AQUEOUS SOLVATION OF PROTEIN SECONDARY STRUCTURES: DENSITY FUNCTIONAL THEORY STUDY

by

MATEUSZ MARIANSKI

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2013

This manuscript has been read and accepted for the
Graduate Faculty in Chemistry in satisfaction of the
dissertation requirement ot the degree of Doctor of Philosophy.

Chair of Examining Committee:

22^{nd} of April, 2013 — Dr. Joseph J. Dannenberg

Executive Officer:

23^{rd} of April, 2013 — Dr. Maria C. Tamargo

Supervisory Committee:

1. Dr. Yujia Xu
2. Dr. Seogjoo Yang

THE CITY UNIVERSITY OF NEW YORK

Abstract

Aqueous Solvation of Protein Secondary Structures:
Density Functional Theory Study

by

Mateusz Marianski

Adviser: Professor Joseph J. Dannenberg

In recent years, van der Waals forces have received considerable attention among the scientific community. It is hard to overestimate the significance of dispersion forces which are thought to play important roles in the energetics of biological molecules, such as DNA and peptides. However, the weakest of interactions is also the most difficult to approach by theoretical methods and has been troubling computational chemists for at least last two decades. In my thesis I will answer how well recently developed density functionals deal with the dispersion in the case study of dispersion-enhanced induction complexes, relative stability of π-stacking and hydrogen bonded dimers, and protein secondary structures. The presented results undermine the belief that recent widely-parametrized and/or dispersion-corrected functionals outperforms older well-established functionals, like famous B3LYP.

In the second part of my thesis I will focus on the influence of aqueous solvent on protein structures. Water is present in all biological systems, where it is not only a static medium of the reaction, but also an active part of the process called life, and it requires careful treatment. I compare models of implicit and explicit solvation for β-turns, α-helices, and β-sheets. I find that solvation by small water clusters can alter the molecular properties of gas phase molecules and continuous methods are not able to model all effects.

Acknowledgments

First I want to thank my mentor, Dr. Joseph J. Dannenberg, for shaping my view on the science, for giving me his guidance in the fields of theoretical and experimental chemistry, and being supportive on the personal level during the challenging years of my life.

I would like to thank Dr. Yujia Xu and Dr. Seogjoo Jang for their important insight into my work and Dr. Neepa Maitra for the great introduction into the physics of Density Functional Theory. I would also like to thank all current and former group members: Dr. Joshua Plumley, Dr. Midas Tsai, Dr. Nadya Kobko, Dr. Dipankar Roy, Dr. Jorge Ali-Torres and Dr. Gabor Pohl for their guidance and fruitful discussions.

Finally, I would like to thank my family for their support and understanding for the long process of education towards the Ph.D. degree.

Preface

Some of the work presented in my thesis has been already published elsewhere. Here I list the publications which base on the data included in the Chapter III and Chapter IV:

Section 3.2: "Comparison of some dispersion-corrected and traditional functionals with CCSD(T) and MP2 ab initio methods: Dispersion, induction and baisi set superposition error" by Dipankar Roy, Mateusz Marianski, Neepa T. Maitra, and J. J. Dannenberg, Journal of Chemical Physics, Volume 137, Pages 134109/1 - 134109/12, Year 2012

Section 3.3: "A Reinvestigation of the Dimer of para-Benzoquinone and Pyrimidine with MP2, CCSD(T), and DFT Using Functionals Including Those Designed to Describe Dispersion" by Mateusz Marianski, Antoni Oliva, and J. J. Dannenberg, Journal of Physical Chemistry A, Volume 116, Pages 8100-8105, Year 2012

Section 3.4: " Comparison of some dispersion-corrected and traditional functionals as applied to peptides and conformations of cyclohexane derivatives" by Mateusz Marianski, Amparo Asensio, and J. J. Dannenberg, Journal of Chemical Physics, Volume 137, Pages 044109/1 - 044109/7, Year 2012

Section 4.2: "A density functional theory evaluation of hydrophobic solvation: Ne, Ar and Kr in a 50-water cluster - Implications for the hydrophobic effect" by Nadya Kobko, Mateusz Marianski, Amparo Asensio, Robert Wieczorek, and J. J. Dannenberg, Computational & Theoretical Chemistry, Volume 990, Pages 214-221, Year 2012

Section 4.3: " Density Functional Theory Study of β-Hairpins in Antiparallel β-Sheets, a New Classification Based upon H-Bond Topology" by Dipankar Roy, Gabor Pohl, Jorge Ali-

Torres, Mateusz Marianski, and J. J. Dannenberg, Biochemistry, Volume 51, Pages 5387 - 5393, Year 2012

Section 4.4: "Aqueous Solvation of Polyalanine α-Helices with Specific Water Molecules and with the CPCM and SM5.2 Aqueous Continuum Models Using Density Functional Theory" by Mateusz Marianski, and J. J. Dannenberg, Journal of Physical Chemistry B, Volume 116, Pages 1437 - 1445, Year 2012

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The presented thesis constitutes the summary of my four years of work in Dr. Joseph J. Dannenberg group. The group is focused on "*the importance of H-bonding and cooperativity in determining the structures of natural (biological) materials such as peptides and unnatural materials such as self assembling nanomaterials*" [1] using Molecular Orbitals (MO's) and Density Functional Theory (DFT). The thesis is organized as follow:

In Chapter 2 I present the theoretical background of methods I use trough my thesis. First I introduce the foundations of Density Functional Theory, the milestone in the computational chemistry and physics [2]. It significantly reduces the cost of ab initio calculations comparing to wavefunction methods and without DFT I would not be able to carry out most of the calculations presented in the thesis. I also discuss the continuous solvation models, widely used in calculations in solution.

Modern DFT faces many challenges, one of them being description of the dispersion forces. In recent years many theoretical groups [3] tried to asses the problem of dispersion and designed a new generation of so-called dispersion-corrected funtionals. In Chapter 3 I evaluate performance of those functionals, along with functionals not designed to treat dispersion, for the biochemically related systems. First I took a look on complexes where dispersion and induction operate together to see how important dispersion is in the presence of other interactions. Then I look into the complexes that form different isomers (π-stacking and Hydrogen Bonded (HB)) stabilized by dispersion and electrostatic-induction, respectively. Since the detailed data on these structures has been published I was able to evaluate which functionals behave with agreement to the experimental

data. At last I apply the DFT to the protein secondary structures - α-helix and β-sheets – and examine which functionals performs the best with respect to the experimental data. All those thest provide me the important insight into the performance of the dispersion-corrected and traditional (non dispersion-corrected) functionals in the systems of the biological relevance. They also help me choose the best functional for study of protein hydration in Chapter 4.

Protein crystal structures consist many ordered (in specific positions, active centers, first hydration shells) or disordered (in spaces between crystallized peptides, not observed directly in experiment) water molecules. These crystallized waters are essential for proper functioning of proteins, providing the stabilizing factor to the protein structure. In Chapter 4 I focus on interactions between water and protein secondary structure. However first I analyze the water itself. The hydrophobic effect, responsible for many unique effects in water (for example protein folding), has been troubling scientists for decades and I try to explain it without invoking the entropic argument Next I shift to the hydration of protein secondary structures: β-turns, α-helices and β-sheets. I reinvestigate the standard classification of β-turns proposed by Venkatachalam[4] in the gas phase and in the solution. Then I try to answer what is the mechanism of hydration of α-helices and, as the last project, I study the the β-sheets formation in the aqueous solvent.

My conclusions are summarized in Chapter 5.

Chapter 2

Theoretical Background

Human subcultures are nested fractally.
There's no bottom.

Crazy Straw Design[5]

2.1 Introduction

The foregoing chapter does not aim to review vast literature of the subjects by any means. Quoting Kieron Burke from his brilliant Perspective on Density Functional Theory[6]: "*Because DFT is now applied so broadly, no such article can hope to be comprehensive*". I rather plan to present a summary of my own understanding of the fascinating world of the modern theories than follow rigorously their derivations. If the potential reader is more interested in the formal aspects I encourage him to look into literature. Many brilliant reviews on Density Functional Theory (DFT) and Continuous Solvation Models (CSM's) were published recently (for example look: [7] for DFT and [8], [9], [10], and [11]). Hereafter, I simply want to support the reader with knowledge necessary to understand concepts presented in the following chapters.

The chapter is organized as follows:

1. At first in section 2.2 I want to familiarize the reader with basic concepts of Density Functional Theory (DFT), prove Hohenberg-Kohn Theorem, and derive Kohn-Sham equation for the systems of non-interacting electrons. Although DFT is recognized as an incredible success story, distinguished by Nobel Prize in Chemistry for Walter Kohn And John Pople in 1998 it is

not a complete story yet. Since the exact functional is not known (and, probably, it will never be known), DFT developers design their own approximations of the density functionals. The current challenges involve the stability of the negative ions, prediction of the band gaps in insulators and semiconductors, description of strongly correlated systems like metals and metal oxides, and weak dispersion forces. The last challenge lies in the special focus of my thesis since its close relationship to proteins and DNA stability ([12], [13]).

2. In section 2.3 I am going to introduce the concept of Continuum Solvation Models (CSM's). They consist of very efficient approach to treating the solvent without the necessity to significantly increase the complexity of the system. I will present the basic idea behind them and derive the milestone model in condensed phase calculation - the family of Polarizable Continuum Models (PCM). Moreover, in chapter 4 I will try to relate these models to real solutions, arguing how well they can reflect to our current understanding of the condensed phase chemistry.

2.2 Foundations of Density Functional Theory

Physicist's perspective: Machinery

The starting point of any discussion of DFT is the Hohenberg-Kohn theorem [14], which states that stationary many-particle system can be fully characterized by the ground state density. Let me briefly present the proof of Hohenberg-Kohn theorem on the following pages. In order to proove it, first (1) I will show that two unique potentials V_1 and V_2 cannot be mapped to the same ground state wavefunction ψ (left part of figure 2.1) and then (2) I will prove that there is one-to-one correspondence between wavefunction ψ and ground state density n (right side of the figure 2.1).

First, let's choose two physically different potentials, $V_1(r)$ and $V_2(r)$ ($V_1(r) = V_2(r) + c(r)$ where $c(r)$ is not a constant) so that two energy operators $\hat{H}_1$ and $\hat{H}_2$ associated with these potentials have the same ground state density $n_1(r) = n_2(r) = n(r)$. Let's write Schroedinger's equation for these Hamiltonians (using Dirac's notation):

$$\hat{H}_1|\psi_1\rangle = E_1|\psi_1\rangle \tag{2.1}$$

$$\hat{H}_2|\psi_1\rangle = E_2|\psi_1\rangle \tag{2.2}$$

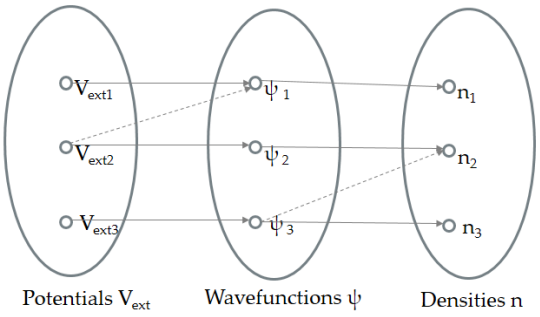


Figure 2.1: Correspondence between external potentials, ground state wavefunctions and ground state electron density. The solid arrows indicate one-to-one mapping, while the dashed arrows represent forbiden correspondence.

where ψ's and E's are the eigenfunctions and eigenvalues, respectively, of corresponding Hamiltionians. If we subtract the bottom equation from the top, we will receive:

$$(\hat{H}_1 - \hat{H}_2)|\psi_1\rangle = (E_1 - E_2)|\psi_1\rangle. \tag{2.3}$$

The left hand side of above equation depends on r as we chose the form of the potential, while the right side of the equation is the constant. Therefore we see that two unique potentials cannot correspond to the same wavefunction ϕ as shown in the left part of the figure 2.1.

In the next step I show the proof for the right side of the figure, i.e. two different non-degenerated wavefunctions cannot yield the same electron density (n).

From the variational prenciple we know that if the ψ_1 minimizes the E_1 then any other wavefunction will yield larger energy:

$$E_1 = \langle\psi_1|\hat{H}_1|\psi_1\rangle < \langle\psi_2|\hat{H}_1|\psi_2\rangle = \langle\psi_2|\hat{H}_2 + \hat{H}_1 - \hat{H}_2|\psi_2\rangle \tag{2.4}$$

$$E_1 < E_2 + \langle\psi_2|\hat{V}_1 - \hat{V}_2|\psi_2\rangle = E_2 + \int n(r)(V_1 - V_2)\,\mathrm{d}^3r \tag{2.5}$$

where the $\langle\psi|\hat{V}_k|\psi\rangle = \int n(r)(V_k)\,\mathrm{d}^3r$ is the definition of the one-body potential operator. Simultaneously we can conduct the same argument for E_2:

$$E_2 < E_1 + \langle\psi_1|\hat{V}_2 - \hat{V}_1|\psi_1\rangle = E_1 + \int n(r)(V_2 - V_1)\,\mathrm{d}^3r \tag{2.6}$$

If we add the above equations, we will receive:

$$E_1 + E_2 < E_1 + E_2 \tag{2.7}$$

which is an obvious contradiction, therefore we see that there is one-to-one correspondence between the wavefunction ψ and the density n(r).

Therefore, we learned that the electron density, in principle, contains the same amount of information as the wavefunction does and the only challenge is the way of extracting this information. In other words, V_{ext}, ψ and n(r) determine each other uniquely and knowledge of one of these terms give us the full information of the system. From the point of view of chemistry, the most interesting information is the energy and here comes the help from another brilliant work which is the foundation of the modern density functional theory – Kohn-Sham equations [15].

Let's go back to the Schrodinger equation rewritten with the use of Hohenberg Kohn theorem. We look for the electron density minimizing the energy of the system at some given external potential V_{ext}:

$$E = \min_n \langle \psi[n] | \hat{T} + \hat{W} + \hat{V}_{ext} | \psi[n] \rangle = \min_n (F[n] + \langle \psi | \hat{V}_{ext} | \psi \rangle) \tag{2.8}$$

The F[n] functional is called universal functional which does not depend on the additive potential V_{ext} but combines only the kinetic energy term ($\hat{T}$) and the electron – electron interaction ($\hat{W}$)). The exact form of the F[n] functional is not known, however, even if we ever knew it (and thus be able to solve Schrodinger's equation exactly) it would probably have very complicated form and it would be impractical to use. Therefore we have to find the clever way of treating the F[n].

Kohn and Sham proposed to represent the system as a set of the non-interacting particles that yields the same electron density $n(r)$ as the fully interacting system. If we do so, we will receive the set of the one-particle equations that we are able to solve:

$$(-\frac{1}{2}\nabla^2 + V_s(r))\phi_i(r)) = \epsilon_i \phi_i(r) \tag{2.9}$$

and:

$$V_s[n](r) = V_{ext}(r) + V_H[n](r) + V_{xc}[n](r) \tag{2.10}$$

where V_{ext} is the external electrostatic potential (in the chemistry - due to the nuclei), V_H

is the Coulomb repulsion between two electrons and V_{xc} is the exchange-correlation potential responsible for all non-coulombic iterations in the $\hat{W}$ term.

The energy of the non-interacting electron is simply given by:

$$E_s = T_s[n] + \int n(r)(V_s)[n]\,\mathrm{d}^3 r \tag{2.11}$$

It is very important to point out that I did not make any approximation till this point. DFT is, in principle, an exact theory, the reformulation of the Quantum Mechanics using electron denisty $n(r)$ in place of the wavefunction ψ, which are equivalent due to Hohenberg-Kohn theorm. It should be also understood that DFT is a mathematical construct representing the non-interacting particles yielding the same electron density as the interacting system. However, what the wavefunction mechanics is, if not the mathematical construct as well? The wavefunction cannot be observed explicitly either (only its squared modulus, the probability of finding an electron), it is only the mathematical model reflecting the properties of the atomic system.

Now, the choice of the form of the exchange-correlation potential, V_{xc}, is the moment when we do make the first approximation. Since the form of the potential is not known, we are forced to approximate it in some plausible form. This is the moment when a casual user touches the essence of Density Functional Theory – the choice of the functional used is the choice of the form of exchange-correlation potential used for solving the Kohn-Sham equations and calculating the energy of the system. I will continue my discussion in the next section.

Chemist's perspective: Taxonomy

The big advantage of the wavefunction theory over the DFT is the clarity. Let me quote the fragment from Kieron Burke's perspective on DFT again: "*Throughout its history, DFT has provided approximations that work for some problems and fail for others, in largely mysterious way*".

In general, while the wavefunction theory provides a clear way of improving accuracy of the results by simply adding additional corrections to the Hamiltonian/wavefunction, the casual DFT user stands in the front of the sea of possible choices of approximations to the exchange-correlation potential. In fact, everybody can create their own form of the functional and use it to solve specific problem, however it is not possible to state that one functional will be more accurate than another.

This explains the popularity of a very specific class of the scientific papers which benchmark functionals against some reference data in order to find the one best suited for a given task (for example please look into: [16], [17], [18], [19], and [20]).

I think the DFT can be compared to modern Internet, where everything is equally accessible for everybody. However, like in Internet which concentrates traffic around the most popular websites [21], most users choose the most popular functionals, like B3LYP or M06-2X in chemistry or PBE and HSE in solid state physics [6].

One of the most famous examples of the attempts to introduce some order into the anarchy of DFT methods is the concept introduced by Perdew called "Jacob's ladder" [22]. It places the functionals on different rungs of the ladder connecting the "Hartree World" and the "Chemical Heaven". The functionals are categorized based on the input imformation:

LDA: The lowest rung is occupied by the Local Density Approximation. Despite the equation uses only the local density to calculate the exchange and correlation energies, it proved to work surprisingly accurately [23]. It is still widely used in the solid state because it provides very accurate results at the low computational cost.

GGA: The next step was the incorporation of the (reduced) gradients of the electron density in the approach called General Gradients Approximation. These functionals are semi-local, they use both the local density and its first gradient in the calculations of E_{xc} term. It took a long time to understood why the first attempts to incorporate gradient of the electron density had failed [23], however once they were successfully implemented, they proved to be the big improvement over the LDA.

meta-GGA: They use the second derivative of the electron density (related to the kinetic energy term), however meta-GGA's did not provide a significant increase in the accuracy over the GGA's.

hybrid-GGA: It has been understood via adiabatic connection that [23] in order to improve accuracy of the exchange functional which underestimate the exchange energy it should be mixed with the exact exchange from Hartree Fock calculations [24]. The most popular functionals widely used in the chemistry and physics (B3LYP, PBE1PBE) belong to this rung.

Beyond... The last rung contain functionals that go beyond the standard approximations and use additional input besided the election density like unoccupied orbitals (Random Phase Approximations [25]), or correlation energy calculated from MP2 contribution (so called double-hybrids, for example B2PLYP functional, [26]).

Above classification does not mean to state that hybrid GGA's will always be performing better than the LDA, however if one measures an average error over a big data set, he should observe statistical improvement. It is essential for DFT user to know for what system the functional are suited best, for instance the LDA will perform reasonably well for metal clusters but the exact exchange in hybrid GGA's will almost certainly fail.

At this point I would like to introduce my own classification of functionals. It is based upon the philosophy the functionals were constructed rather than their formal equations. As a useful comparison I would like to mention the development of semi-empirical molecular orbitals theories. Dewar et al. used gas phase enthalpies of formation taken from reliable experimental data as their primary source of data. Until they developed AM1 [27] semiempirical methods of Dewar et al. did not provide reasonable descriptions of molecules and reactions. Even then, the method only worked well for molecules containing atoms for which there were good data in differing molecular environments. For example, molecules containing halogens were not as well described as those containing only C, H, O, and N, probably due to the combination of their mono-valency and relative lack of reliable experimental heats of formation. Furthermore, Dewar et al. used a training set of molecules to parametrize and a large set to test. They insisted that the parameters for each atom follow a logical progression expected from their positions in the periodic table. Thus, they did not simply use the best fit to the data.

Keeping in mind what I learned from the closely related field, I would like to propose classification of functionals based on the number of parameters they include in the building the exchange-correlation term rather than they performance (in parenthesis I list the functionals studied in the chapter 3 of my thesis):

Non-empirical functionals (PBE1PBE): They do not utilize any empirical parameters in the equations but are solely based on the first principle calculations. Few properties of the exchange and correlation terms are known from the studies of homogenous electron gas [23] and

non-empirical functionals are designed to fulfill these properties and intrapolate behavior of the electrons between the known boundaries. The examples are the PBE, PW91 and the exchange part (X) in the X3LYP functional. Despite that they are not parameterized against any data sets, their robust construction enable their agreement with high level calculations and experimental data as accurately as their empirical cousins.

Semi-empirical functional (B3LYP, X3LYP, B2PLYP): These functionals use some empirical data as a training set and they contain limited number of parameters. They B3LYP is the best example of the semi-empirical functional. The expression for the exchange-correlation energy is given:

$$E_{xc} = E_{xc}^{LDA} + a_0(E_x^{HF} - E_x^{LDA}) + a_x(E_x^{GGA} - E_x^{LDA}) + a_c(E_c^{GGA} - E_c^{LDA}) \qquad (2.12)$$

where $a_0 = 0.20$, $a_x = 0.72$, and $a_c = 0.81$ are empirically fitted parameters to atomization energies [24] which mix the energies contributions from the theoretical calculations (LDA, B88 exchange functional, LYP correlation functional[28] and Hartree-Fock exact exchange). Two important things should be noted here: (1) The parameters combine the energies calculated from theoretically derived equations for the energies, and (2) the training (small) and testing (big) sets are separated, the functional is not trained against the large data set, but just on a small set of data and then its performance is compared with larger testing set [29].

