not true, and it is only the dispersion energy that is covered or deleted. When the dispersion energy is not covered, the remaining SCC-DFTB interaction energy fully covers all the other interaction energy terms such as the electrostatic (however, not only the charge−charge (term) as in the MM approach, but all the terms in the expansion), exchange−repulsion, induction, exchange−induction, and charge transfer. The agreement between the QM/MM and MM simulations thus indicates that the deep changes in the structure and the folding/unfolding equilibrium of the protein can be ascribed solely to the missing dispersion term and not to the physically oversimplified interaction energy in the MM approach when the dispersion and consequently repulsion terms were deleted (see above).

The reduction of the atomic charges directly correlates with a decrease of the electrostatic energy, which supposedly affects mainly the H-bonding. The green line in Figure 4 indicates the mean distance between the Trp and Pro systems obtained from the simulations with the standard (unmodified) empirical potential. The distance of around 6 Å remains stable in the 100 ns simulations. When the atomic charges are reduced to 10%, i.e., the electrostatic interaction is reduced to 1% (the red curve), the mean distance between the Trp and Pro systems oscillates a great deal between the states belonging to folded or completely random protein structures. Figure 5 shows snapshots of the Trp-cage structures obtained from the simulations after about 20, 40, and 60 ns, respectively. Figure 5 also shows that the α helical structure, which is supposed to be held by H-bonds, is surprisingly retained at a noticeable level of stability. It seems that an important part of the stabilization of the α helix originates

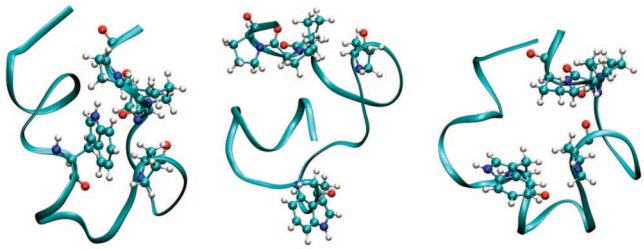

Figure 5. Snapshots of the protein structures obtained from the simulations with the modified electrostatic term (with the atomic point charges scaled by 0.1) after about 20 (a), 40 (b), and 60 ns (c), respectively.

also in dispersion energy, which is fully covered in these simulations (contrary to electrostatic energy, which is reduced to 1%).

Upon analogous reduction of the electrostatic energy (and thus H-bonding) in the case of DNA,[4] a different picture resulted. The DNA double-helical structure was denaturated to two separate strands. This was, however, not surprising since H-bonds between DNA bases are much stronger than these in the protein α helix. On the other hand, helicity within single DNA strands was retained to high degree, which corroborates the above-mentioned conclusion that the structure of the α helix is determined not only by H-bonding but also by dispersion forces.

The simulation in proteins is different and apparently more complex. The protein structure deteriorates completely upon the reduction of either electrostatic or dispersion energy, and not a trace of the former structural arrangement remains. This is evidently due to the fact that both H-bonding and stacking are stronger in DNA than in proteins.

Conclusion

We have shown that the folding/unfolding equilibrium of the Trp-cage protein can be affected not only by the physical factors such as the temperature or pH but also by the presence of dispersion energy. When this term was included neither in the empirical potential function nor the QM treatment, the folding/unfolding equilibrium of the protein was changed and the protein stayed in the nonnative state. When the dispersion energy was recovered, the native structure was reestablished in a rather short time. We should recall again that the accurate description of dispersion energy is rather difficult and many nonempirical and semiempirical QM methods fail in this respect; the use of these procedures for investigating protein dynamics is thus completely misleading and cannot be recommended. An unexpected finding was that the lack of dispersion energy affected also the Trp-cage α helix, which seems to indicate that this structure is held not only by H-bonding but also by dispersion forces. The role of the electrostatic forces (and thus H-bonding) was surprisingly less apparent. When the electrostatic interactions were practically eliminated, the Trp−Pro distance oscillated between folded and unfolded structures and the α helical structure was retained to a surprisingly high degree. All these findings indicate that dispersion energy played an important role not only in the determination of the structure but also in the folding/unfolding equilibrium of the Trp-cage protein and probably also more generally in all proteins.

Acknowledgment. This work was a part of Research Project No. Z40550506 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and was supported by Grants No. LC512 and MSM6198959216 from the Ministry of Education, Youth and Sports of the Czech Republic and Grant No. 203/06/1727 from the Czech Science Foundation. This work was supported by the Instutional Research Concept No. AV0Z505200701 of the Academy of Sciences of the Czech Republic. The support of Praemium Academiae, Academy of Sciences of the Czech Republic, awarded to P.H. in 2007 is also acknowledged.

References and Notes

(1) Müller-Dethlefs, K.; Hobza, P. *Chem. Rev.* **2000**, *100*, 143.
(2) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608.
(3) Vondrášek, J.; Bendová, L.; Klusák, V.; Hobza, P. *J. Am. Chem. Soc.* **2005**, *127*, 2615.
(4) Černý, J.; Kabeláč, M.; Hobza, P. *J. Am. Chem. Soc.* **2008**, *130*, 16055.
(5) Malham, S. W. R.; Johnstone, S.; Bindgham, R. J.; Barratt, E.; Phillips, S. E. V.; Laughton, C. A.; Homans, S. *J. Am. Chem. Soc.* **2005**, *127*, 17061.
(6) Paschek, D.; Nymeyer, H.; Gardia, A. E. *J. Struct. Biol.* **2007**, *157*, 524.
(7) Paschek, D.; Hempel, S.; Gardia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17754.
(8) Biedermannová, L.; Riley, K. E.; Berka, K.; Hobza, P.; Vondrášek, J. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6350.
(9) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
(10) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701.
(11) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149.
(12) Case D. A. University of California: San Francisco, 2008.
(13) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1.
(14) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.

JP9004746