Empirical functionals (M05, M05-2X, M06, M06-2X, ωB97x-D, B97D): They use excessive number of parameters (up to 39 in M06-2X functional [30]), that provide very good fit to big training sets [25]. The equation for the energy combines many terms that are responsible for specific interactions between the atom and for additional dispersion corrections. Since they are widely parameterized, they usually provide (statistically) very good fit to the test set (which are usually included in the training sets as well).

In the first part of the thesis I will try to answer if the newer functionals (mostly empirical) provide better assessment of the studied molecules than the non-empirical ones when compared to the experimental data rather than theoretical test sets.

2.3 Continuous Solvation Models

How to approach the problem

I stated in the introduction that water is an extremely important piece in the puzzle of protein world. However its unique properties like poralizability, hydrogen bonding and hydrophobic effect make it extremely difficult to reliably include in calculations. There are few common approaches in computations. In general we can divide them into implicit, explicit and hybrid models, where each of them posses their own advantages and disadvantages.

In the most coarse approximation (in terms of simplicity model, not cost of accuracy of results) a solute in placed in a cavity in a continuum medium which is described with a dielectric constant. In this case the solvent does not posses any discrete structure and is described with a set of arbitrary macroscopic parameters and the solute is treated explicitly with desired method. These method might vary from molecular mechanics among semiempirical methods up to wavefunction theory. Implicit solvent, in comparison to explicit solvation, drastically reduces the computational cost and allows to describe desired part of the system using high level molecular orbital theory. There are many successful models that vary in the approximations and complexity. However properly used, the most popular implicit solvent models (SMx family [31], poralizable continuum models [9] and COSMO models [32]) provide an accuracy better that 1 kcal/mol for test sets of different classes of molecules.

The explicit methods constitute the opposite approach to the problem. In this case solvent is treated explicitly which means solute is put in a box full of solvent molecules. Again there is a full spectrum of methods available for this approach: solvent might be treated with molecular mechanics (popular TIPxP models for water [33]) or desired quantum mechanical levels (even including dynamics using Car-Parrinelo methods). However explicit consideration of the solvent dramatically complicates the problem by increasing number of degrees of freedom. Also it is not clear how many waters are necessary to simulate both first hydration shell and bulk hydration. The former introduces local effects connected with discrete structure of the solvent while latter adds macroscopic properties like bulk dipole moment. This methods are the most popular in molecular dynamics where water structural properties are approximation as three fixed charges in given geometry allows quick solution to Newton's equations of motion. Recently attention was

brought to new approach of embedding high level quantum methods with low level MM mechanics calculations however these sophisticated methods require great attention in avoiding artifacts at the border between different levels of theory.

The third method combines advantages of both previous methods while reducing their problems. It combines the first hydration shell which is considered explicitly with implicit solvent for bulk solvation. It allows to reduce major problem of continuum methods – poor description of specific (H-bonding) interactions – by enhancing critical region with explicit waters [34], [35], [36], [37]. Few additional waters does not increase computational cost significantly but it can greatly change how solvation affects the solute. It is successfully used in calculations of free energy of transfer of charged species which provide technical problems in implicit model approach.

Basic Formulation of Continuous Solvation Model

The continuous solvation models (CSM) divide the solvated system into two subsystems: solute (M) and solvent (S). The Hamiltionian might be written as:

$$\hat{H}^{MS}(R,r) = \hat{H}^{M}(R) + \hat{H}^{S}(r) + \hat{H}^{int}(R,r) \tag{2.13}$$

where R and r are, respectively, the solute's and solvent's coordinates and $\hat{H}^{int}$ is the coupling between two subsystems. However since we are just interested in the solute M, we can reduce the full Hamiltionian to effective hamiltionan, written as:

$$\hat{H}^{MS}_{eff}(R,r) = \hat{H}^{M}(R) + \hat{H}^{int}(R,r) \tag{2.14}$$

As one can see, I simplified the system by removing the Hamiltonian of the solvent and assuming that it is sufficient for the good descitpion of the interation. $\hat{H}^{int}(R,r)$ still have the dependence on the solvent's coordinate r, which still increases the number of degrees of freedom. However we can even further eliminate the solvent by introdicing some appropriate solvent response function based on the solute's coordinate R. Then, the $\hat{H}^{int}(R,r)$ will have a form of the potential, descibing the solvent's influence on the solute.

Now, we can define the basics of the CSM. The solute (or simply some charge distribution ρ_M is placed in the cavity under an infinite dilution in the homogenous continuum medium descibed

by the set of parameters. The solute can be treated on any level of theory, but the solute - solvent interaction are limited just to electrostatic. The contribution form the other sources (dispersion, repulsion, cavitation energies, entropy) requires separate treatment.

I want to emphasize that we represent the solvation problem as the placing the solute in the cavity digged in the inifinite continuum medium described with some dielectric constant ϵ. This is the model we use and the size (or the shape) of the cavity is one of the most important parameters we are going to use because it affects the boundary position and the strength of the interaction. The atomic radii - the experimental van der Waals Pauling or Bondi [38][39], defined for force fields UFF and UA0[40], or parameterized from ab initio calculations uaks and uahf [41] - are used for the cavity's definiton, sometimes corrected by Solvent Accessible Surface (SAS) or Solvent Excluded Surface (SES). In general, anyone can define their own set of parameters, however it is suggested that van der Waals radii have the most physical meaning [32] since they provide good contacts between the electron density and the surface of the cavity.

Derivation of the Polarizable Continuum Model

The basic solvation model requires the solution to a classical electrostatic problem (Poisson equation) nested within a quantum mechanic framework. For the homogenous continuum solvent we can simplify Poisson equation:

$$-\overrightarrow{\nabla}[\epsilon(\overrightarrow{r})\nabla\overrightarrow{V}(\overrightarrow{r})] = 4\pi\rho_M(\overrightarrow{r}) \tag{2.15}$$

into:

$$\nabla^2 V(\overrightarrow{r}) = 4\pi\rho_M(\overrightarrow{r}) \tag{2.16}$$

within the cavity, and:

$$\epsilon\nabla^2 V(\overrightarrow{r}) = 0 \tag{2.17}$$

outside the cavity, where the V is the sum of the elctrostatic potential V_M of the charge distribution ρ_M and the reaction potential V_R generated by the polarization of the dielectric medium.

In order to solve the Poisson equation we have to satisfy the jump conditions at the cavity surface Γ:

$$[V] = V_{inside} - V_{outside} = 0 \tag{2.18}$$

$$[\partial V] = (\frac{\partial V}{\partial \overrightarrow{n}})_{inside} - (\frac{\partial V}{\partial \overrightarrow{n}})_{outside} = 0 \tag{2.19}$$

The first jump condition ensures the continuity of the potentail accross the surface, while the second ensures the continuity of the gradient of the potenital that is perpendicular to the cavity surface ($\overrightarrow{n}$ is the normal outward pointing vector at the cavity surface). The above equations are the basic foundation of any continuous solvation model.

From the jump conditions, we can derive the apparent surface charge $\sigma(r)$ distribution spread on the cavity surface (where $\overrightarrow{s}$ is the position on the space limited to the surface Γ). The apparent charge defines the potential:

$$V_\sigma(r) = \int_\Gamma \frac{\sigma(\overrightarrow{s})}{|\overrightarrow{r} - \overrightarrow{s}|} \mathrm{d}^2 s \tag{2.20}$$

which is exactly the reaction potential (V_R) defined at the begining of the section. The basic Polarizable Solvent Model definition is given by the equation:

$$\sigma(s) = \frac{\epsilon - 1}{4\pi\epsilon} \frac{\partial}{\partial n} (V_M + V_\sigma)_{inside} \tag{2.21}$$

which is solved iteratively till the convergence between the potential $V_M + V_\sigma$ and apparent surface charge is achieved.

In my studies I use conductor-like [42],[43] reformulation of PCM (CPCM). CPCM assumes that the medium is a perfect conductor and can be described with dielectric constant $\epsilon = \infty$, which strongly modifies the boundary conditions. The total potential $V(r)$ cancels out at the cavity surface and the new apparent charges ($\sigma^*(s)$ can be determined from the local value of the potential. In order to recover the charges at the finite value of ϵ, the $\sigma^*(s)$ is scaled by the $\frac{\epsilon-1}{\epsilon+k}$, where k is determined to be 0.5 [43]. CPCM model provides better numerical stability and decreases the CPU time for the solvent calculations when compared to the standard PCM method. Moreover, for high dielectric constant it predicts essentially the same values for electrostatic solvation.

At the end I want to remind that the continuous solvation models calculate only the electrostatic contribution to the solvation energy. Other contributions (cavitation [44], repulstion, and

dispersion [9]) have to be taken into account by another formula which are not clear because they do not relate to the physically measurable quantity. Truhlar proposed an interesting solution to the problem by gathering all terms in one “CDS“(Cavitation-Dispersion-Entropy) term which is parameterized to, together with electrostic term, yield the experimental free energies , as implemented in the SMx.x [31] and SMD solvation models [45].

Chapter 3

Critical Evaluation of the Dispersion Corrected Density Functionals

"And now what's two and two?"
"Seven!" replied the machine.

The Cyberiad [46]

3.1 Introduction

In the world of Density Functional Theory, the methodology and development permeate each other a lot. It is important to keep track of new trends in the constantly evolving field of the computational chemistry. Few years ago several seminal papers (for example please look into [47], [48], [49]) foreshadowed a rapid change in the methodology. However, the world of DFT is governed by its own rules (vide Burke perspective on DFT [6]) and systematic and clear way to improvement is not one of them. For years it was known that current state-of-the-art functionals, while successful in description of a variety of the systems - starting from solid state, through catalysts up to proteins, fail in the description of the weakest forces known as the van der Waals dispersion forces. Just recently, new class of dispersion corrected functionals (in theory they assess the problem of the weakest interactions) gained on popularity. In general they provide better agreement with the high level ab initio calculations collected in various data sets which describe different kinds of properties or interactions. Nonetheless, the question how important is dispersion still remains not

answered, it has not been showed if current models of simple dependence of dispersion forces on R^{-6} are sufficient and accurate.

Therefore the proper evaluation of new so-called dispersion corrected functionals is an important part of my work; a careful and wise choice of the method might determine the output (or even the success) of the entire project. In essence, I want to avoid a common mistake made by many theoreticians – too deep trust in infallibility of their methods. In my approach validation against experimental data, not high level calculations, is the primary factor in the critical examination of the results.

Three systems were chosen for examination of various new and traditional functionals, two of them being systems of biochemical relevance: (1) complexes when the dispersion is coupled with induction, (2) a dimer of p-benzoquinone and pyrimidine, and (3) a model polyalanine secondary structures of α-helix and β-sheet. The first model gives us the insinght into the importance of dispersion interactions when other forces are present in the complex. The second system can measure the relative strength of the electrostatic (H-bonded) and dispersive (π-stacking) interactions while the third model provides insight into the stability of proteins secondary structures as described by traditional functionals and newer generation of dispersion corrected ones.

3.2 Methods

All calculations were performed using the Gaussian 09 suite of computer programs [50]. Unless it is explicitly stated, all structures were optimized with the respect to all internal coordinates and the harmonic frequencies were calculated to: (1) confirm the existence of the minimum on the Potential Energy Surface (PES) and (2) obtain the gas phase enthalpies that correspond to those structures. The counterpoise (CP) correction (due to the incompleteness of the basis set - Basis Set Superposition Error, BSSE) was calculated either on the CP-corrected surface or as a single point a posteriori calculations or both. If the CP-correction procedure was used, the frequencies were recalculated on the corrected surface.

I divide functionals into two groups: (1) Dispersion – Corrected Functionals which are abbreviated hereafter as DF's; and (2) Traditional Functionals (indicated as TF's) that were not

specifically designed to treat the dispersion interactions. I considered the following DF's (sections where these functionals were used are shown in parenthesis): B97D (3.3, 3.4, 3.5) [51], ωB97x-D (3.3, 3.4, 3.5) [52], M06 (3.3, 3.4, 3.5), M06-2X (3.3, 3.4, 3.5), M06L (3.3, 3.4) [30], B2PLYP (3.3)[26], B2PLYP-D (3.3)[53]; and the following TF's: B3LYP (3.3, 3.4, 3.5)[24], X3LYP (3.3, 3.4, 3.5)[54], PBE1PBE (3.3, 3.4, 3.5)[55]. Various ab initio calculations were performed where it was possible and reasonable: Hartree-Fock (3.3), MP2 (3.3, 3.4) and CCSD(T) (3.3, 3.4). For the more detailed description of the listed DFT methods please refer to the section 2.2 and the original papers.

Besides the full DFT calculation, the ONIOM method was used in the section 3.5 as a comparison. The ONIOM divides the system into up to three layers and each of them can be addressed with the different level of theory. Thus, we are able to treat the essential part of the system at the high level theory (the network of the peptide hydrogen bonds, herein) and use the less computionally demanding semi-empirical AM1 method for the less critical parts (alanine side chain).

The correlation-consistent aug-cc-pVXZ (X=D, T, Q and 5) family of basis sets was used in section 3.3. The small size of the system allowed me to use larger basis sets up to the 5-zeta quality. In section 3.4, two basis sets were used: the d95++** double zeta with extra sets of polarized and diffused functions on all atoms, and the triple zeta correlation-consistent cc-pVTZ. These two basis sets allow the observation of the effects of improved contraction and the importance of extra diffuse functions. Size of the system studied in section 3.5 imposes serious limitation for plausible basis sets, therefore I decided to use the well-established double zeta d95** basis set which was shown to work extremely well for the hydrogen bonds in water [56] and peptides [57] [58].

For many DF's I found that the standard grid used in integration of electron density is not sufficient. The optimized structures contained low lying imaginary frequencies and predicted vibrations were inconsistent. Therefore all complexes in sections 3.3 and 3.4 were recalculated using the finest '99974' grid available in Gaussian 09, and the very tight and tight convergence criteria in sections 3.3 and 3.4 respectively were applied. In section 3.5 only structures containing imaginary frequencies were recalculated using '99974' grid and the tight convergence criteria due to the size of the system. This procedure fixed the imaginary frequency problem. I will comment on this issue in more detail in the following sections.

In section 3.3 I discuss more detailed analysis of a molecular dipole moment and a magnitude of a charge transfer between species. The MP2 dipoles were obtained using the "density" keyword. The Mulliken populations used for charge transfer were calculated the following way: at first, orbitals were projected to the minimum basis set (pop=MBS) and then artefactual electron transfer to ghost orbitals was removed via CP procedure.

3.3 The Role of Dispersion and Induction in van der Waals Complexes

Introduction

Dispersion can be thought of as a time dependent polarization, when electron density in both entities is shifted in the same direction creating a dipole-dipole (or higher multipole) interaction. In case of two molecules not containing any permanent multipole moment the probability of a shift in one direction is equal to the probability of a shift in another, and thus complex does not contain any multipole moment. The mutual dispersive polarization requires electrons to respond to the density changes that occur on another molecule which is a fully non-local effect. Therefore, the electron correlation is necessary to describe the dispersive effects and, in principle, the Single Slater Determinant (SSD) based Hartree-Fock method is not able to capture this interaction.

Nevertheless, when the entity is placed in some external potential, the polarization of the electron density is no longer symmetrical and the electric field will induce a molecular dipole moment. Now, moving on to the molecular world, let me consider the non-polar electron density a noble gas atom, for example neon, and the electric field might be produced by a molecule with a permanent dipole moment, for example by hydrogen fluoride, HF. The polarization at neon caused by HF will result in the attractive interaction which is called induction. The induction does not require the electron correlation and, in principle, it can be described by SSD methods.

In the cases both of dispersion and induction, the energy of the system will depend upon the polarizability of the non-polar atoms and the -6th power of the distance r between the entities. Since the polarization depends on the polarizability ρ x r^{-3}, the strength of the dispersive interaction depends on ρ_a x ρ_b x r^{-6} (vide Casimir-Polder equation in physics [59]). On the other hand, the induction depends on the induced dipole of non-polar atom, ρ x r^{-3}, interacting with the electric field produced by a molecule with a dipole moment, μ x r^{-3}. In both examples we obtain the r^{-6}

dependence.

It is important to point out that the dispersion might be influenced by the induction. As pointed in the previous paragraph, the polarization between two neons in the van der Waals complex is symmetrical, having equal probabilities of on-axis electron polarization in either direction. In the case of Ne-HF complex, the electron polarization at neon might either increase the dipole moment and lower the overall energy, or decrease it and increase the energy of the complex. The different energy output breaks the symmetry in polarization, favoring the former and augmenting the inductive interaction.

In this section I look at various functionals to evaluate and examine if they properly account for dispersion, especially when the dispersion is augmented by the induction. The work is divided into two parts:

1. Calculations on the van der Waals He_2, Ne_2 and Ar_2 complexes which interact solely via dispersion.
2. Calculation on the complexes where the dispersion and induction operate together. The neon, methane and 2-butyne were chosen as a non-polar entity and placed in the presence of HF and LiF.

I want to emphasize that the main focus lies on newer functionals designed to treat dispersion and my goal is to compare their performance with the high level ab initio theory, up to CCSD(T)/aug-cc-pV5Z level.

The detailed methodology is described in preceding section. All calculations were performed using the finest grid available in Gaussian 09, optimizations were carried on the non-CP corrected and the CP-corrected potential energy surfaces and harmonic vibrations were calculated to confirm stable minima. The complexes of polar and non-polar molecules were optimized using aug-cc-pVTZ basis set and single point calculations using larger basis sets were performed on the optimized geometries. In order to avoid ambiguity between hydrogen fluoride (HF) and Hartree-Fock method (abbreviated as HF too) I will use RHF (Restricted HF) abbreviation for the latter.

Noble gas dimers

In principle, Molecular Orbitals (MO's) methods should predict stronger interactions with larger basis sets and the BSSE should decrease as the basis functions span the vector space more efficiently. In essence, the BSSE should disappear for the infinite basis set. Along with decreasing interaction energy, the intermolecular distance should decrease as well. Despite all ab initio methods follow these qualitative expectations, I observed that none of the DFT methods meet these predictions.

The highest level CCSD(T)/aug-cc-pV5Z calculations predicted He_2, Ne_2 and Ar_2 interactions to be -20, -74 and -260 cal/mol respectively and experimentally determined values for these dimers are -0.002, -86 and -262 cal/mol. While the Ar_2 dimer stands in a perfect agreement with the experimental data and Ne_2 complex is somehow smaller (by 12 cal/mol which is equal to the dimer's BSSE and Complete Basis Set extrapolation would improve the strength of predicted interaction), the He_2 dimer is 10^3 times overestimated. However, I should note that the reported experimental distance, 52Å, includes zero point vibrations, while we calculated the depth of the potential well, therefore a direct comparison to experiment is impossible. Nevertheless we leave calculated energy for the He_2 dimer as a reference point for other DFT and ab initio calculations.

Other frequently used reference method, MP2, predicts weaker interactions for He_2 and Ne_2 (-13 and -49 versus -20 and -74 cal/mol) than CCSD(T), but overestimates it by 42 cal/mol for Ar dimer. Among all these dimers, the Ne_2 dimer represents the second row atoms the best and the finding that the absolute energy of interaction between two neons is smaller than one found experimentally and calculated by CCSD(T) contradicts the common knowledge that MP2 in principle overestimates dispersion.

For the Ne_2 dimer, the X3LYP, PBE1PBE, M06L and ωB97x-D functionals predict energies in a good agreement with CCSD(T) method (-73, -72, -69 and -69 cal/mol respectively). Remaining TF - B3LYP and the double-hybrid B2PLYP do not or hardly do find any attraction between two neons. On the other hand remaining DF's severely overestimate the interaction energy, listing them in the order of decreasing energy (all values are given in cal/mol): B2PLYP-D (-108), M05-2X (-118), M06 (-128), B97D (-156), M06-2X (-165) and M05 (-176). Accurate performance of few functionals might be attributed to the fact that neon dimer is often present in the training sets because the neon's electrons configuration resembles that for closed shell second row atoms in molecules.

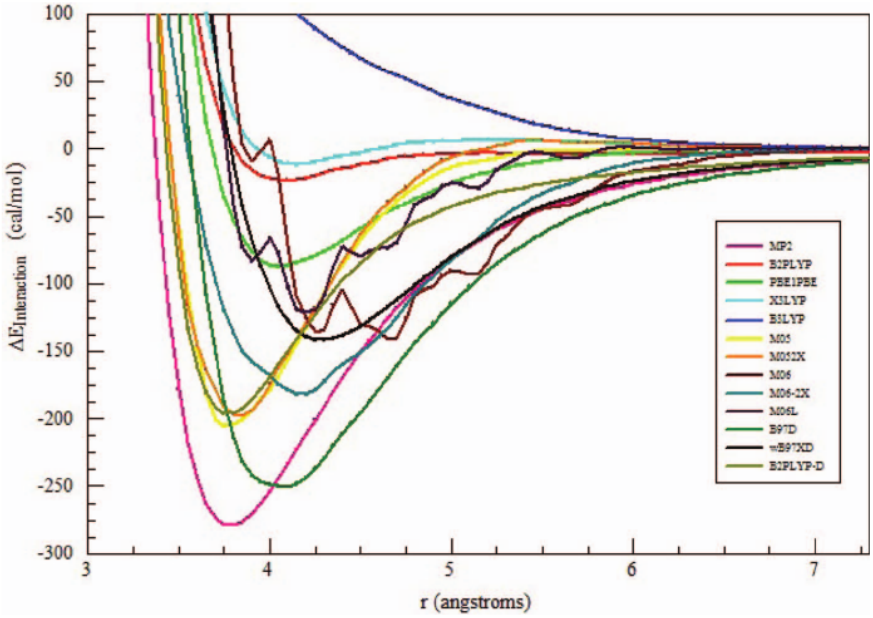


Figure 3.1: Comparison of scans for Ar_2 using different methods

It is worth noting that many of DF's give unexpected results with different basis sets which could not have been attributed to BSSE or energy changes and I decided to investigate burdensome complexes in greater detail. I found that some of the PES's contain multiple minima, for example the argon dimer's binding curve is shown in the figure 3.1. One can see that M06, M06L and M06-2X do not predict smooth binding curves and first two find multiple minima. This is a common problem for meta-GGA's which require second derivative of electron density as an input for energy evaluation. The problem was observed before and finer grid was suggested as a solution to the problem [60], but the proposed remedy does not help in all cases. Although multiple minima are easy to analyze for the two atoms energy curve, in a case of more complicated systems the problem is more cumbersome and undermines the confidence into predicted results.

Polar and non-polar species

In the next step I analyze the complexes in which the dispersion is coupled with the induction. I performed the calculations using few non-polar species (neon, methane and 2-butyne) in the presence of a molecule with a permanent dipole moment (HF and LiF with either H or Li, or F facing the non-polar entity). Figure 3.2 shows the structures of the investigated complexes and the results are presented in the tables 3.1 - 3.4.

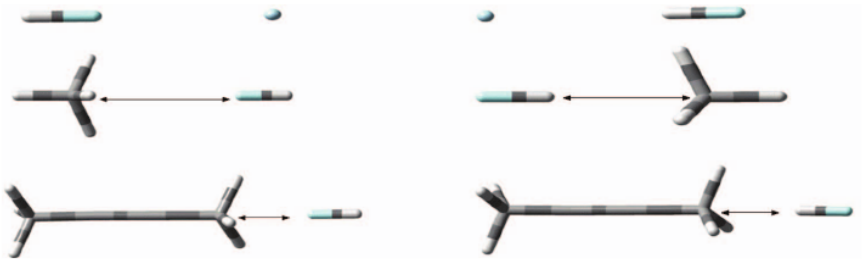


Figure 3.2: Interaction geometries for HF with Ne, Methane and 2-Butyne.

Table 3.1: Interaction energies and BSSE (kcal/mol) for complexes of Ne, methane, and 2-butyne with HF and LiF using CCSD(T)aug-cc-pVXZ with different values of X. Geometries are fixed at the MP2/aug-cc-pVTZ optimized structures on CP-corrected surface.

	HF						LiF					
	F-facing			H-facing			F-facing			Li-facing		
X	T	Q	5	T	Q	5	T	Q	5	T	Q	5
ΔE_{int}												
Ne	-0.10	-0.12	-0.13	-0.21	-0.22	-0.23	-0.14	-0.17	-0.18	-0.62	-0.65	-0.68
Methane	-0.26	-0.27		-1.56	-1.62		-0.43	-0.45		-3.74	-3.76	
2-Butyne	-0.71	-0.73		-0.43	-0.45		-2.10	-2.15		-1.24	-1.24	
BSSE												
Ne	0.07	0.04	0.01	0.23	0.12	0.04	0.08	0.04	0.02	0.045	0.29	0.10
Methane	0.11	0.05		0.36	0.18		0.17	0.45		0.57	0.18	
2-Butyne	0.19	0.08		0.41	0.17		0.34	0.11		0.67	0.21	

Table 3.2: Interaction energies for Ne/HF complex using different methods in aug-cc-pVTZ basis set. Energies are given in cal/mol.

	CP-optimized			No-CP		
	ΔE_{int}	Ne--F/H	BSSE	ΔE_{int}	Ne--F/H	Ratio $\mu\mu$
F-facing						
CCSD(T)	− 102	3.10	73	− 178	3.02	1.0
MP2	− 80	3.17	61	− 145	3.08	1.0
RHF	−2	4.20	7	− 11	3.75	0.2
B2PLYP	− 20	3.12	31	− 51	3.08	0.3
X3LYP	− 79	3.03	18	− 97	3.01	0.7
B3LYP	0			0		
M05	− 208	2.87	31	− 239	2.86	1.0
M05-2X	− 200	2.88	29	− 229	2.87	1.3
M06	− 160	3.48	34	− 183	3.47	0.8
M06-2X	− 215	3.09	20	− 235	3.08	1.6
M06L	− 196	3.09	34	− 230	3.09	2.4
B97D	− 174	3.32	16	− 190	3.29	0.9
B2PLYP-D	− 150	2.92	39	− 190	2.89	0.7
H-facing						
CCSD(T)	− 212	2.40	253	− 506	2.23	
MP2	− 168	2.47	210	− 422	2.28	
RHF	− 31	2.94	28	− 70	2.66	
B2PLYP	− 169	2.34	122	− 299	2.26	
X3LYP	− 295	2.25	71	− 368	2.22	
B3LYP	− 98	2.37	59	− 160	2.31	
M05	− 415	2.27	96	− 515	2.22	
M05-2X	− 306	2.28	107	− 418	2.23	
M06	− 176	2.95	31	− 207	2.94	
M06-2X	− 251	2.50	65	− 327	2.34	
M06L	− 123	2.71	65	− 193	2.56	
B97D	− 454	2.46	63	− 519	2.43	
B2PLYP-D	− 412	2.31	127	− 543	2.27	

As seen from the table 3.1, the stability of the F-facing complexes (both HF and LiF) increases with a size (and a polarizability) of the non-polar molecule. However, if the polar molecule faces the non-polar entity with its less electronegative atom (H or Li), the order of stability changes to Ne < 2-butyne < methane which indicates that other effects like a charge transfer (CT) influence the interaction. What is more important, the interaction energies converge much faster than BSSE. For example, the increase of the basis set from aug-cc-pVTZ to aug-cc-pV5Z for the Li-facing Ne-LiF complex decreases the interaction by just 0.06 kcal/mol, while the BSSE decreases from 0.45 kcal/mol to 0.10 kcal/mol. Therefore, the aug-cc-pVTZ basis set seems to work reasonable well when used on the CP-corrected surface. If one considers the percent change of the interaction upon the rise of the basis set, the slowest convergence is observed for neon, probably due to its low polarizability which requires more flexible basis sets to fully converge.

The analyzed inductive interaction is accompanied by the dispersion and the charge transfer. The table 3.4 presents calculated induced dipoles and the charge transfers between entities. The induced dipoles increase in the order Ne < methane < 2-butyne and are greater for LiF than for HF, as expected. The MP2 dipoles are mostly smaller than those calculated by DFT. When F faces the non-polar molecule the smaller induced dipoles are observed which correlate with the longer equilibrium distance. To calculate the charge transfer, the electron density was first projected to the minimal basis set and then divided using the Mulliken analysis. The projection is necessary since the Mulliken analysis equally divides the electron density between the adjacent atoms and some artefactual charge transfer due to the presence of diffuse functions is observed. Projection to the minimal STO-3G basis set is a common way of handling the problem[61]. The calculated charge transfer was corrected for the unphysical transfer to ghost orbitals using the CP procedure. All calculated transfers are negative indicating the electron density shifts from the neutral to the polar molecule. The charge transfer for the H and Li facing complexes exceeds that for F-facing and those predicted for methane exceed those for 2-butyne, which correlates with the results in table 3.1 and support hypothesis of the underlying charge transfer augmenting the inductive interaction.

In the table 3.2 I gathered results for Ne-HF complex calculated using various DFT and ab initio methods on the CP-corrected surface. CCSD(T)/aug-cc-pVTZ predicts interaction energies of -102 and -212 cal/mol, respectively, for F and H facing complexes. In general, all DF's overestimate the energy of interaction by from 50% (B2PLYP-D) to 100% (M06-2X), while MP2 and X3LYP

Table 3.3: Interaction energies (kcal/mol) of HF and LiF with Ne, methane and 2-butyne for structures optimized on CP-corrected surface using aug-cc-pVTZ basis set for all methods. In case of MP2 and CCSD(T), aug-cc-pV5Z and aug-cc-pVQZ basis sets were used for Neon, and Methane and 2-Butyne, respectively.

	F-facing			H or Li-facing		
	Neon	Methane	2-Butyne	Neon	Methane	2-Butyne
			HF			
B3LYP	0	0	− 0.19	− 0.76	− 1.00	0
RHF	0	0	− 0.28	− 0.48	− 0.49	0
X3LYP	− 0.08	0	− 0.39	− 1.01	− 1.33	− 0.10
M06-2X	− 0.22	− 0.24	− 0.73	− 0.84	− 1.99	− 0.58
B97D	− 0.17	− 0.26	− 0.68	− 2.54	− 2.20	− 1.05
MP2	− 0.09	− 0.19	− 0.66b	− 0.64a	− 1.50b	− 0.38b
CCSD(T)	− 0.13	− 0.27	− 0.73b	− 0.68a	− 1.62b	− 0.45b
			LiF			
B3LYP	− 0.02	0	− 1.28	− 0.10	− 3.81	− 1.33
RHF	− 0.03	0	− 1.34	− 0.03	− 3.08	− 0.49
X3LYP	− 0.16	− 0.09	− 1.58	− 0.30	− 4.24	− 1.71
M06-2X	− 0.28	− 0.41	− 2.33	− 0.25	− 4.32	− 1.69
B97D	− 0.21	− 0.32	− 1.78	− 0.45	− 8.64	− 6.62
MP2	− 0.13	− 0.29	− 1.92b	− 0.18a	− 3.64b	− 1.20b
CCSD(T)	− 0.18	− 0.45	− 2.15b	− 0.23a	− 3.76b	− 1.24b

% error compared to CCSD(T)

	F-facing		H or Li-facing		Overall	
	MSE	MUE	MSE	MUE	MSE	MUE
B3LYP	40	40	17	24	29	32
RHF	38	38	57	57	47	47
X3LYP	27	27	− 11	34	8	30
M06-2X	−5	7	− 28	28	− 17	18
B97D	15	18	− 225	225	− 105	121
MP2	11	10	7	7	9	9

Table 3.4: Induced dipole moments (in Debyes) and Charge tranfer (in electrons x 10^3) for different complexes. Please see details in the text.

		Induced Dipole moment:				Charge tranfer in electrons x 10^3			
	X	HF		LiF		HF		LiF	
Facing:		F	H	F	Li	F	H	F	Li
Neon									
RHF	T	0.01	0.04	0.04	0.19		− 0.0	− 0.0	− 8.0
X3LYP	T	0.03	0.12	0.08	0.26	− 0.0	− 2.7	− 0.2	− 12.8
B3LYP	T		0.11	0.06	0.25		− 2.0	− 0.1	− 12.2
M06-2X	T	0.02	0.08	0.07	0.22	− 0.0	− 0.6	− 0.1	− 10.1
B97D	T	0.02	0.10	0.06	0.26	− 0.0	− 1.7	− 0.1	− 12.1
MP2/CCSD(T)	T	0.02	0.08	0.06	0.22	− 0.0	− 0.3	− 0.0	− 7.7
MP2/CCSD(T)	Q	0.02	0.08	0.07	0.22	− 0.0	− 0.3	− 0.0	− 8.0
MP2/CCSD(T)	5	0.02	0.08	0.07	0.22	− 0.0	− 0.3	− 0.0	− 8.0
Methane									
RHF	T		0.29	0.21	0.87		− 1.9	−0.0	− 38.5
X3LYP	T	0.09	0.50	0.34	0.95	− 0.1	− 13.3	− 1.3	− 51.9
B3LYP	T		0.45		0.95		− 10.6		− 51.5
M06-2X	T	0.10	0.54	0.41	0.94	− 0.3	− 14.0	− 3.0	− 49.4
B97D	T	0.10	0.51	0.33	0.94	− 0.2	− 15.5	− 1.0	− 49.7
MP2/CCSD(T)	T	0.11	0.46	0.36	0.87	− 0.4	− 6.4	− 1.3	− 39.5
MP2/CCSD(T)	Q	0.11	0.46	0.36	0.88	− 0.4	− 6.4	− 1.3	− 39.4
2-Butyne									
RHF	T	0.17		0.78	1.40	− 0.1		− 1.9	− 34.9
X3LYP	T	0.26	0.62	0.89	1.68	− 0.6	− 6.5	− 4.1	− 52.8
B3LYP	T	0.23	0.56	0.87	1.64	− 0.2	− 5.0	− 3.5	− 50.0
M06-2X	T	0.29	0.73	1.00	1.61	− 1.3	− 9.5	− 6.6	− 46.9
B97D	T	0.27	0.66	0.91	1.84	− 0.6	− 8.5	− 3.5	− 55.6
MP2/CCSD(T)	T	0.27	0.57	0.89	1.51	− 0.8	− 3.1	− 3.8	− 37.4
MP2/CCSD(T)	Q	0.27	0.57	0.89	1.51	− 0.8	− 3.1	− 3.8	− 37.3

slightly underestimate by 20%. The B2PLYP predicts the interaction to be very weak and B3LYP does not find any stable minimum on the PES. RHF converges to a very shallow minimum of -2 cal/mol at 4.2Å. When the F faces the noble gas, the induction is augmented by the charge transfer (see table 3.4) which result in stronger attraction. Despite, five of seven DF's overestimate the interaction by from 25% (M06-2X) to 100% (M05, B2PLYP-D, and B97D) and other two, M06L and M06, underestimate it respectively by 89 and 36 cal/mol. All other methods, except X3LYP which predicts the interaction to be too strong by 83 cal/mol, underestimate it. MP2 and B2PLYP perform reasonably well predicting the energies to be 168 and 169 cal/mol, but for B3LYP the attraction becomes too weak, while RHF hardly finds any.

In table 3.3 I examine some more popular functionals in somewhat less details. The analysis of mean signed errors (MSE's) and mean unsigned errors (MUE's) shows that X3LYP provides the lowest MSE with respect to the CCSD(T) energies among all considered functionals. M06-2X follows X3LYP with slightly higher MSE caused by consistent overstabilization of Li or H facing complexes, another DF - B97D – drastically overestimates mentioned interaction placing it at the last position, behind B3LYP and even RHF. Interestingly, MP2 consistently underestimates the interaction by 10% which does not follow reports that MP2 in general overestimates pure dispersion complexes [62].

Discussion

When compared to CCSD(T) calculations, the functionals designed to treat dispersion interactions behave rather erratically, but with a tendency to overestimate the strength of these interactions for most of the cases that are studied in this work. The functionals do not follow the expected trends with respect to increasing size of basis sets and dispersion interactions and BSSE. In several cases, the distances between the interacting entities increased when the interaction energies became stronger (as the basis set was varied). All of these observations suggest that these functionals do not provide reasonable physical models for the non-bonded interactions that they aim to assess. Erratic results involving apparent anomalous multiple minima on potential energy surfaces can arise with some of these functionals with normal and even relatively fine integrations grids. In the present work, we found this problem with B2PLYP, M06, M06L, and B2PLYP-D. Even with the finest grid available in Gaussian 09 (99974) the anomalies did not completely disappear. Sherrill

and co-authors have reported the ability of several of these (and other) functionals to reproduce the values of many interactions in the datasets which were used to parametrize many of them [20].

The high level calculations on the systems where induction occurs tend to support the idea that induction enhances dispersion since the induced permanent dipole will more likely be increased than diminished by the time dependent electron correlated motions that lead to dispersion. The B3LYP and X3LYP functionals can describe the interactions in several of these systems as well or better than some of the functionals specifically designed to address dispersion.

I suggest using better data sets that are based on good experimental data, and using fewer parameters that are chosen with preservation of the physical model in mind when designing better functionals.

3.4 Reinvestigation of the p-Benzoquinone and Pyrimidine Dimer

Introduction

Interactions due to the dispersion and induction might play an important role in the energetic of biological molecules such as DNA and peptides. For example, Phenylalanine and Tyrosine are commonly found in amyloid β-sheets which suggest their possible stacking (dispersive) interaction might be a part of problematic stability of these aggregates (work has not been published yet). Even in a wider view, the dispersion might provide a significant stabilization in the protein structure, interlocking side chains in rigid minima inside the protein's hydrophobic core [63]. Recently, the stacking interactions in another important class of biological molecules - DNA and RNA - have received considerable attention. While each base is hydrogen bonded to its complementary partner, the stacking interactions with an adjacent pair might provide another source of the stabilization. Moreover, these close contacts are the source of unusual for biological molecules properties like possibility of the charge transfer along the DNA chain for long distances [64]. This process is thought to have the critical meaning in the localization and repair of mismatched DNA pairs. Therefore, it is important to answer the question what is the strength of π-stacking interactions, especially when compared to other forces.

In the experiment, like the protein crystallography, one might observe the proximity of stacking rings. However, this observation does not tell us much about the reasons these rings were brought

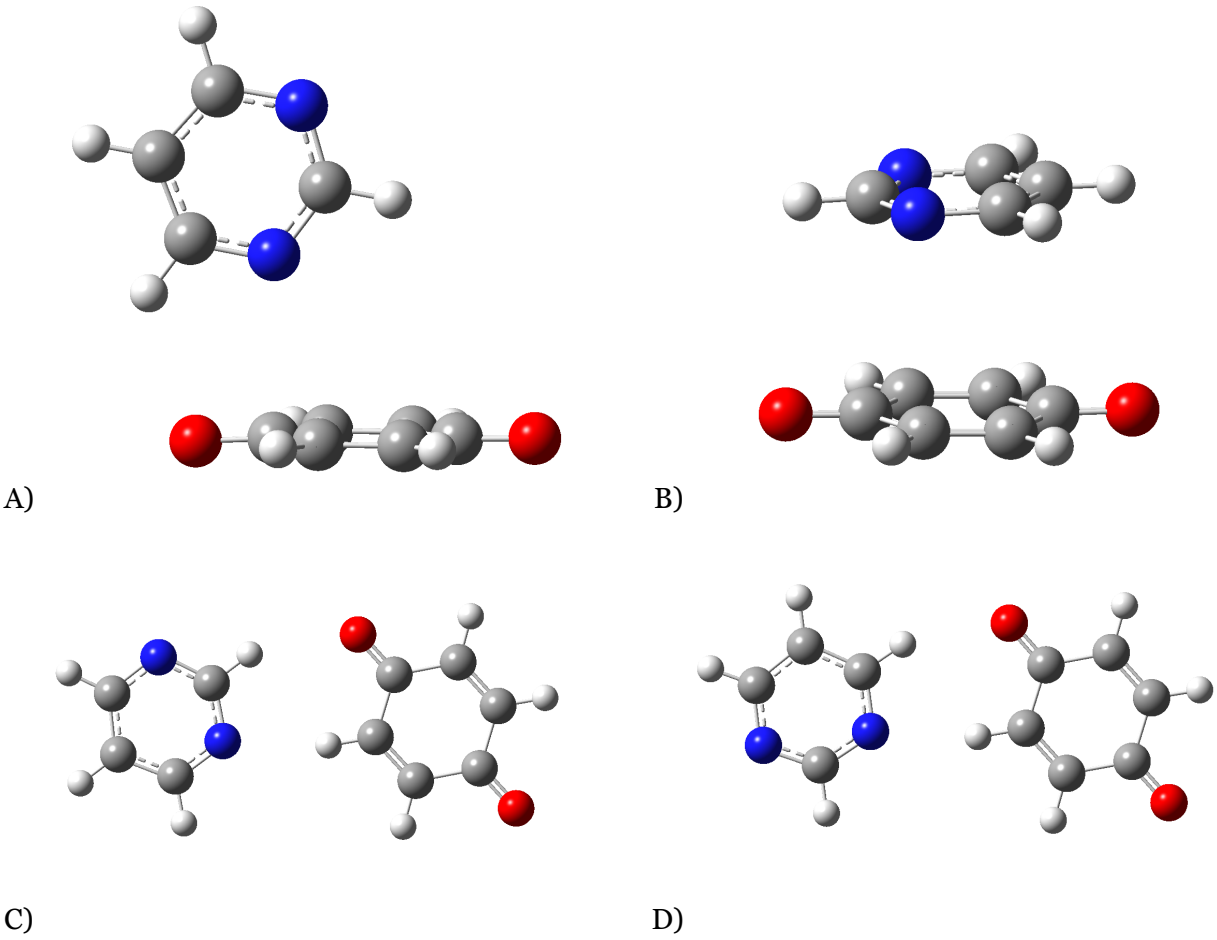


Figure 3.3: Structures of A) T-shape, B) Stacking, C) planar A and D) planar B dimers

together. Indeed, they might occur to be in a close contact due to the attractive dispersion interaction, but at the same moment other conformations could destabilize the system. Theoretical methods help to "number" these interactions, introducing some mathematical order into dominion of biochemical chaos. When interactions are "numbered", one might feel more confident and can qualitatively (or even quantitatively) calculate their influence. However, at first it is important to know, how good the theoretical methods are. In other words we have to know if the results are trustworthy, because the wrong answer might make more harm than the lack of the answer; and here, one has to turn back into the experiment again.

In this section I compare predictions of two classes of newly developed DF's and TF's with high level ab initio CCSD(T) calculations on the complex formed between pyrimidine and p-benzoquinone, for which the detailed experiment was reported [65]. I also identified new T-shape structure which was added to previously found π-stacking and two planar structures (see figure

3.3). The experiment, with support by ab initio calculations, identified the hydrogen bonded complex to be the only conformer found in the argon matrix at 12K; now I want to examine various DFT methods if their predictions will agree with the experimental findings.

Results

The approach which is described in the methodology part of the current chapter needed two modifications: (1) the enthalpies were evaluated at 12K which was the temperature of the argon matrix, and (2) the MP2/cc-pVTZ vibrational corrections were used for calculating the CCSD(T) enthalpies.

According to Adamowicz's report and my predictions, I was not able to find minimum on PES for the π-stacking dimer using B3LYP, X3LYP and PBE1PBE functionals (collectively called TF's). Multiple optimization attempts consistenlty leaded to the T-shaped complex which contains bent C-H --- O hydrogen bond and possible electrostatic (or dispersive in case of DF's) interaction between the nitrogen in pyrimidine and the positive carbon in p-benzoquinone's carbonyl group (or the electron cloud of the π-system). The results for all methods (except B97D) do not differ between d95++** and cc-pVTZ basis sets and I will analyze them collectively.

As seen from the table 3.5 there is no agreement between different methods upon the stability of the different conformers and only one apparent trend can be found: all methods predict the planar A structure to be less stable than other hydrogen bonded conformer, planar B, and therefore I will remove the planar A dimer from the further consideration. The reason might be a shorter distance between the non-bonding nitrogen and the oxygen or lower acidity of the hydrogen involved in the hydrogen bond. However this subject is out of the scope of my study.

At this point few observations about the most stable structures predicted by DF's might be done:

1. B97D, ωB97x-D and M06-2X functionals find the π-stacking complex to be the most stable.
2. The T-shaped dimer has the lowest ΔE (and ΔH) for M06 and M06L functionals.
3. M05 and M05-2X methods predict the planar B structure to have lowest ΔE (and ΔH).

Moreover, if one disregards the planar A structure, the DF's do not agree about the least stable structure either:

Table 3.5: Energies and enthalpies of interaction for presented dimers in d95++** and cc-pVTZ basis sets

	Interaction Energies (CP-Opt)				Interaction Enthalpies (CP-Opt) at 12 K			
Functional:	basis set: d95++**							
	T-shape	stacking	planar A	planar B	T-shape	stacking	planar A	planar B
B3LYP	-1.97		-2.69	-3.49	-1.61		-2.22	-2.95
X3LYP	-2.48		-3.15	-3.97	-2.09		-2.66	-3.41
PBE1PBE	-2.65		-3.31	-4.14	-2.27		-2.84	-3.57
B97D	-4.87	-5.06	-4.20	-5.00	-4.37	-4.57	-3.69	-4.42
ωB97xd	-4.86	-5.36	-4.34	-5.18	-4.37	-4.91	-3.83	-4.63
M05	-3.51	-2.08	-3.57	-4.41	-3.10	-1.79	-3.08	-3.83
M05-2X	-4.52	-3.85	-4.00	-4.87	-3.97	-3.51	-3.53	-4.32
M06	-4.39	-3.52	-3.47	-4.35	-4.01	-3.15	-2.98	-3.80
M06-2X	-4.84	-5.00	-3.71	-4.54	-4.33	-4.55	-3.29	-4.04
M06L	-4.81	-4.35	-3.41	-4.25	-4.34	-3.94	-3.00	-3.73
MP2	-4.97	-5.38	-4.04	-4.80	-3.89	-4.42	-3.06	-4.34*
	basis set: cc-pVTZ							
	T-shape	stacking	planar A	planar B	T-shape	stacking	planar A	planar B
B3LYP	-1.73		-2.36	-3.09	-1.40		-1.93	-2.56
X3LYP	-2.23		-2.80	-3.56	-1.87		-2.36	-3.04
PBE1PBE	-2.45		-3.01	-3.76	-2.10		-2.58	-3.25
B97D	-4.75	-5.06	-3.92	-4.67	-4.27	-4.58	-3.42	-4.12
ωB97xd	-4.80	-5.46	-4.15	-4.96	-4.35	-4.95	-3.63	-4.41
M05	-3.25	-1.92	-3.36	-4.15	-2.90	-1.64	-2.90	-3.58
M05-2X	-4.41	-3.85	-3.66	-4.44	-3.95	-3.48	-3.21	-4.18
M06	-4.54	-3.78	-3.47	-4.32	-4.10	-3.04	-2.99	-3.77
M06-2X	-4.80	-5.16	-3.46	-4.26	-4.36*	-4.73	-3.07	-3.69
M06L	-4.84	-4.42	-3.16	-3.95	-4.40	-4.00	-2.74	-3.44
MP2	-4.92	-6.23	-3.77	-4.54	-4.56	-5.90	-3.36	-4.06
CCSD(T)[a]	-3.88	-3.08		-4.52	-3.52	-2.74		-4.04

[a] - used with the MP2 enthalpic corrections

* - 1 imaginary frequency

1. M06, M06L, and B97D (with cc-pVTZ basis set) find the planar B isomer to be the least stable.
2. The π-stacking complex is found to have the highest ΔE (and ΔH) for M05, M05-2X and M06 functionals
3. B97D (with d95++** basis set) and ωB97x-D predict the T-shaped structure to be the least stable.

On the other hand, all TF's predict the planar B structure to be more stable than T-shape and do not find the stable minimum for the π-stacking complex. Moreover, MP2 reverses this order, predicting the π-stacking structure to be the most stable, followed by the T-shaped and the hydrogen bonded planar B being the last one.

Clearly these results miss consistency, and what is more, most of the methods including MP2, do not agree with the experiment. Therefore, I decided to (1) recalculate energies using CCSD(T), and (2) fit the calculated shifts in frequencies to those reported in the experiment. From the analysis of the CCSD(T) results, included in the table 3.5, one can see that the planar B structure becomes the most stable, followed by the T-shape and π-stacking conformations. These predictions stand in the agreement with the findings by Adamowicz's group.

I chose to use seven frequency shifts for further analysis. The relative shifts in the frequencies upon formation of various complexes are reported in the table 3.6, while detailed analysis for MP2/cc-pVTZ shifts is presented in table 3.7. The p-benzoquinone contains one imaginary frequency at MP2/d95++** level of theory due to use of diffuse functions with double zeta basis sets, as it was observed for other arenes [66], and it is excluded from the further analysis.

MP2 produces the lowest mean unsigned error (MUE) for the planar B structure. The detailed examination shows that the stacking structure predicts shifts in the wrong direction for four of the seven frequencies. The largest discrepancies appear for the out-of-plane bending mode of p-benzoquinone at 886 cm^{-1} (-5.9 cm^{-1} instead of +6 cm^{-1}) and of pyrimidine at 719 cm^{-1} (0.5 cm^{-1} instead of +3 cm^{-1}), while the planar B conformer's shifts correlate with experimental results (+6.7^{-1} and 4.0 cm^{-1} respectively). The T-shape structure experiences similar mismatch at 886 cm^{-1}, being redshifted by 6.3 cm^{-1}.

The table 3.6 presents the MUE's for the assignment of the seven reported shifts considered here. All the functionals have the lowest MUE for the planar B structure, despite their inconsistent

Table 3.6: Band shifts calculated by various functionals (in cm^{-1}) compared to those obtained from the experiment [65]

Basis set:	d95++**			cc-pVTZ		
Functionals:	T-shape	Stacking	planar B	T-shape	Stacking	planar B
MP2	2.7	3.5		3.5	4.4	1.0
B3LYP	2.2		1.4	2.7		1.6
X3LYP	2.2		1.6	2.6		1.6
PBE1PBE	2.5		1.5	2.9		1.4
B97D	2.8	2.9	1.6	2.7	3.0	1.8
ωB97x-D	2.4	2.9	1.8	2.7	3.0	2.1
M05	2.7	2.7	1.6	2.7	2.8	1.9
M05-2X	2.7	3.1	1.6	3.0	4.1	1.9
M06	3.6	3.9	1.7	3.5	3.5	2.2
M06-2X	2.9	3.2	1.5	3.3	3.7	2.7
M06L	3.9	3.8	1.7	4.3	3.5	2.1

Table 3.7: MP2/cc-pVTZ calculated unscaled frequency shifts (in cm^{-1}) compared to those obtained from the experiment

		T-shaped	Stacking	planar B	exp[65]
1400	p	1,4	-3.1	7.7	5.0
1301	q	3.7	7.3	4.6	4.0
1223	p	0.7	-0.5	4.3	4.0
886	q	-6.3	-5.9	6.7	6.0
719	p	0.6	0.5	4.0	3.0
621	p	2.4	-1.2	1.9	1.0
407	q	2.9	0.5	2.6	2.0

predictions about the lowest enthalpy conformer. The functionals have larger MUE than MP2 calculations which in fact stands in perfect agreement (MUE of 1.0 cm^{-1}) with the experiment. In addition, in order to achieve these results for DFT methods I needed to use the finest '99974' grid available in Gaussian, otherwise the calculated MUE's were much larger.

Discussion

At this point one might suspect that there might be something wrong with the experiment. However, the methodology describes that it was performed very carefully [65] and the fact that all methods provide the best fit of the frequencies' shifts with the planar B structure also supports the hydrogen bonded structure. Therefore, the problem lies in the DFT methods, few of which are not able to correctly estimate the energy of the most stable conformer. Moreover, the MP2 and CCSD(T) ab initio calculations do not shed any light on either, and the former method finds π-stacking structure to be the most stable. Are we able to precise where the problem appears to be?

The difficulty lies in the proper evaluation of the relative stability of the hydrogen bonded (mainly electrostatic, but with a significant covalent character [67]) and π-stacking/T-shaped (mainly dispersion-driven) conformers. It has been shown that TF's behaves extremely well for hydrogen bonded complexes [56] and proteins secondary structures [57], [68] Likewise, from analyzed data one can see that DF's stands in a good agreement with the CCSD(T) calculations for the hydrogen bonded compound. Therefore the real problem lies specifically in the estimation of the dispersive interactions' strength.

In case of MP2, it was previously observed that it can drastically overestimate dispersion in the π-stacking complexes; however it does not appear to be true for all dispersion stabilized aggregates, like noble gas dimers studied in the previous section. My results support these findings, since the enthalpy of hydrogen bonded dimer at MP2/cc-pVTZ level of theory differs just by 0.02 kcal/mol from values predicted by CCSD(T)/cc-pVTZ, while the π-stacking dimer is overestimated by more than 100%.

The foregoing questions the quality of methods used in calculations on biochemical systems. They rather base on MP2 level of theory extrapolated to Complete Basis Set (CBS), since CCSD(T) calculation are prohibitively expensive for the biochemical systems [2]. Even when CCSD(T)

corrections are included, they use double zeta basis sets and the accuracy of which was recently impaired by the work of Scherill et al [20].

The presented results undermine the choice of dispersion-corrected functionals as a cheap and effective alternative to ab initio methods which is able to assess influence of the dispersion. As one can see, the relative influence of the dispersion (π-stacking) and the electrostatic (hydrogen bonding) is imbalanced, predicting artefactual structures based on erroneous emphasis on the dispersion in aromatic compounds.

In the next section I will extend my study to systems more directly connected to the biochemistry. I will investigate the stability and the geometries of various secondary motives using TF's and DF's to examine how the dispersion influences complexes which potential energy (hyper)-surfaces' are more complicated than those of p-benzoquinone and pyrimidine dimers.

3.5 Comparison of some Dispersion-corrected and Traditional Functionals Applied to Peptides

Introduction

We know that traditional functionals (like studied here B3LYP, X3LYP and PBE1PBE) are not able to describe pure dispersion and the newer generation of functionals have been designed to assess this problem. The temptation to use these to biochemical calculations is obvious. The parameterization against big training sets (like M06 – series of functional having up to 39 different parameters) and small average errors for test sets provide a solid guarantee for more accurate results. However, in section 3.3 I argued that the dispersion becomes less significant when coupled with induction and the secondary structures which are supported by the network of hydrogen bonds might be the induction-dominated structure.

In this section, I compare several functionals designed to treat dispersion with the older generation of hybrid General Gradients Approximations (GGA's) functionals used for determination of the α-helical and the β-sheet peptide secondary structures of the capped polyalanines and available experimental results.

I use methods described in the methodology section of this chapter. In addition, α-helices were additionally corrected for the internal BSSE by cutting out individual chains of the hydrogen

bonded formamides and capping them with hydrogen atoms. These three chains of six formamides each were corrected for the BSSE using a posteriori method and the computed BSSE was used as a CP-correction for the predicted enthalpies of folding.

α-helix

First I compare the ΔH/residue for folding from the Fully Extended Strand (FES) to the α-helix (see table 3.8). The 17 alanines α-helix was chosen to limit terminal effects and achieve the necessary length required to obtain the stable α-helix strucure [57]. The FES which contain all Ramachandran dihedrals ϕ and ψ close to 180^o was chosen for the model of unfolded state.

The calculated ΔH's/residue fall into two ranges: all TF's predict ΔH's to be about -1 kcal/mole (-0.8, -1.2 and -1.3 kcal/mole for B3LYP, X3LYP, and PBE1PBE, respectively), while DF's are more stabilizing and the estimated ΔH/residue to be about -4 kcal/mol (-3.9, -4.2, -3.9 and -4.3 for M06, M06-2X, B97D, and ωB97x-D, respectively). Various experiments conducted to measure ΔH (collected in table 3.8) estimated experimental enthalpy in the range of -0.9 – -1.2 kcal/mol which agrees with the predictions by TF's. The influence of the solvent on difference between pp2-like and α-helix solubility was found to be quite small [69], therefore it cannot alter the presented results.

It is interesting to note that despite different peptides were used in the reported experiments all results fall into very small range of energies. This suggests that the unusual stability of an α-helix is mainly due to hydrogen bond network and while some amino acids itself might enhance helix formation (results has not been published yet), the side chains interactions might not be as important (except some rare cases like formation of extra hydrogen bonding chains [76]) as generally assumed.

In table 3.9 I gather structural parameters for the optimized helices and extracted from the crystallographic x-ray structures. I discarded 6 terminal and averaged 9 internal hydrogen bonds to include the environment effects, since the calculations are performed in the gas phase, but in the proteins helices might interact with the surrounding which will affect helix's geometry as I show in case of interaction with water [79] in section 4.4. For the comparison I used helices containing 15 hydrogen bonds. It should be mentioned that since the positions of hydrogens are not directly observed in the x-ray experiment and they are added later using some structural parameters, my

Table 3.8: ΔH/residue for the folding of the 17-alanine α-helix from extended strand.

Method:	ΔH/residue:
B3lYP	-0.8
X3LYP	-1.2
PBE1PBE	-1.3
M06	-3.9
M06-2X	-4.2
B97D	-3.9
ωB97x-D	-4.3
Other theoretical:	
ONIOM/B3LYP:AM1	-0.7[58]
Thermodynamic cycle	-0.8[70]
Experimental:	
Kemp 2006	-1.0[71]
Baldwin 1991	-1.1[72]
Makhatadze 2005	-0.9[73]
Rialdi 1996	-1.1[74]
Scheraga 1971	-1.2[75]

Table 3.9: Geometric parameters for the optimized structures of the 17-alanine α-helix

	Distances(Å)			Angles(^{o})		
Method:	H-O	N-O	CH-O	C=O---H	O---H-N	C=O --- N
X3LYP	2.01	3.01	2.86	151.2	164.5	155.8
B3LYP	2.03	3.03	2.90	151.0	164.6	155.5
PBE1PBE	2.03	3.01	2.84	149.3	163.1	154.3
M06	1.97	2.96	2.62	151.2	162.1	156.5
M06-2X	1.98	2.95	2.61	151.2	159.8	157.7
B97D	2.04	3.00	2.59	146.8	155.7	154.0
ωB97x-D	1.94	2.92	2.64	149.5	161.4	155.3
Experiment [77]	2.06	2.99±0.01		149±7	157±9	
Experiment [78]		2.97±0.01				153±2

confidence in the parameters involving hydrogen atoms is lowered.

Due to the network of the hydrogen bonds, the α-helix is a very rigid structure and only small derivations in the geometries bring no surprise. Only the N-O distance for ωB97x-D is somehow shorter by 0.05Å and C=O – N angle for M06-2X is larger by 5.7^o than the experimentally determined values. Remaining parameters are within or very close to the experimental range. Also, I used the full alanine helix while the data extracted from the databases contain all range of different helices; therefore some small deviations are expected. The only big difference between TF's and DF's is observed in the C-H – O distances for which I was not able to find any experimental values. DF's tend to tilt methyl side chains to form C-H – O bond, while TF's do not pick this interaction. Unfortunately, I am not aware of any experimental data that allows me to comment on this matter in more detail.

One might argue that erroneous ΔH's for DF's might be an effect of the double zeta basis sets which are not the best suited for dispersion, as DF's are meant to work with "properly polarized triple zeta basis set" [80]. However, if one increases the basis set, the more exhaustively spanned vector space will make the overlap of basis functions more efficient which will increase the strength of dispersive interactions, making the α-helix even more stable. Moreover, the DFT is superior to ab initio wavefunction methods due to its great performance in the moderate basis sets (vide section 3.3 and [81]), while the wavefunction methods require expensive triple and quadruple zeta basis sets to reassure convergence of the electron correlation.

β-sheet

In the next part I compare the performance of TF's and DF's (pure and in combination with ONIOM method) in the calculations on the antiparallel β-sheets consisting of four strands of $Ac(Ala)_6NHMe$. For the calculation details please look in the methodology section. The energies are gathered in table 3.10 and the structural data are presented in tables 3.11 and 3.12. Figure 3.5 shows a side view on the sheets optimized with different methods, and figure 3.4 is a legend to the table 3.11.

The energies presented in the table 3.10 fall into two different ranges, somehow similar to those observed for the α-helix. The DF's predicts β-sheets to be more stable than TF's, estimating the CP-corrected energy of the inter-strand interaction about 46-52 kcal/mole, 50% more than

30-34 kcal/mol calculated for TF's. These data reflects the finding for the α-helices and, while we lack the direct experimental data for the enthalpies of the antiparallel β-sheets formation, I can conclude that due to the similarity of the systems energies predicted by DFs are too stabilizing.

Moreover, let me note the difference between calculations performed with the ONIOM and pure DFT functionals. DF's energies are significantly lower (by 5 – 9 kcal/mol) than those obtained with the ONIOM, while the difference for TF's is quite small. Since I use the semi-empirical AM1 method in the ONIOM just for methyl side chains, I attribute the difference to the methyl – methyl attraction between two adjacent strands which gets emphasized in DF's calculations. Figure 3.5 shows how the methyls tilt toward and away from each other in order to increase the dispersive interactions. I will analyze the structures in more detail in the following paragraphs.

The optimized β-sheets fall into three distinct categories presented in figure 3.5. The structures calculated using pure TF's converge to typical textbook planar pleated sheets. When the methyl side chains are described using the AM1 in the ONIOM approach the predicted structures remain pleated, but instead of being planar, a significant twist (25 degrees, see row "Dihedral" in table 3.11) is observed. Finally, both pure DF's and when used with ONIOM (except B97D:AM1 combination) predict the structures more flat than the standard β-sheets. The exemption, B97D:AM1, behaves like other TF's:AM1 methods.

I used the two innermost strands to measure Ramachandran ϕ and ψ dihedrals for different methods (see table 3.12) and compare them with some experimental data. Three reports were used for comparison:

1. Creighton in his book [82] recommends using values of ϕ and ψ for β-sheets to be -139^o and 135^o respectively.
2. NMR study of a correlation between chemical shifts and dihedrals in high quality crystal structures by Bax et al [83] found substantial variation for ϕ and ψ, reporting -114$\pm$32 and 142$\pm$20 for ϕ and ψ angles respectively.
3. Meier et al [84] work on spider silk protein (which is composed of large antiparallel β-sheets consisting mainly alanines and glycines) using NMR technique finds ϕ and ψ angles about alanine residues to be -135^o and 150^o respectively.

TF's predict the dihedral ϕ and ψ angles to be in the good agreement with the experiment.

Table 3.10: Interaction energy and distortion energy per interaction in 4A6 AP β-sheet using different functionals with and without ONIOM

Method:	b3lyp	x3lyp	PBE1PBE	M06	M06-2X	B97D	ωb97-D
			ONIOM				
ΔE(opt)	-41.5	-45.0	-43.0	-50.0	-53.7	-55.2	-54.2
ΔE(frozen)	-54.8	-58.5	-56.3	-57.1	-60.1	-63.4	-61.9
ΔE(CP SP)	-30.1	-33.2	-33.8	-40.2	-44.9	-44.2	-44.8
ΔE(CP SP frozen)	-43.6	-46.7	-47.1	-47.3	-51.4	-52.3	-52.5
Distortion	13.3	13.4	13.3	7.1	6.3	8.2	7.7
			Full DFT				
ΔE(opt)	-42.7	-46.3	-44.2	-57.1	-58.9	-61.2	-62.9
ΔE(frozen)	-53.2	-56.8	-54.3	-64.2	-65.1	-66.6	-70.2
ΔE(CP SP)	-30.1	-33.2	-34.0	-46.1	-49.0	-48.3	-52.4
ΔE(CP SP frozen)	-40.5	-43.7	-44.0	-53.2	-55.3	-53.7	-59.7
Distortion	10.4	10.5	10.0	7.1	6.3	5.4	7.3

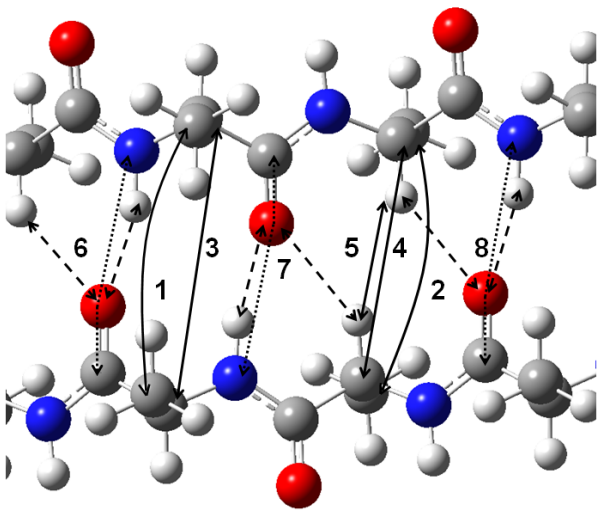


Figure 3.4: Close up view of the center of the β-sheet (the middle of the interaction between the 2nd and 3rd strands). The numbers correspond to distances listed in table 3.11

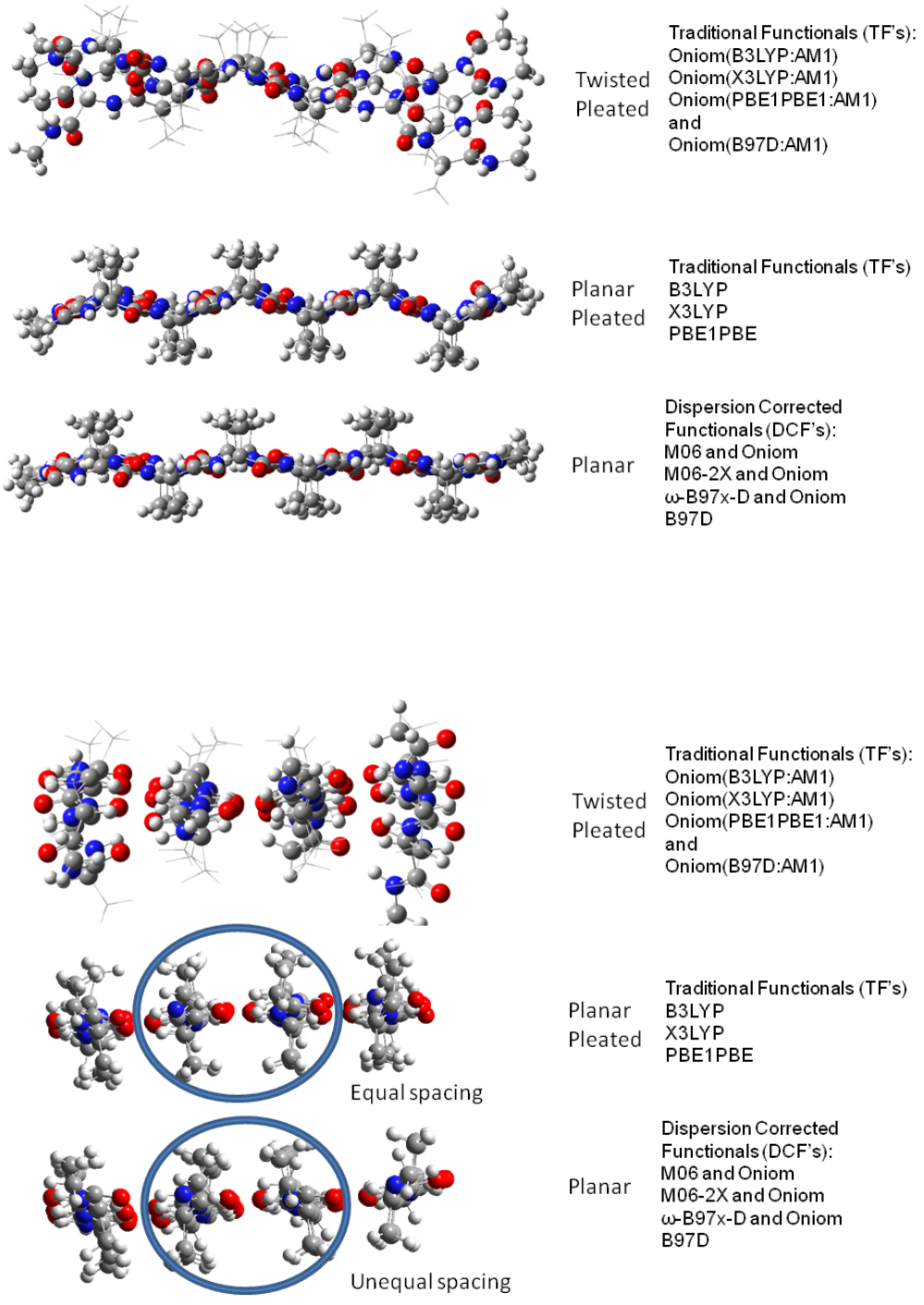


Figure 3.5: Three types of geometries for AP β-sheet classified by the calculation that give raise to them. Top - view from end of sheets; bottom - view from edge of sheet

While it is not clear if the twist in the β-sheets is an intrinsic property of the sheet or the effect of the interaction with the environment (for example interactions with water – please see section 4.5), the twisted TF's with ONIOM (and B97D:AM1) structures are within the experimental range too. However, the dihedrals predicted by DF's are significantly larger and, while ψ angles are on the high limit, the ϕ angles are out the experimental range reported by Bax.

Not only dihedral angles suggest that DF's predict uncommon structures. Other apparent feature is the behavior of alanine's methyl groups. As seen from figure 3.5, methyls in the adjacent stands tilt toward and away from each other in a manner that optimizes contacts (hence increase the dispersion interaction) between them. Figure 3.4 and table 3.11 presents more details about this and other features. The antiparallel β-sheets contain cyclic hydrogen bonds composed from 14 (large ring) and 10 (small ring) atoms (see the figure 3.4). The rows 1 and 2 in the table 3.11 shows that for TF's the distance between the methyls in the small ring (1) is slightly larger (by 0.2 – 0.6Å) than in the large rings (2), while this difference increases to 2.0 – 2.2Å when DF's are used. On the other hand, when TF's (and B97D) are used with the ONIOM, the Me-Me distance in the large ring becomes larger, probably due to the twist in the sheet. When the remaining DF's are used with the ONIOM, the methyls behave identically with pure DF's. Moreover, DF's predict closer C_α – C_α and C-H – C-H (rows 4 and 5) interactions at the expense of a larger O – H and O -- N distance for the hydrogen bonds (rows 6, 7, 8). In general, the individual strands within the β-sheet optimized with DF's seem to be less distorted from the FES than their TF's conformers which is reflected by the smaller distortions energies reported in table 3.10. The predicted N – O distances agree with the experimental values of 2.903 +-0.004Å [77], except for ωB97x-D, which, like in the α- helices, predicts shorter hydrogen bonds. At last, the experimental C=O – N angle reported by Klebe et al [78] to be $153^o\pm$ from 0.2 to 5.26 depending on number of strands agrees with these calculated by ONIOM (150-154^o). TF's predicts slightly larger values of 157-160^o, while DF's predict even larger angles (162-167^o).

Summarizing, I was not able to find any experiment suggesting the existence of the unequally spaced side chains which should be the easiest feature of the structures optimized with DF's to observe. However, the structural data from the protein crystals lacks resolution available for small molecules and this might have been overlooked. Despite, I support my conclusion of artefactual structures with few other observations: the dihedral angles, twists and hydrogen bond parameters

Table 3.11: Geometrical features of $(Ac(Ala)_6NHMe)_4$ antiparallel β-sheets

		B3LYP	X3LYP	PBE1PBE	M06	M06-2X	B97D	ωB97x-D
ONIOM								
	Dihedral	-24.0	-23.6	-25.0	-5.50	-0.4	-24.9	-1.0
	1	4.463	4.466	4.440	6.135	6.050	4.346	5.990
	2	5.403	5.371	5.357	3.500	3.525	5.444	3.556
	3	5.130	5.124	5.089	5.620	5.630	5.089	5.595
	4	4.611	4.591	4.567	3.960	3.967	4.648	3.980
	5	2.510	2.493	2.459	2.112	2.151	2.452	2.136
6	C=O – H-N distance	1.890	1.881	1.872	1.940	1.954	1.889	1.933
	O – N distance	2.893	2.883	2.873	2.924	2.934	2.893	2.915
	C=O --- N angle	153.3	153.3	151.7	163.0	164.2	149.9	164.7
	C=O – H-C distance	2.376	2.363	2.357	2.456	2.418	2.359	2.402
7	C=O – H-N distance	1.870	1.862	1.854	1.929	1.947	1.857	1.921
	O – N distance	2.876	2.867	2.858	2.920	2.930	2.867	2.908
	C=O --- N angle	154.4	154.4	152.7	163.8	164.2	152.3	165.4
	C=O – H-C distance	2.400	2.385	2.379	2.438	2.424	2.400	2.399
8	C=O – H-N distance	1.900	1.891	1.884	1.943	1.935	1.888	1.918
	O – N distance	2.902	2.892	2.883	2.931	2.926	2.890	2.910
	C=O --- N angle	152.3	152.3	150.7	162.5	167.3	150.6	165.6
	C=O – H-C distance	2.379	2.365	2.359	2.452	2.427	2.351	2.423
Full DFT								
	Dihedral	2.4	4.6	2.8	0.9	3.8	-0.2	-3.9
	1	4.971	5.077	5.167	5.815	5.594	5.887	5.804
	2	4.812	4.662	4.459	3.752	3.603	3.744	3.688
	3	5.339	5.365	5.377	5.550	5.582	5.596	5.530
	4	4.422	4.360	4.273	4.007	3.959	4.067	3.993
	5	2.383	2.337	2.265	2.119	2.104	2.183	2.113
6	C=O – H-N distance	1.929	1.926	1.924	1.909	1.941	1.934	1.905
	O – N distance	2.925	2.919	2.915	2.897	2.916	2.925	2.887
	C=O --- N angle	157.0	159.6	159.2	164.1	161.8	163.9	162.9
	C=O – H-C distance	2.363	2.342	2.347	2.345	2.402	2.386	2.363
7	C=O – H-N distance	1.917	1.912	1.905	1.901	1.933	1.910	1.887
	O – N distance	2.917	2.910	2.900	2.887	2.912	2.907	2.875
	C=O --- N angle	159.7	160.3	160.4	164.4	162.3	167.2	164.7
	C=O – H-C distance	2.376	2.358	2.359	2.331	2.390	2.372	2.361
8	C=O – H-N distance	1.922	1.913	1.916	1.914	1.936	1.928	1.905
	O – N distance	2.921	2.912	2.910	2.900	2.917	2.919	2.890
	C=O --- N angle	158.1	159.1	159.1	163.8	163.5	165.6	163.1
	C=O – H-C distance	2.386	2.378	2.367	2.352	2.391	2.376	2.374

Table 3.12: Dihedral angles in optimized $(Ac(Ala)_6NHMe)_4$ antiparallel β-sheet using different functionals

Method	Position	ONIOM(DFT:AM1)				Full DFT			
		2^{nd} strand		3^{rd} strand		2^{nd} strand		3^{rd} strand	
		ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
B3LYP	1	-127	140	-121	141	-141	142	-142	142
	2	-128	140	-132	140	-140	143	-144	145
	3	-135	139	-133	139	-144	143	-143	141
	4	-132	138	-134	141	-142	140	-143	144
	5	-132	134	-135	137	-140	138	-143	137
	6	-127	131	-131	128	-131	130	-137	130
X3LYP	1	-127	140	-122	141	-144	144	-145	144
	2	-128	140	-132	140	-144	145	-146	147
	3	-135	139	-134	139	-146	145	-145	143
	4	-132	138	-135	141	-144	141	-145	145
	5	-132	135	-135	137	-141	138	-144	138
	6	-127	131	-131	129	-132	130	-139	130
PBE1PBE	1	-126	141	-122	142	-147	148	-146	146
	2	-128	141	-133	142	-146	148	-147	149
	3	-136	141	-134	141	-148	148	-147	146
	4	-133	140	-135	143	-145	144	-147	148
	5	-133	137	-136	139	-144	141	-146	142
	6	-128	132	-132	131	-134	133	-140	134
M06	1	-156	156	-154	154	-156	156	-154	154
	2	-160	146	-156	157	-156	158	-156	157
	3	-156	157	-154	154	-157	157	-154	155
	4	-154	156	-155	156	-155	156	-155	156
	5	-155	156	-154	155	-155	156	-153	154
	6	-151	150	-153	150	-151	150	-153	145
M06-2X	1	-159	161	-152	152	-160	162	-157	155
	2	-160	161	-159	160	-159	162	-160	162
	3	-159	160	-152	155	-159	161	-158	159
	4	-150	152	-152	155	-155	155	-156	158
	5	-150	150	-146	139	-153	153	-154	153
	6	-147	147	-149	146	-150	146	-151	147
B97D	1	-121	144	-118	141	-157	163	-154	151
	2	-125	137	-129	141	-156	156	-155	157
	3	-133	136	-134	139	-153	154	-154	153
	4	-131	138	-134	141	-150	152	-154	156
	5	-130	128	-136	137	-150	151	-152	152
	6	-118	125	-131	124	-145	143	-147	144
ωB97x-D	1	-158	159	-151	151	-157	163	-146	143
	2	-159	160	-157	159	-155	159	-154	157
	3	-157	158	-152	154	-153	156	-153	155
	4	-149	150	-153	155	-149	153	-154	157
	5	-149	148	-150	150	-150	153	-149	151
	6	-143	145	-149	141	-146	145	-145	144

which together allow me to draw the conclusion that DF's fail at prediction of the reasonable β-sheet secondary structures.

The calculations on the ΔH of folding per residue for the polyalanine α-helix clearly indicate that DF's grossly overestimated the magnitude of the helix stabilization. For the β-sheets, again, I observe much larger absolute interaction energies for the newer DF functionals. Unfortunately, I am not aware of any experimental data that could be used for the direct comparison, but the error should be similar to the one observed in the α-helices. However, unlike the helices, the β-sheets have much more geometric flexibility and DF's predict structures which parameters are out the range of the values observed experimentally.

Conclusions

The X3LYP and PBE1PBE functionals gave the best results and B3LYP performed almost as well. Interestingly, using AM1 in the ONIOM gave different results indicating that side chain interactions influence the β-sheet structure. Moreover, the results presented above does not support the assumption that the dispersion corrected (B97D, ωB97x-D), and widely parameterized (M06, M06-2X) functionals will provide better results for the α-helical and the antiparallel β-sheet motives. In fact, they failed in predicting both folding energies and structures that would agree with those found experimentally. Therefore, DF's should not be recommended for calculations on the peptides and other functionals that might be developed should be carefully evaluated using the available experimental data before they can be used in confidence in the field.

3.6 Concluding remarks

In the preceding sections I studied three different classes of molecules in order to compare performance of the traditional and dispersion-corrected functionals. Wherever it is possible I tried to compare the calculations to the experimental results (sections 3.4 and 3.5) to ensure the existence of a link between the theory and the experiment. The results discussed here helped me to evaluate the significance of the dispersion, and what is more important, the study on peptides allowed me to choose the best method to use in the following study of the aqueous solvation of the proteins secondary structures.

At first I showed that the dispersion becomes less important when coupled with induction, because other effects, like interaction between the induced dipoles or the charge transfer, which are described by traditional functionals, augment the interaction. We saw that for the set of dispersion-induction complexes (see table 3.3) the DF's performed worse than some of TF's and even for the pure dispersion complexes (noble gas dimers) they overestimate the attraction. Therefore, nevertheless still being aware of their limitations, I can conclude that in the systems where dispersion and induction cooperates together, TF's are reliable and can outperform the widely-parameterized dispersion-corrected functionals.

In the next step I compared the performance of DF's and TF's for the p-benzoquinone and pyrimidine complex for which reliable experimental data exists[65]. This is the system of special interest for us because the most stable complex might be formed either by the electrostatic (hydrogen bonding) or dispersive (π-stacking) forces and the experiment definitively proved that only hydrogen bonded complex exists in the argon matrix at 12K. The experimental findings might be related to the importance of dispersion forces in the base-stacking in the DNA and it is crucial to see if the dispersion-corrected functionals can reproduce the experimental finding. In all cases, except for M05 and M05-2X, we observe that DF's overestimate stability of the dispersion-driven (either π-stacking or T-shape) complexes, predicting them to be more stable than the hydrogen bonding dimer. The results suggest that DF's (and ab initio MP2 method) do not perform well when compared to the experimental data despite their perfect match to the theoretical databases.

In section 3.5 I apply different methods to the two most abundant protein secondary structures – α-helix and β-sheet. As I emphasized in the previous paragraph, the dispersion methods provide the better fit to the theoretical databases therefore they use should, in principle, improve our results. However, I decided to evaluate various DFT methods against available experimental data: the enthalpies of folding, and structural parameters, and found that DF's do not agree with the experiment, while TF's do. In essence, DF's overestimate folding enthalpy of the α-helix by 400% and predict artefactual structures of antiparallel β-sheets, while calculations with TF's stand in the agreement with the available experimental data. In fact, this is not surprising result when one analyzes it in the context of the previous work. Dispersion is not the dominant force in the peptides and we saw that it loses its significance when coupled with induction, the interaction present due to the abundant amide groups.

Chapter 4

Aqueous Solvation of Protein Secondary Structures

"Curiouser and curiouser!"

Alice in Wonderland[85]

4.1 Introduction

Proteins are specialized micro-machines which evolved during eons to achieve necessary complexity. It would be naive to think that during evolution Nature did not take any advantage of unique properties [86] of most abundant molecule present in living organisms – water. In fact mutual interactions with water are essential in protein structure; therefore functioning. We can distinguish two kinds of water molecules: (1) solvent molecules which hydrate surface of a protein modifying the potential energy landscape (and therefore structure and dynamics of whole protein) and (2) specific waters bonded in hydrophilic cavities inside the protein where it plays crucial role in controlling functions of protein. For example water found in the active site of ribosome is essential for the catalytic formation of peptide bond in protein expression from mRNA [87]. However both kinds of waters require different treatment in molecular simulation. While the latter type of water can be explicitly included in high level molecular calculations, the former, due to the computational limitations, usually needs approximate methods to make the problem approachable by modern computers. Some of the approximations are discussed in section 2.3.

The goal of this chapter of my thesis is the study of the hydration of the secondary protein structures. However, first I will start from the rethinking the hydrophobic effect. It is thought that the positive free energy of the solvation of non-polar species in water is an effect of the decreasing entropy during the transfer from the gas to the aqueous solution. I will try to approach the problem with regard to the solvent's Cohesive Energy Density which can be related to the solvent's internal pressure and I will show that there is no need for invoking the entropic argument when the energy required to make a hole in water is taken into the account. Then, I will investigate the stability of the β-turns both in the gas phase and in the solution. I will show that the classification based on the hard-sphere modeling [4] is outdated and does not agree with the recent findings by high level theory. Next I will study the solvation of the α-helix which might undergoe two types of solvation – first, the cooperative solvation by strongly bonded waters at the helical termini that tend to augment the molecular dipole moment, and then the solvation by the bulk solution. As the last project, the solvation of the β-sheets will be discussed. In the contrast to the α-helices, the β-sheets do not experience cooperative solvation. The waters hydrating the termini of the sheets are found crucial to prevent the aggregation of strands and I find that the continuous solvent models are not able to properly simulate the solvation of the strand.

4.2 Hydrophobic Solvation: Implications for the hydrophobic effect

Introduction

The hydrophobic effect causes nonpolar molecules to aggregate or fold in aqueous solution. This effect has been invoked to explain the acceleration of certain organic reactions in water (compared to organic solvents), the formation of micelles in water, and (perhaps most importantly) protein folding. However, there appears no universal agreement on what the physical explanation for this effect might be.

The debate about the hydrophobic effect might be tracked down back to the paper by Frank and Evans about solubility of noble gases in water published in 1945 [88]. They observed that the nonpolar molecules have a positive free energy of transfer from the gas phase to the aqueous solution ($\Delta G_{gas \rightarrow water}$), although the heat was released. This observation begs for invoking the ΔG=ΔH-TΔS relation which allows to conclude that the entropy of the process must be negative.

They physically interpreted the results as a formation of “icebergs” in the solvent around the solute. The problems with this interpretation had been already recognized by Kauzmann [89] in his chapter in 1959, however he did not reject the entropic explanation of the process.

On the other hand, physical organic chemists seem not to follow the entropic paradigm, but rather cite Abraham’s [90] work who only reports measured free energies. He noticed that the noble gases and methane experience different effect than larger non-polar hydrocarbons; the former follows the slight decrease in $\Delta G_{gas \rightarrow water}$ with increasing the atomic radii [90], while for the latter ΔG increases by 0.54 kcal/mol per each CH_2 group. Thus, why does hydrophobic effect act differently on those non-polar molecules?

Another interesting consequence of the hydrophobic effect is how water affects the reaction rates, for instance the Diels-Alder reaction. The aqueous solvation speeds up reactions with negative volume of activations and affects the distribution of endo and exo products since they are formed via different transition states [91]. Furthermore, different solvent’s composition or electroconstraining salts can affect the reaction rate as well, by accelerating it similarly to the increased pressure. Physical chemists use as a measurement of the internal pressure of the solvent the Cohesive Energy Density (CED) defined as $(\Delta H_{vaporization}\text{-}RT)/V_{molar}$. For water it is estimated to be 0.55 kcal/cm^3 or 22,700 atmospheres, while an average CED of organic solvents is measured to be around $\frac{1}{10}$ of the water’s CED. Gajewski [92] found that the Diels-Alder reaction rates correlate best with the solvent’s CED.

The $\Delta G=\Delta H\text{-}T\Delta S$ relation holds only if the transfer is done under a constant (in this case atmospheric) pressure. To transfer a molecule, first we have to make a hole in the solvent, which requires work, $P\Delta V$. If the work is done against atmospheric pressure, both terms, one due to removing a hole from the gas phase and another responsible for making a hole in the solvent, cancel out. However if we consider it as a work against the internal pressure of water (CED), the term gains importance.

From the standard thermodynamic textbook we can learn that $G = H - TS = U + PV - TS$. We also know that:

$$dH = TdS + VdP \text{ (at defined S and P)} \tag{4.1}$$

$$dG = -SdT + VdP \text{ (at defined T and P)} \tag{4.2}$$

and $\left(\frac{dG}{dP}\right)_T = V$ and $\left(\frac{dH}{dP}\right)_S = V$.

If the external pressure during the transfer changes, for example by removing a sphere with volume V from one phase (for example a gas phase with p_1) and placing it in another phase (for example water with p_2), the required work is given by $V(p_1\text{-}p_2)=V\Delta P$. Similarly, if one simply integrates the above equations, the $G=V\Delta P$ (at constant T and V) and $H=V\Delta P$ (at constant S and V) relations would be restored. However the estimation of these terms is not simple since we do not really know what the pressure is against which the work is done and what the volume of the hole is.

In this study, I calculated what is the effect of making a hole inside a water cluster within the framework of Density Functional Theory. Using three noble gases: neon, argon and krypton, I estimated what is the energetic penalty making a hole big enough to accommodate the entity. The noble gas atoms constitute perfect candidates for this study since the detailed data on the free energies and enthalpies are available [88],[90], they do not contain internal structure which simplifies a conformational entropy problem, and due to underestimation of dispersive interactions by functional chosen for this work (B3LYP), we minimize the stabilizing forces between the solute and the solvent and we can focus on the energy required just to make the hole in the solvent.

Methods

The calculations were performed using G09 and G03 [50] software packages using B3LYP functional and 6-31G(d,p) basis set. I needed to choose a different basis set than that used in other sections (d95(d,p)) of the thesis due to lack of basis functions defined for the noble gases in d95(d,p) basis set. Multiple optimizations of each system (50 waters cluster and 50 waters cluster with the solute inside) were performed in order to localize the most stable minima. Each cluster was then reoptimized on the counterpoise (CP) corrected surface to remove the artificial binding due to the basis set superposition error (BSSE). Harmonic vibrations were calculated to confirm the stable minima (no imaginary frequencies) and to evaluate enthalpies and entropies at 298K. The energies and the enthalpies of the gas $\rightarrow$ solvent transfer were calculated as a difference between the optimized water cluster – noble gas complex (which is a local minima since the global

minimum would contain the noble gas outside the cluster) and the fully optimized 50 waters cluster and noble gas separately. The energies of making the hole were defined as a difference between 50 waters cluster in the geometry of the complex (with the noble gas removed) and fully optimized 50 waters cluster. Since the former is not a real minimum (and in principle the hole collapses when one tries to optimize the structure) we cannot evaluate enthalpies of this structure.

Results

I will focus on the results for the neon atom first and, since argon and krypton follow the same reasoning, I will summarize all of them in the further part. To estimate the volume of the hole I use van der Waals radii (1.54 Å for Neon) as reported by Bondi [39]. However, we know that the hole is not perfectly spherical and I decided to assume the sphere to be the lower bound of the hole's volume, while a cube with an edge equal to two times the van der Waals radii will consitute the upper bound.

In the table 4.1 I collected all data relevant for the following reasoning. The energy necessary to make the hole of a size of 9.2-17.6 $Å^3$ was calculated to be 7.0 kcal/mol and energy of the transfer, i.e. making the hole and placing there the neon atom is 7.5 kcal/mol which reflects repulsive interaction between the solute and the water cluster, as expected form B3LYP. Now, having the energy required to make the hole and its volume, we can calculate what a pressure we work against is. Using ΔE of 7.5 kcal/mol, these values are 0.82 kcal/cm^3 and 0.43 kcal/cm^3, respectively, for spheres and cubes, which do bracket the water's experimental CED of 0.55 kcal/cm^3.

Moreover, the ΔH of transfer at the constant V and S (thus $TdS = 0$, what I show to be the case for the non-polar gas molecule tranfer) is given by VΔP + q if the change includes the pressure. If I use the calculated ΔE_{hole} and the experimentally measured heat q=-1.88 kcal/mol, I get ΔH=+5.12 kcal/mol which is very close to the experimental free energy (ΔG=6.94 kcal/mol) of transfer that suggests that the entropic TΔS term might be small or negligible.

The careful examination of the water cluster (see figure 4.1) shows no change in the number of hydrogen bonds around the solute suggesting that it only weakens the surrounding bonds in respect to those in the bulk solution, which is reflected in increasing distance between the oxygens around the solute.

Argon and krypton have lower measured free energy of transfer from the gas to water which

Table 4.1: Calculated properties for Noble Gases in a 50 water cluster and reported experimental values for aqueous noble gas solutions. Please see details in text.

	Ne	Ar	Kr
$\Delta E_{tranfer}$ (kcal/mol)	7.5	12.3	14.6
$\Delta H_{tranfer}$ (kcal/mol)	9.2	13.8	16.3
ΔS_{vib} (cal/degree·mol)	1.5	-0.1	-0.2
van der Waals Radius (Å) [39]	1.54	1.88	2.02
Volume cm^3/mol (sphere ↔ cube)	9.2 ↔ 17.6	16.8 ↔ 32.0	20.8 ↔ 39.7
Volume (T) cm^3/mol (liquid, experiment [93])	16.8 (25 K)	24.2 (84 K)	32.2 (116 K)
Heat $\Delta H_{literature}$ (kcal/mol)[88]	-1.88	-2.73	-3.55
$\Delta S_{literature}$(cal/degree·mol)[88]	-28.8	-30.2	-32.3
ΔS_{hole} (kcal/mol)	7.0	7.7	8.2
$-T\Delta S_{literature}$ (298 K) [88]	8.6	9.0	9.6
ΔH (as ΔS_{hole} + q)	5.2	5.0	4.6
K_H ($atm/10^4$) [94]	12.30	3.96	2.23
$\Delta G_{literature}$ (kcal/mol) [90]	6.94	6.27	5.93
Effective CED (kcal/cm^3) from ΔE_{hole}	0.82 ↔0.43	0.73 ↔ 0.38	0.70 ↔ 0.37

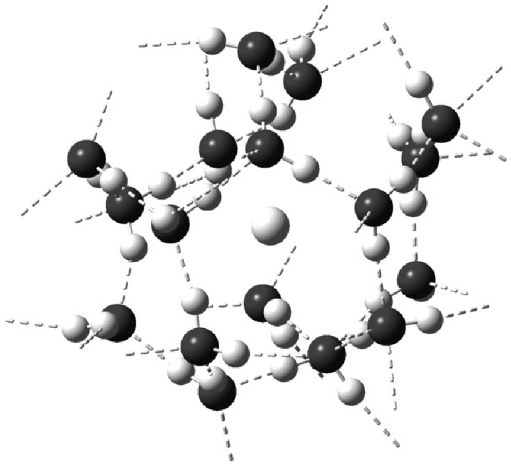


Figure 4.1: Inner solvation shell taken from 50 waters + Neon. Note that each water near the Neon participates in four H-bonds.

indicates that they are more soluble. On the other hand, they have larger van der Waals radii (1.88 and 2.02 Å) so in principle the hole required to accommodate them should be bigger and it would require more energy to form it. Nevertheless, the larger atoms have also stronger dispersion interaction with waters as experimentally measured by released heat (q) which compensates for the larger holes. Following the same reasoning as for neon, the calculated boundaries for the water's CED are 0.73 and 0.38 kcal/cm^3 for argon and 0.70-0.37 kcal/cm^3 for krypton. Furthermore, calculated ΔH's of transfer from $\Delta H=q+\Delta E_{hole}$ (5.0 and 4.6 kcal/mol, respectively, for argon and krypton) restore the decreasing trend in free energy of transfer of noble gases.

Discussion

The foregoing results probably triggered many questions that I will try to assess in the following part.

First we should realize that we are trying to combine two opposite regimes: the macroscopic kingdom of the classical thermodynamics, homogenous on the macroscopic level, and the granular world of quantum chemistry where properties do not change in a continuous manner. This is an important realization, especially when one tries to understand the difference between the pressure applied on an object in water, as described by the Archimedes law known from the common sense, and the internal pressure (CED) of water acting on molecules in the microscale. In principle, there should exist some connection between these two regimes, for example like that suggested by [95]. In the microscale we work against the strong hydrogen bonds surrounding the solute (somehow similar to the model proposed by Lee [96]) instead of the atmospheric pressure.

The next important question is what the magnitude of the pressure to be considered in water is. If the work is done just against the atmospheric pressure, then the VΔP term should cancel out. However it does not explain the magnitude of the calculated energy necessary to make the hole in the water cluster. We know that the internal energy (E in the calculations) U equals q+PΔV and making the hole in the cluster must correspond to reversible work. However, only the CED allows to explain the magnitude of the energy necessary to make the hole.

The next important question is what happens to the entropy of the system. In the biochemists' explanation decrease in the entropy leaded to a formation of "icebergs" around the solute, however we postulate that the change is negligible and the difference between experimental free energy and

enthalpy should be attributed to the work done by the solute.

In general, the entropy can be broken down to the contributions from the configurational entropy (number of local minima of equivalent energy) and local entropy (electronic, vibrational, rotational and translational) of the individual minimum. If we make a reasonable assumption that the number of low energy configurations is constant for all solutes (neon, argon and krypton) then the entropy of any equivalent minimum should be equal to its internal entropy plus a constant. Since the translational and rotational entropies do not vary significantly between clusters with and without solutes (just by 0.3 and 0.1 cal/(degree·mol)), and we restrict our discussion just to the electronic ground state, the major contribution to the ΔS for the individual low energy configurations with and without the solute should come from the vibrational entropy. ΔS of transfer from the gas phase to the solution also includes the entropy of mixing (always positive), the reduction of the entropy upon decreasing the number of entities, and the increase in the entropy due to formation of additional vibrational modes. The first two contributions are the same for all solutes and are not included in this analysis. In order to be able to evaluate the influence of the last contribution, I subtract the three lowest vibrations from the solute-water cluster complex. Please note that since the lowest vibrations contribute most to the overall vibrational entropy, I calculated the lower limit to the ΔS_v.

The calculated ΔS_v's vary from 1.5 cal/degree·mol to -2.0 cal/degree·mol but they are much larger than approximate value of -30 cal/degree·mol reported by Frank and Evans [88]. Moreover, my estimated values constitute the lower limit to the entropy so the difference can only get larger. Thus, I do not find that entropy is reduced due to increased ordering of waters around the solute which should be reflected by the decrease of the vibrational entropy of the water cluster.

I also note that my argument is valid only for the noble gases. The larger molecules and proteins can experience different entropic effects (like measurable change in the heat capacity during protein folding) due to restriction of internal motions [63] and the reduced exposition of large hydrophobic surfaces which increase mobility of waters [95]. Even the long hydrocarbon chains have the conformational flexibility that the noble gases (and methane) lack. Therefore I want to distinguish between hydrophobic effect experienced by the large molecules which are able to assume more compact conformations in water; and hydrophobic solvation experienced by the simple noble gases and methane which have no capacity for the conformational change. This might

explain the different behavior of $\Delta G_{gas \rightarrow water}$ for the rare gases and larger hydrocarbons in water as reported by Abraham [90].

Conclusions

The interpretation of the hydrophobic effect as being driven by the change in entropy of the aqueous solvent derives primarily from the work of Frank and Evans [88] who showed that transfer of non-polar solutes to water has $\Delta G > 0$, but heat is released, q (not the ΔH) < 0. However literature supports us with many examples against the entropic interpretation of the hydrophobic effect. For example, Abraham showed that the measured free energy of transfer from the gas to the aqueous solution for the noble gases and nonpolar organic solutes is the subject to the two different relations [90] which is inconsistent with the entropic model which should be universal. Also, Le Noble et al have exhaustively studied Diels Alder reaction, which $\Delta V_{activation}$ of the transition state is negative, and found that the reaction rates behave similarly in water and under applied pressure which suggest the common source of the effect[91]. In general, the rate of the reaction correlates with the CED's of the solvent. Moreover, CED can be affected by other factors, like constricting salts (for instance LiCl) and different solvent compositions and this correlation with reaction rate is observed as well [97]. At last, the well-known constants in the Henry's law equation for helium and neon (which describes the equilibrium between the liquid and the partial pressure of the vapors) are the same order as the water's CED which implies that in order to maintain the concentration of the gas in the solution and the gas phase in equilibrium ($\Delta G=0$) the partial pressure of the gas should be similar to water's CED.

In my work I analyze few more arguments against the entropic interpretation of the hydrophobic effect. Using DFT I show that energy required to make a hole in a water cluster is not negligible but agrees with the values predicted by the CED of water. I also point out that the vibrational analyze shows no significant change in the vibrational entropies which should be observed if the "icebergs" were formed. At last, I show that there is no need to invoke entropy if one considers $\Delta H=q+w$ (where w is work necessary to make a hole in the liquid) instead of $\Delta H=q$.

4.3 Study of β-hairpins in Antiparallel β-sheets

Introduction

The idea behind the project was born during an investigation of the stability of the antiparallel β-sheets having β turns connecting adjacent stands. We observed that many structures contain C_7 instead of the expected C_{10} hydrogen bond (HB) postulated by standard biochemical textbooks (for instance, please look into [98], the work on β-sheets has been published elsewhere [99], second part of work has not beet published yet). This discrepancy motivated me to investigate the problem in greater detail.

The antiparallel β-sheets consist of two subunits - small (C_{10}) and large (C_{14}) rings (please see figure 3.4 in section 3.5) which are composed of 10 (small) and 14 (large) atoms connected by two hydrogen bonds into a cyclic structure. Visvanathan et al. showed that the interaction enthalpy of a glycine diamide dimer changes from -4.85 to 13.99 kcal/mol between the small (SR) and large (LR) rings, respectively. Larger β-sheets experience similar difference in stability depending on the number of SR's and LR's in the sheet. Another example, the smallest possible β-hairpin motif contains one large and one small ring, in which the latter is a part of the postulated C_{10} hydrogen bond. From the foregoing, one might expect that this HB in the small ring is relatively weak and thither weakening or even breaking is possible. Furthermore, other possible hydrogen bonding entities, for example: non-bonded amide between i+1 and i+2 residues, other parts of the protein or water; might facilitate formation of a new hydrogen bond in place of the weaker C_{10} HB.

The different types of the β-turns, originally described by Venkatachalam [4] and reviewed by Rose [100], are defined with the assumption that the C_{10} HB between residues i and i+3 is intact, and they are divided into classes based upon the ϕ and ψ dihedral angles on residues i+1 and i+2. This straightforward definition requires detailed knowledge of all atomic positions which is problematic for the proteins due to (1) low resolution of protein crystal structures, and (2) the fact that hydrogen atoms positions are not accessible in the electron diffraction experiment. In essence, the inaccuracy in N--- O distance (which is by far the best measurement of the existence of HB) might not allow to determine between the putative C_{10} HB and, for instance, the potential C_7 HB with non-bonding peptide, because the N---O distance would not change drastically during a rotation of the carbonyl group.

Furthermore, the original definition divides the β-turns into four classes (or six if one includes type III and III' helical turns which are essentially short 3_{10}-helices and will be disregarded in this work) which are required to have specific chirality on the i+1 and i+2 residues. The type I turn contains two L amino acids and two D enantiomers are required by type I' turn. The other two types, type II and type II', require these two residues to have the opposite chirality or one of them being the achiral glycine.

Discussed classification of Venkatachalam is not the only criteria used for the β-turns. Another definition involves the $C_\alpha - C_\alpha$ distance [101] or other structural parameters [102] and do not distinguish between different classes of the turns. However, the original definition seems to be the most popular among biochemists.

The smallest peptide necessary to mimic the β-turns like those found in β-sheets must contain at least 6 amino acids capped with NH_2 and acetyl groups, otherwise missing C_{14} ring (if the peptide is shorter) would make the C_{10} hydrogen bonds stronger. The capped alanine hexapeptide, both in the gas phase and in the aqueous solution, are considered in this section. I explore its conformational space in order to localize all possible minima. Calculations in the gas phase were performed for four different diastereomers with different chirality of alanines at the i+1 and i+2 positions as specified by the original definition. Harmonic vibrations were calculated in order to evaluate enthalpies and confirm existence of the stable minimum on the potential energy surface. All calculations were performed using B3LYP functional with d95** basis set combined with the semiempirical AM1 method for methyl side chains as implemented in the ONIOM approach in Gaussian 09 package of software [50]. This method was shown to be accurate for peptides (please refer to chapter 3 and [57], [58], [68]). The details on calculations in solvent are included in the further part of the section.

Gas phase

Figure 4.2 presents all structures and their relative enthalpies which were found during the exploration of $Ac(Ala)_6NHMe$ PES. The turns are divided into two classes: (A) those that contain the cyclic C_{10} HB bewteen the residues i and i+3, and (B) those that do not contain that cyclic C_{10} HB. The table 4.2 presents calculated relative enthalpies for all peptides with different chirality at i+1 and i+2 positions, while table 4.3 lists mean unsigned errors (MUE's) with respect to the

standard ϕ and ψ dihedrals as defined in the literature [100]. All attempts to optimize the B_{III} structure with DD and DL amino acids converged to the B_I conformer. I also could not find fourth possible structure with the C_7 HB involving the i+2 position which converges to the A_I turn.

As can be seen from table 4.2, the B_I structure is the most stable turn for all diastereomers and it contains the cyclic C_7 HB in place of the C_{10} HB. Besides the common most stable structure, the order of stability changes between different diastereomers; for the LL amino acids the B_I structure is followed by A_V, which contain two extra C_7 HB's at the strand's termini, B_{IV} with the second C_7 HB at the i+2 position and B_V, which also contains two C_7 HB's but in the opposite puckering. The lowest energy structure which contains only C_{10} HB, A_{II}, is 3.8 kcal/mol less stable than the B_I turn. For the DD stereoisomer, the B_I turn is followed by A_I with the single C_{10} HB, and then by the previously mentioned B_{IV} and A_V turns; and the LD stereoisomer is followed by three B-type of turns (B_{II}, B_V and B_{IV}), all of which are within 1 kcal/mol above the B_I minima, and the first A-type turn is 2.2 kcal/mol above the B_I turn. Finally, in the case of DL polypeptide the B_I is followed by the A_V (1.6 kcal/mol), B_{IV} (2.7 kcal), and A_{II} (3.0) kcal/mol turns.

Table 4.2: Relative Enthalpies (in kcal/mol) for isomers of $Ac(Ala)_6NH_2$ designed by the chirality of residues i+1 and i+2

Turn:	LL	DD	LD	DL
A_I	6.1	1.4	2.9	4.6
A_{II}	3.8	3.5	4.4	3.0
A_{III}	7.2	9.8	7.5	9.5
A_{IV}	7.9	2.8	3.1	7.6
A_V	2.3	1.6	2.2	1.6
B_I	0.0	0.0	0.0	0.0
B_{II}	4.8	2.1	0.2	6.7
B_{III}	8.5	-	7.7	-
B_{IV}	2.1	1.6	0.9	2.7
B_V	3.1	2.6	0.3	5.4

Table 4.3 presents how different turns correlate with the standard dihedral angles as defined by Ventakatachalan [4]. The A_I turn correlates best with the type I' turn, A_{II} correlates best with II', and A_{III} correlates best with angles defined for type I. The A_{IV} and A_V turns which contain two extra C_7 terminal HB's that are not present in the original definition, correlates the best with I' and II' turns receptively, and the latter has the MUE smaller than that for the A_{II} turn. I want to

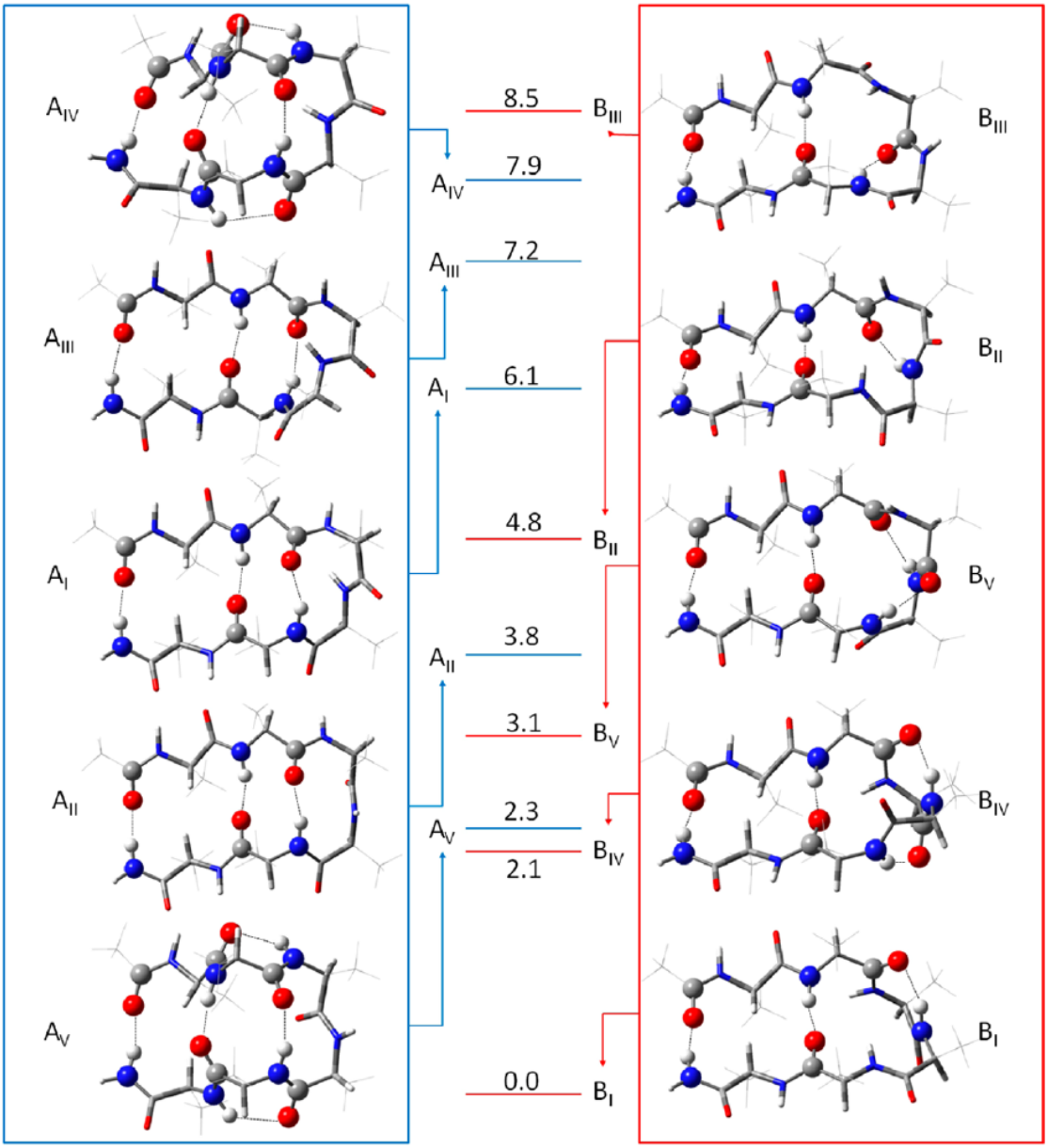


Figure 4.2: Type A and B turns with relative enthalpies (in kcal/mol) for the all-L-Ala structures

note that the multiple optimizations of the LL and DL hexapeptides using partial constrains and dihedrals defined for the type II β-turn did not lead to stable minima but converged to B_{II} turn. Also, B_I and B_{II} turns have MUE's between 20 and 30 for II' and II type of turns, respectively, and the former's MUE is only 10 degrees larger than that for the A_{II} turn, while the latter has the lowest MUE for the type II class.

Table 4.3: Mean Unsigned Errors (MUE's) with respect to the standard ϕ and ψ dihedrals for Type I, I', II, and II' β-turns

	Standard Dihedral Angles:							
	ϕ_1	-60	-30	-90	0			
	ψ_1	60	30	90	0			
	ϕ_2	-60	120	80	0			
	ψ_2	60	-120	-80	0			
	MUE's with respect to standard ϕ's and ψ's:							
Stereoisomer:	I	I'	II	II'	I	I'	II	II'
	A_I				B_I			
LL	94	6	61	84	43	73	123	25
DD	97	17	72	87	52	76	126	23
DL	97	14	69	87	47	77	127	26
LD	93	9	64	83	48	72	122	23
	A_{II}				B_{II}			
LL	60	90	140	11	77	48	21	127
DD	57	87	137	7	81	51	32	131
DL	61	91	141	11	73	43	22	123
LD	57	88	137	8	85	55	32	135
	A_{III}				B_{III}			
LL	19	97	87	74	51	138	128	62
DD	8	93	83	63	-	-	-	-
DL	15	97	87	70	-	-	-	-
LD	15	94	84	67	50	133	123	59
	A_{IV}				B_{IV}			
LL	106	18	73	96	64	89	139	35
DD	102	16	71	92	68	88	138	36
DL	104	17	72	94	68	92	142	36
LD	102	15	70	92	64	85	135	35
	A_V				B_V			
LL	55	82	132	5	89	67	28	139
DD	64	84	134	9	94	67	31	144
DL	47	84	134	4	85	62	28	135
LD	48	82	132	7	89	62	30	139

It is worth noting that Venkatachalam specified that different β-turns should contain only amino acids with specific chirality at the positions i+1 and i+2, but in my calculations I find that the each conformer (except B_{III}) posses the stable minima for all four stereoisomers. While for the A-type, B_I, and B_{II} turns the most stable stereoisomer is the one expected for the turn which have the lowest MUE's for that structure, the differences are sometimes insignificant. For instance, the DD stereoisomer is only 0.5 kcal/mol less stable than the DL for A_{II} turn. Moreover, if one narrows down the classification to the $A_I - A_{III}$ turns, which resemble best those originally specified by Rose [100], only DD and DL stereoisomers adopt geometry expected for the given chirality. The LL peptide in the A_{II} conformation (closest to the type I turn) is more stable than A_{III} by 3.4 kcal/mol and LD's native structure is not a stable minimum on PES.

Our data suggests that the most stable turns are the B-type and contain the C_7 hydrogen bond in the place of C_{10}, while the A-type of turns are higher in energy. Therefore, any change in the structure leading to the formation of an unfavorable C_{10} hydrogen bond is an effect of the environment, not the intrinsic property of the turn. For example crystal structure of VV-*D*-PGVV (type II' turn) solved by Karle et al [103] contains an excessive network of HB's, not only in the (infinite) plane of the turn, but also in the perpendicular direction extended by water molecules. These waters bind to the amides which are not involved in the intramolecular HB and bridge two turns from separate crystal lattices. This network can stabilize the turn in energetically less favorable geometry due to the cooperative HB's. In case of proteins the situation is even more complicated, the turns can be exposed to the solvent or interact with other parts of the peptide which would provide the stabilization of the turn [100], somehow resembling Etter's rule in the crystallography that all hydrogen bond donors and acceptors have to be satisfied [104]. In fact, the data presented in table 4.3 shows that few different structures have low MUE with regard to one classically defined turn. Thus, some dynamic in the C_{10}/C_7 hydrogen bond might be anticipated depending if additional interactions would promote formation of one of the specific conformations. For example the β-sheets can be thought of as an one-dimensional crystal structure in which sheets expand infinitely in the each direction. In those sheets the adjacency of another strand influences the turns' conformation. Ali-Torres et al recently investigated stability of various antiparallel β-sheets [99] and found that the most stable isomers contain a mix of B_I and A_{II} turns which are the most stable turns in their classes (when one disregards A_{IV} turn which has two terminal C_7

and cannot exist in the sheet). The A-type turn is formed in the middle of the sheet where its C_{10} HB can anticipate in sheet's cooperative HB network, while the B-type turn was present at the terminal strands. The sheets which contain the A_{III} turns (one that correlate best with type I turn) were found to be much higher in the energy.

Aqueous Solvation

Another important question is the importance of the aqueous environment. Turns in the proteins are usually exposed to the solvent which stabilizes the non-bonding peptide group (all HB's have to be satisfied). The differences in the solvation free energies can affect the turn's conformational space and change the order of stability. Moreover, water constitutes a different environment than the one-dimensional β-sheets which can interact only by the constrained peptide groups, in water the solvation is anisotropic and all broken HB's can be potentially satisfied by water molecules. Therefore, the potential breaking of the C_{10} HB will not bear a penalty associated with breaking two other HB's because solvent will stabilize the newly exposed polar group.

Nonetheless, one should not forget that the amide groups in peptides are in close proximity from each other (4-5Å) and water has a finite size (2-3 waters fit in a space between two peptide groups), thus the solvation process at one amide can contribute to the solvation of another. In other words, the solvation of polyamides is nonadditive; however, lack of experimental data does not allow estimation to what extent. In fact, the lack of reliable experimental data is a significant problem in accurate prediction of solvation energies, only four experimentally measured free energies for monoamides are available and these three data points are linearly related, because they involve very similar molecules – Acetamide, trans-N-methylacetamide, cis-N-methylacetamide and N,N-dimethylacetamide [105]. The only available data for the polyamides come from theoretical predictions (for instance: [106], [69]) which often do not include all effects, like the cavitation energy or the dispersion, but focus on the electrostatic part of the solvation (vide DelPhi program used by Baldwin [107] or default implementation of solvent models in Gaussian [50]).

In my study I use implicit and explicit solvent model approaches. Unfortunately, the technique that combines the first explicit solvation shell with the implicit solvent model for bulk water proved to be plagued by numerical instabilities which generate many low lying imaginary frequencies. I use the Conductor-like Polarizable Continous Model (CPCM) which is a revised version of

the Integral Equations Formalism - Polarizable Continous Model (IEF-PCM) with the infinite dielectric constant (for more detail please look into section 2.4). The Pauling radii [108], which have the lowest MUE for calculated free energy of solvation for four amides and water, were used to build the cavity around a solute. Moreover, there are some theoretical arguments that cavities based on van der Waals radii are the best suited for work with the polarizable solvent models [32]. Also, I want to point out that definition of the cavity is the most significant parameter in the solvent models, the calculated free energies vary substantially with the radii used [109], and solvent models are usually parametrized to work with a specific class of optimized parameters (for example SMD [45], SMx.x models [110] [111] [31], MST models [112]).

For the solvation by explicit water molecules I decided to use 12 waters to simulate the first hydration shell around the i – i+3 residues, since the smaller numbers did not provide efficient solvation of the turn. However, 12 waters are not enough to stabilize the reference polyproline II (PP2) structure which requires at least 11 waters per residue [113] to become more stable than the fully extended strand (FES), therefore I use the PP2 structure only with the continuous model. All structures were fully optimized and the vibrational frequencies were calculated to confirm the stationary points on the PES and to evaluate enthalpies. The BSSE between the waters and the turn, and between the waters themselves were added as a single point. Multiple optimizations were performed to fully explore the conformational space, but the main focus was put on solvation of the C_{10} HB in the turn. The complexity of the PES does not allow me to state that I found the most stable complexes, however, I point out that the multiple optimizations always lead to structures higher in the energy within a narrow range. Thus, while I might not be able to indicate the most stable conformer between the few with close energies, I can estimate general stability of turns if differences are significant.

Table 4.4 presents the results for the continuous solvation. Since the structures were reoptimized in the presence of solvent, I specify the initial (starting) and optimized (final) geometries for each turn. The asterisk indicates that the dihedrals at the i+1 and i+2 residues did not change substantially, but the initial HB's were broken and the polar groups were exposed to the solvent. For the not labelled structures, the HB's in the optimized geometries remained intact.

Comparing these results with those for the gas phase, two observations are clear: (1) the B_I turn is still the most stable conformer, and (2) the difference in enthalpy between the most and

Table 4.4: Relative Enthalpies (in kcal/mol) for LL isomer of $Ac(Ala)_6NH_2$ in implicit and explicit solvation models

Turn structure Initial→Final		ΔH	Turn structure Initial→Final		ΔH
CPCM					
A_I	$\rightarrow A_I$	1.3	B_I	$\rightarrow B_I$	0.0
A_{II}	$\rightarrow A_{II}$	1.8	B_{II}	$\rightarrow B_{II*}$	2.2
A_{III}	$\rightarrow$B	3.5	B_{III}	$\rightarrow B_{III*}$	0.2
A_{IV}	$\rightarrow A_I$	1.3	B_{IV}	$\rightarrow B_{IV}$	3.7
A_V	$\rightarrow A_V$	1.7	B_V	$\rightarrow B_{V*}$	2.8
PP2		2.8	B_{VI}		1.9
12 waters					
A_I	$\rightarrow A_I$	15.5	B_I	$\rightarrow B_{I*}$	2.9
A_{II}	$\rightarrow A_{II}$	0.0	B_{II}	$\rightarrow B_{II*}$	8,1
A_{III}	$\rightarrow A_{III}$	11.7	B_{III}	$\rightarrow B_{III*}$	2.7
A_{VI}	$\rightarrow A_{VI}$	1.6	B_{IV}	$\rightarrow B_{IV*}$	15.1
			B_V	$\rightarrow B_{V*}$	13.9
			B_{VI}	$\rightarrow B_{VI}$	22.2

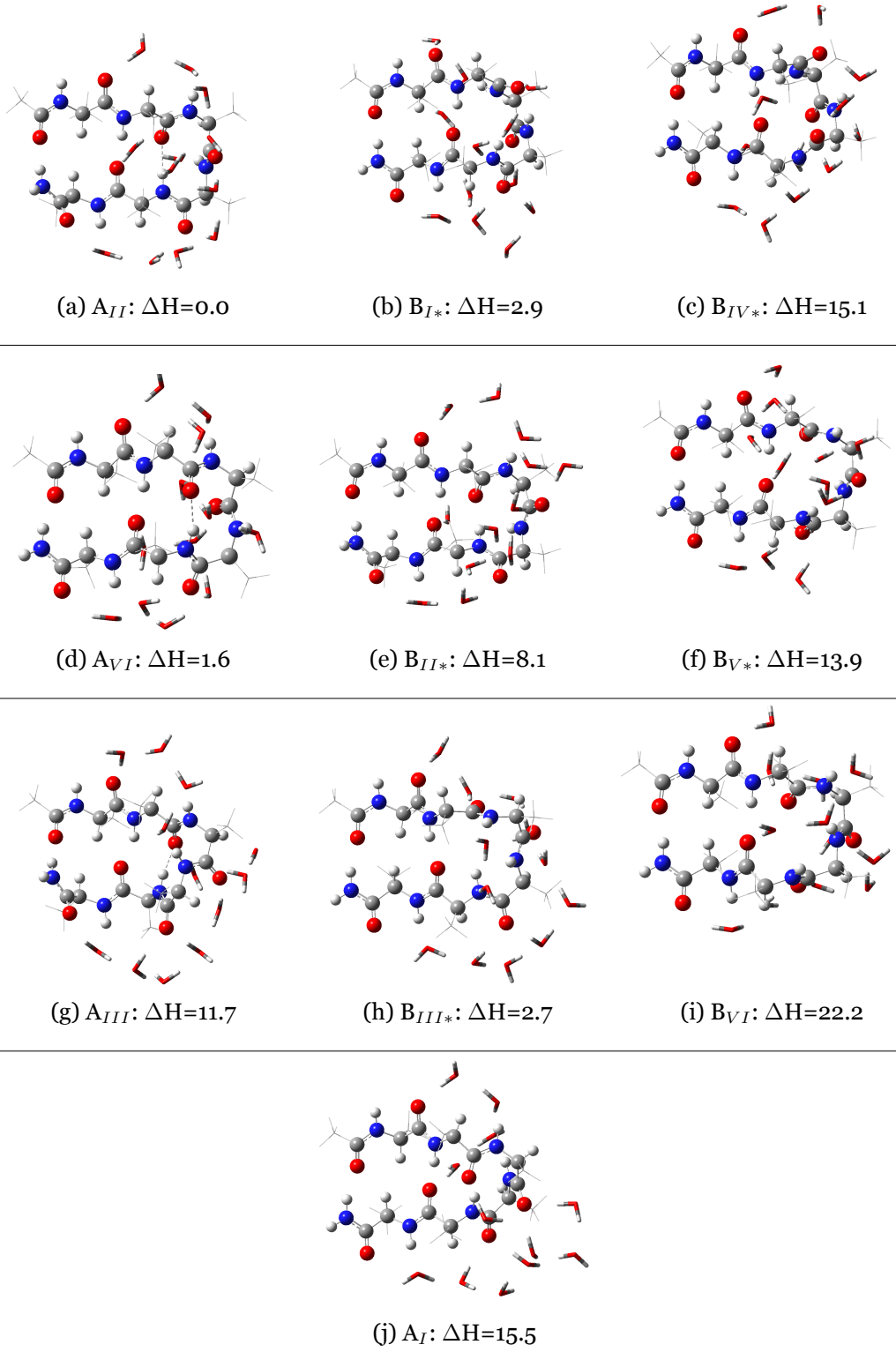
the least stable turns decreases from 8.5 to 3.7 kcal/mol. The B_I structure is followed by B_{III*} which experiences some changes in CPCM: the C_7 HB is broken and two amides were exposed to the solvent, pointing in the same direction, which lowers the MUE for type I turn to 17 (despite the C_{10} HB specified in the definition of Type I turn is not present). On the other hand, the A_{III} turn which correlates best with the Type I in the gas phase converges to the B-kind of the structure which is 3.5 kcal/mole above the B_I minimum. B_{III*} is followed by the A_I (1.3 kcal/mol), A_V (1.7 kcal/mol), and A_{II} (1.8 kcal/mol) turns which are not experiencing significant changes. Next is the newly identified B_{VI} structure that does not contain any HB and has all amides exposed to the solvent in the same direction, with the C=O pointing below the plane of the turn. Despite that the B_{VI} turn does not contain any stabilizing interaction between the adjacent peptide groups, it is only 1.9 kcal/mol above the lowest B_I conformer. The following structure, the B_{II*} turn, in which C_7 HB is broken, was found to be 2.2 kcal/mol above B_I. Finally, the B_{IV*} and B_V turns, which were the second and the fourth most stable structures in the gas phase, became the least stable turns (I disregarded the unspecified B structure resulted from A_{III} optimization); and the A_{IV} turn converged to the A_I type. For comparison I added the PP2 strand which is supposed to be predominant alanine structure in water [114]. PP2 is more stable than FES in CPCM/Pauling

model (which is not true for UFF and UA0 radii) but it is predicted to be less stable than most of the turns what confirms inadequacy of the solvent model to correctly mimic the aqueous environment. Conformational search for turn that would correlate with the type II turn did not find a stable minimum.

Summarizing, the CPCM predicts the B-type turns to be the most stable, followed by the two A turns which are 1.3 and 1.8 kcal/mol above the lowest lying conformer. Interestingly, none of this structures resemble the standard type I turn which is supposed to be predominant for the L amino acids, while the closest B_{III*} turn does not have C_{10} (or any other) HB.

In table 4.4 I present results using another common approach to the solvation problem: calculations using explicit waters. As one can see, substitution of the continuous medium by water molecules qualitatively changes the results (All hydrated structures are shown in figure 4.3). The B_I structure is not the most stable turn any more, and two A-type structures becomes two the most stable complexes. Moreover, among these two A-type turns is a newly identified A_{VI} turn which correlates best with type II β-turn, while in the gas phase and in the CPCM the search for the stable minimum around dihedral angles defined for the type II always converged to the B_{II} turn which contains the C_7 instead of C_{10} HB, but has the smallest MUE's for the type II turn.

The most stable structure, A_{II}, is followed by A_{VI} turn which is 1.6 kcal/mol less stable than the A_{II} minimum. The next turns are B_{III*}, which was very unstable in a gas phase but it is somehow stabilized in the solution, and B_{I*} which apparently is less effectively hydrated by the explicit waters than the A turns. They are 2.7 and 2.9 kcal/mol above the A_{II} minimum respectively. These turns are followed by B_{II*} which is 8.1 kcal less stable, and the remaining turns are more than 10 kcal/mol less stable than A_{II} minimum. In two of them, B_{IV*} and B_{V*} which contained two stabilizing C_7 HB's in the gas phase, water inserts into those bonds breaking them and forming a bridge between two polar groups. However, the overall effect does not provide efficient hydration and they become enthapically very unfavorable, placing them 15.1 and 13.9 kcal/mol, respectively, above the lowest enthalpy structure. Another B-type turn, B_{VI}, which has all C=O's pointing in one direction and no internal HB's was also found to be 22.2 kcal/mol less stable than A_{II}. The A_I and A_{III} turns remain in the initial geometry with C_{10} HB intact however they are significantly less stable (by 11.7 and 15.5 kcal/mol) than the most stable hydrated turn. On the other hand, A_{IV} and A_V turns converged to A_I and A_{II} turns respectively since the water inserted into the C_7 HB



(a) A_{II}: ΔH=0.0 (b) B_{I*}: ΔH=2.9 (c) B_{IV*}: ΔH=15.1

(d) A_{VI}: ΔH=1.6 (e) B_{II*}: ΔH=8.1 (f) B_{V*}: ΔH=13.9

(g) A_{III}: ΔH=11.7 (h) B_{III*}: ΔH=2.7 (i) B_{VI}: ΔH=22.2

(j) A_{I}: ΔH=15.5

Figure 4.3: Explicitly hydrated β-turns and their relative enthalpies (in kcal/mol) collected in groups with the lowest MUE's with respect to the: Type II', Type II, Type I and Type I' turns, respectively (see table 4.3). The waters are shown using tubes and methyl side chains using wireframe, the C_{10} HB is shown as a dashed line in the A-structures.

at their termini, leaving only the C_{10} HB.

It is important to point out that some of the A and B turns are somehow coupled with each other. In many cases during optimization of the hydrated B-type of turns I observed that structure converged to A-type restoring the C_{10} HB, despite that the initial structure either contained C_7 HB or had a water molecule bridging this interaction; the B_{II*} turn always converged to the A_{VI} structure and B_{I*} to A_{II}. As it can be seen from the table 4.3 (and from MUE's for the explicitly hydrated turns which are not reported here), these pair of turns always correlate best with the same β-turn which suggest that they can potentially easily hop between the minima. This is even more pronounced in vacuum and CPCM where there is no barrier between the putative A_{VI} and B_{II} turns and the former turn always converge to the latter minimum. However, both in the vacuum and the CPCM, the B turn in these pairs have lower energy, whereas with the explicit solvent this equilibrium changes towards the C_{10} hydrogen bonded conformers.

The A_{II} turn was found to be the most stable structure among the three standard β-turns identified in a gas phase and be the one in the most stable β-sheets [99] which suggest its intrinsic stability. The terminal hydration in the A turns recall the cooperative hydrogen bonding in the β-sheets which can potentially make the C_{10} HB stronger whereas the non-bonding peptide can be efficiently hydrated by the remaining water cluster which stabilizes the turn in the "neck" region on the Ramachandran plot [115]. On the other hand, the B turns, that were found more stable in the gas phase than their A type pairs, are less efficiently hydrated. The exposed peptide group in alternating conformation cannot accommodate waters as efficiently as more distant polar groups in the A type. This observation correlates with the non-additive properties of polyamides [107] which probably are not described by the CPCM very well.

Last but not least, I would like to point out that A_{III} turn, the expected turn for the LL peptides, was predicted to be 11.7 kcal/mol less stable than A_{II} conformer, which is even more than calculated for the peptides in the gas phase and the CPCM. This subject requires further careful studies of the protein high-resolution X-ray structures.

Conclusion

I proposed the new classification of β-turns based upon the existence (or not) of a C_{10} HB between the residues i and i+3 in the turn. I found that the most stable conformation in gas phase and in

CPCM model contain the C_7 HB between the i and i+2 residues (non-bonding peptide group) in place of the relatively weak C_{10} HB. The traditional classification into the I, I', II and II' turns, based on hard-sphere modeling, did not agree with the high level DFT calculations. The type I (or the closest A_{III}) turn which was predicted to be a native turn for L-alanines was found 7.2 kcal/mol less stable than the B_I turn and 3.4 kcal/mol less stable than the type II' (A_{II}) turn which was formed despite having only L amino acids. Moreover, the differences in the energies between different stereoisomers were found to be small and other effects like proper solvation or interactions with the environment might play more important role in the determination of the turns' conformation. The study of the turns' hydration using explicit water molecules showed that hydrated turns restore their broken C_{10} HB. The hydration energy gain of the exposed peptide groups in the A-type of turns overcomes the intrinsic stability of the C_7 hydrogen bonded B-turns, which constrains the available space around the polar groups lowering the overall solubility of the turn. However, the A_{III} still lies above (11.7 kcal/mol) the A_{II} conformer.

4.4 Aqueous Solvation of α-helix

Introduction

In 1989 Baldwin [116] first reported solubilized α-helices to be stable in an aqueous solution. Since that time, his group and a group of Kemp have reported quantitative data on the stability of the α-helical polyalanine as the function of the peptide length [117]. What makes an α-helix such an unique system it is the internal network of the highly cooperative hydrogen bonds (HB) which establishes the foregoing property – the unusual stability of the α-helices in the water environment. Our group [57], [58], [118], [119] and others [120], [121], [122], have studied the α-helix using DFT methods in a gas phase, while other groups[123], [124], [106] used molecular dynamics to investigate the secondary motif and its interactions with water.

The latter approach uses empirical force fields both for water and the peptide which are parameterized to reproduce the bulk properties of water and structural parameters of the peptides. Literature about both tools has been published elsewhere (for some reviews see [125], [126]) and I will comment only on the water force fields. In general they do not include the dispersion nor polarization effects, for example the very popular TIP3P model [33] approximates the water

molecule as the three point charges combined with the Lennard-Jones potential (r^{-6} attractive and r^{-12} repulsive terms) where each parameter is set to reproduce reasonable geometries and energetic of gas phase complexes of waters and water – alcohol, and liquid water. Nonetheless, the missing polarizable effects does not allow to examine if interactions between waters, or a solute and water, are calculated in the physically correct manner. In essence I argue that solvation of the α-helix is the highly cooperative process and it requires accounting for the polarization effects due to the redistribution of the electron density during the hydration process.

In 1994 Karplus demonstrated the existence of the cooperative hydration effect for amidic HB's [127], similar to that found in peptides [128] and molecular crystals [129]. Two water molecules placed on the opposite sides of the amide group in such a way that one is acting as the HB acceptor (N-H $\longleftrightarrow$ O_w) and second as the HB donor (O_w-H $\longleftrightarrow$ O) experience extra stabilization due to mutual polarization of amidic group. As the result, the first water increases the propensity of binding another one which implies that the hydration energies are not additive. Now, if such cooperativity would be applied to the α-helices, which consist of three chains of the H-bonded amides along the helical axis, one might expect that the water molecules hydrating unsatisfied amidic groups at the termini of the helix would experience such additional stabilization.

However, to fully take an advantage of the foregoing effect, waters should align parallel with the helical dipole moment causing the increase of the entire aggregate's dipole moment over that of the isolated helix. Thus, this specific interactions will be different from the expectation that the solvation of a dipolar molecule (for instance - the α-helix) would involve ordering of the medium with high dielectric constant (for instance - water) so that the medium would create an opposing dipole moment. The mechanism was originally suggested by Onsager [130] in his seminal work about the solvation of polar species.

In this section I will try to answer the question what kind of solvation of α-helix occurs in the aqueous solution. I combine the explicit solvation by discrete waters at DFT level of theory with the implicit CPCM [42] and SM5.2 [110] solvent models for the bulk water.

Methods

The calculations at DFT level of theory were performed using Gaussian 09 suite of programs [50]. X3LYP functional in d95(d,p) basis set was chosen due to its excellent performance for the water

dimer and larger water clusters ([56], ref [131]). Pre-optimized α-helices from Wieczorek et al [57] were used as initial geometries, and the complexes were reoptimized with the presence of the explicit waters or/and continuous solvent model. Multiple optimizations for several water clusters were performed in order to localize the most stable minima. The vibrational frequencies were computed to confirm the stationary points on the potential energy surface (PES) and to evaluate the enthalpies at 298K. All frequencies were real for the structures discussed here indicating that these are true minima. The counterpoise (CP) correction [132] was incorporated using a single point calculation on non-CP optimized surface rather than performing the full optimization, due to the excessive number of subsystems (up to 18) necessary to be included which would require inaccessible CPU time. In the CP single point calculations all waters except the ensemble of the non-interacting between themselves backbone waters were treated separately.

I use two solvent models to simulate the bulk water: (1) the Conductor-like Polarizable Solvent Model (CPCM, [42]) with the default UFF [40] radii set for the definition of the cavity, and (2) SM5.2 [110] version of the AMSOL solvation model applied to the DFT-CPCM optimized structures using the AM1 semi-empirical method as coded in Ampac 8.15 program package [133]. I used the default UFF radii since it predicted reasonable free energy of transfer for water and by this time I was not aware that the cavity definition influences the calculated results. The second approach was used in the previous study of the α-helix solvation[69]. Both solvent models are constructed to estimate the free energy of solvation rather than the enthalpy or energy which are calculated using specific waters. Also, as I argued in the previous section, the data of good quality for free energy of solvation of polyamides is not available, therefore I cannot evaluate the performance of these methods directly. In addition I decided to use ΔH's instead of ΔE's for calculating the free energy of transfer of the α-helix, since the vibrational correction for molecule this size might not be negligible and the training sets consist mainly small molecules [45], [41].

Results

The Ac$(Ala)_{15}$NHMe α-helix was chosen for this study. It can be divided into three regions:

1. C-termini (C) when the helix acts as the HB acceptor
2. N-termini (N) when the helix acts as the HB donor
3. Backbone (B) as the region in-between two termini.

Each of these regions has different properties and thus, will be treated separately. I also break down my discussion into three parts: (1) solvation by a single water, (2) solvation by small water clusters with different spacial distributions, and (3) solvation by continuous solvent models with and without the explicit first hydration shell.

Single Waters

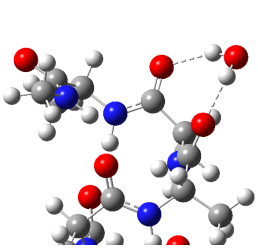
Table 4.5 and figure 4.4 present data on solvation by individual waters. Figure 4.4 shows all stable minima for the single waters and small water clusters complexes. Table 4.5 collects all data on the enthalpies, dipole moments and energies with the breakdown to the contributions from the pure interaction between species in geometry of the complex and the distortion energy.

As I mentioned in the introduction, one can relate the solvation of the α-helix to the solvation of the N-methylacetamide (NMA). Similarly to the simplest protein building block, water molecules prefer to bind to the C-termini (unsatisfied C=O groups) over the N-termini (N-H fragments). However, in the case of the α-helix a single water prefers to form the HB with few C=O's due to the proximity of those groups (please see figure 4.4). The strongest interaction occurs when the water forms three HB (two O_w-H --- O and one N-H --- O_w from the broken helical HB) in the complex with the helical C-termini which has interaction energy of -11.1 kcal/mol, the hydrogen bonding energy -15.1 kcal/mol and large distortion energy 3.9 kcal/mol due to the broken HB. Two other positions where water forms only two HB's with carbonyls have interaction energies -9.3 and -10.1 kcal/mol. Upon the vibrational correction, the enthalpies of those waters become -9.1, -7.4 and -8.0 kcal/mole respectively which is 50%-62% more than -5.6 kcal/mol reported for solvation of NMA for B3LYP/d95(d,p) level of theory[134]. The complexes formed at the N-termini are less stable, however estimated enthalpies, -7.7 kcal/mol and -5.8 kcal/mol, are again lower than -2.9 kcal/mol reported for NMA.

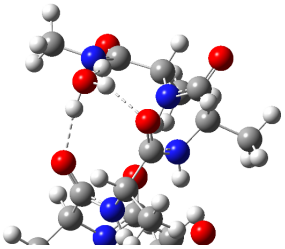
I found three types of structures formed by a backbone water which are shown in figure 4.4 and their energies are also reported in table 4.5. As one can see, the "planar" complex is the most stable due to the distortion energy which is lower than that in the "bridged" structure. For this reason I use planar water for the further analysis of the α-helix hydration. Moreover, energies and enthalpies of the planar backbone hydration vary from -2.8 to -5.5 kcal/mol and correlate with the position of the HB. In essence, if the water is closer to the C-termini, the chain of HB's it interacts

Table 4.5: Counterpoise corrected energies of discrete water hydration (in kcal/mol) and dipole moments (Debyes)

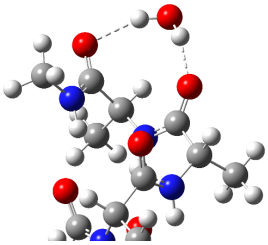
Position	ΔE	ΔE	ΔE_{frozen}	$\Delta E_{Distortion}$			ΔH	μ
	1 water	cluster		helix	water	cluster		
C-termini								
A	-9.3		-11.6	2.2	0.1		-7.4	74.6
B	-11.1		-15.1	3.9	0.1		-9.1	70.0
C	-10.1		-10.8	0.6	0.1		-8.0	72.9
N-termini								
D	-9.4		-9.6	0.2	0.0		-7.7	73.5
E	-7.4		-8.5	1.0	0.0		-5.8	73.0
Backbone								
planar	-6.4		-7.2	0.6	0.1		-4.7	68.5
perpendicular	-4.6		-5.1	0.4	0.1		-3.2	67.9
bridged	-2.5		-7.4	4.7	0.1		-0.4	69.7
1	-4.3		-5.2	0.8	0.1		-2.8	68.5
2	-5.8		-6.6	0.7	0.1		-4.3	68.1
3	-6.5		-7.3	0.7	0.1		-4.9	68.0
4	converges to 3-10 helix							
5	-6.3		-7.0	0.6	0.1		-4.7	68.2
6	-6.5		-7.1	0.5	0.1		-4.7	68.2
7	-6.4		-7.2	0.6	0.1		-4.7	68.5
8	-6.6		-7.3	0.6	0.1		-4.9	68.5
9	-6.7		-7.4	0.5	0.1		-4.9	68.5
10	-6.9		-7.8	0.7	0.1		-5.1	68.5
11	-7.4		-8.3	0.7	0.1		-5.5	69.1
12	-6.7		-7.7	0.9	0.1		-4.9	68.9
Clusters								
4C	-47.2	-15.7	-21.0	2.7	-28.8	2.7	-38.8	79.2
6C	-64.2	-7.8	-40.3	6.0	-29.9	26.5	-41.7	86.6
11C	-131.2	-7.3	-58.8	6.7	-79.2	44.7	-108.5	85.4
3N	-30.9	-11.8	-20.4	1.3	-11.9	7.3	-25.7	80.1
6N	-64.3	-7.8	-25.1	3.1	-42.3	14.1	-38.8	72.7
4C+3N	-78.4	-27.7	-40.8	3.1	-40.6	10.1	-64.8	89.5
6C+3N	-95.8	-20.1	-61.0	7.3	-42.1	33.6	-77.7	97.0
11C+6N	-196.8	-16.4	-85.4	9.8	-121.2	59.1	-163.1	88.9
4C+3N+11B	-144.4	-93.7	-112.8	8.7	-40.3	10.4	-112.7	73.0
11B	helix hydrated in 4C+3N+11B geometry							56.8



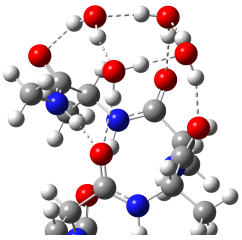
(a) A: ΔH=-7.4



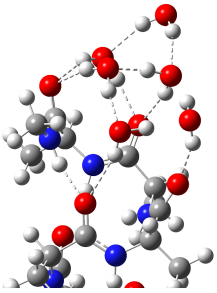
(b) B: ΔH=-9.1



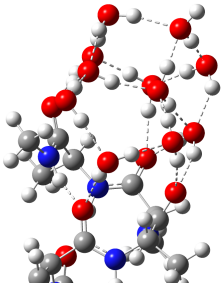
(c) C: ΔH=-8.0



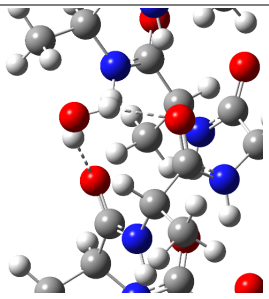
(d) 4C: ΔH=-38.8



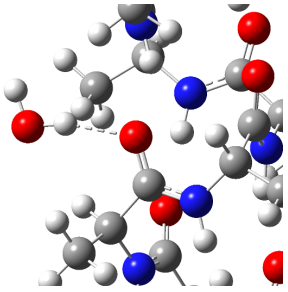
(e) 6C: ΔH=-41.7



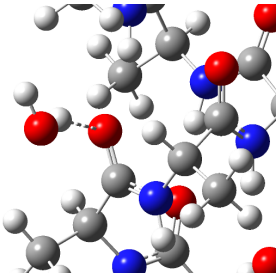
(f) 11C: ΔH=-108.5



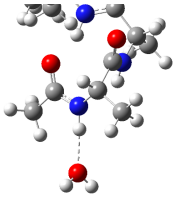
(g) bridged: ΔH=-0.4



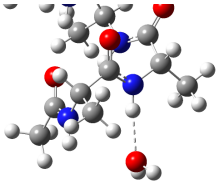
(h) planar: ΔH=-4.7



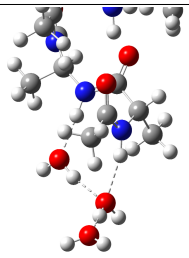
(i) perpendicular: ΔH=-3.2



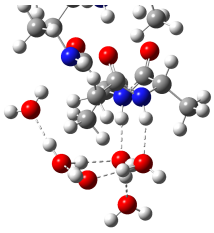
(j) D: ΔH=-7.7



(k) E: ΔH=-5.8



(l) 3N: ΔH=-25.7



(m) 6N: ΔH=-38.8

Figure 4.4: Optimized structures and enthalpies (in kcal/mol) of the solvated α-helix.

with (defined as the number of HB's in the chain below the hydrated peptide group) increases, which affects the cooperative effect imposed on the HB. Nonetheless, the predicted enthalpy is larger than -5.8 kcal/mol as calculated for the single water-water HB in the bulk water, what should be expected from the hydrophobic surface.

It is interesting to point out that waters at the termini and the backbone behave differently. As can be seen from the table 4.5, the solvation of the termini by a single water leads to the increase of the hydrate's dipole moment more than the sum of isolated molecules (with the exception for B-water which breaks one helical HB). Thus, those waters participate in cooperative HB network in the manner originally suggested by Karplus [127]. However, the backbone waters tend to decrease the overall dipole moment by assuming the antiparallel geometry, which on the other hand agrees with the Onsager's definition of the dielectric medium. It is clear that in order to investigate the problem in more detail I have to go beyond the single water solvation.

Small Water Clusters

In case of the small water clusters one can expect two competing processes – each individual water can interact with either the helix directly or with another water molecules. The lowest energy structure will contain a balance of these two types of interactions. To calculate the interaction energy between the water cluster and the α-helix I reoptimized the most stable cluster from the work of Shields et al [135] using X3LYP functional and the CP correction. Table 4.5 contains all interaction, hydrogen bonding, and distortion energies where the last is defined as the penalty required to rearrange the water cluster from the global minima to the geometry assumed in the complex.

Negative energies for all cluster sizes indicate that the formation of the water adducts is preferable. What is more interesting is the behavior of the dipole moment – it is significantly larger for hydrates than for the isolated α-helix. For instance, addition of four waters to the C-termini increases the dipole moment from 69.8 to 79.2D, or by 9.4D, which is just 0.2D less than a sum of the dipole moments of the isolated molecules. The N-terminus experience even stronger effect since three waters increase the dipole moment by 10.4D, which is 3.8D more than the simple sum. However, the simultaneous addition of two water clusters is effectively non-cooperative process since the interaction energies and the dipole moments of the doubly capped helix are the sum of

the two independent contributions. On the other hand, adding 11 backbone waters decreases the molecular dipole by 13.0 D. Thus, the small water clusters align with the dipole of the α-helix due to cooperatively enhanced HB's, but the backbone waters align opposite to the helical dipole, as would be expected from the dielectric medium. Increasing the water cluster at the termini leads to the decrease of the dipole moment, suggesting that the bigger clusters start to incorporate some features of the bulk water.

Continous Solvent Models

As the last step I investigated the effects of adding the continuous solvation model with and without discrete water molecules. The CPCM predicts increase of the dipole moment from 69.8 to 84.2D and calculates the free energy of transfer from gas phase to the CPCM medium (representing water) to be -55.5 kcal/mol which is substantially less than calculated energy of solvation by water clusters. However, we should keep in mind that we compare two different processes and quantities, the continuous models calculate the free energy of transfer while explicit waters gives enthalpies of binding of water clusters in the gas phase.

If the CPCM was the perfect water model, moving a water molecule from the bulk to the first solvation should not cost/provide any energy and the free energy of transfer of any water, regardless if bonded in the first hydration shell or in the bulk, should give us ΔG_{solv}=-4.8 kcal/mol (the value calculated for a single water in X3LYP/d95(d,p) for CPCM-UFF model).

Tables 4.5 and 4.6 present that the overall solvation energy becomes lower for the complexes with the explicit waters. However, as I stated in the previous paragraph, one should expect extra gain of -4.8 kcal/mol for the each water transferred to the CPCM. The lower part of the figure 4.6 shows that the explicit waters provide -40.2/7= -5.7 kcal/mol extra stabilization per water when moved from bulk to the first solvation shell at the termini, and (-55.9+40.2)/11= -1.4 kcal/mol when moved to the backbone. Thus, the calculations predict that the waters hydrating the ends of the helix and the backbone are in the energetically more favorable positions than those in the bulk water, and that those hydrating the termini provide significantly more stabilization.

On the other hand, the results of the SM5.2 using the AM1/AMSOL procedure differ substantially from those calculated by the CPCM. The SM5.2 predicts ΔG_{water} to be -6.3kcal/mol which agrees with the experimental value, although some explicit parameters were used to achieve the

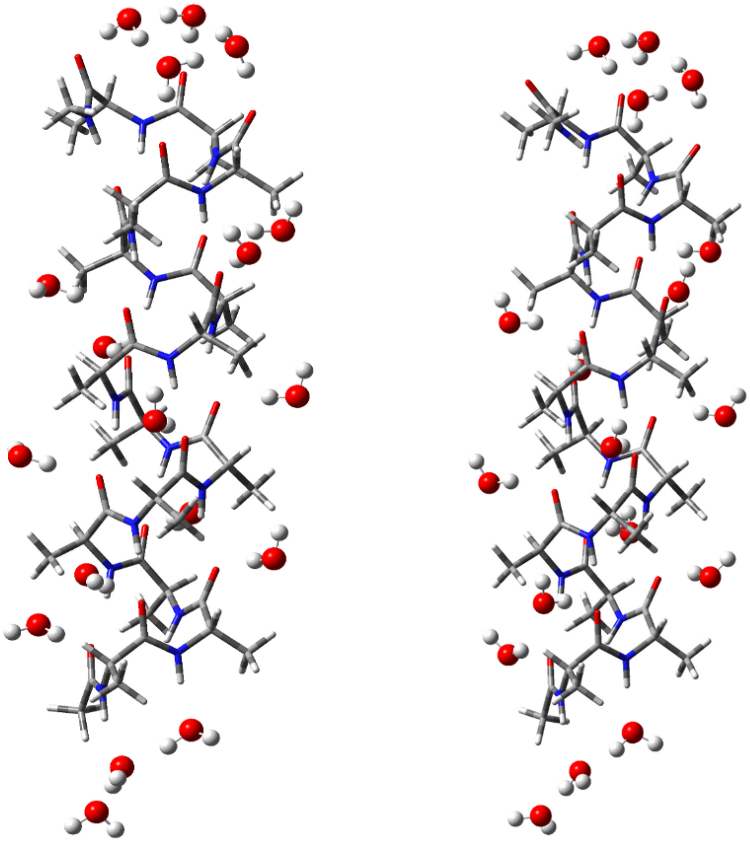


Figure 4.5: Optimized α-helix with 18 waters: left - CPCM aqueous phase; right - gas phase

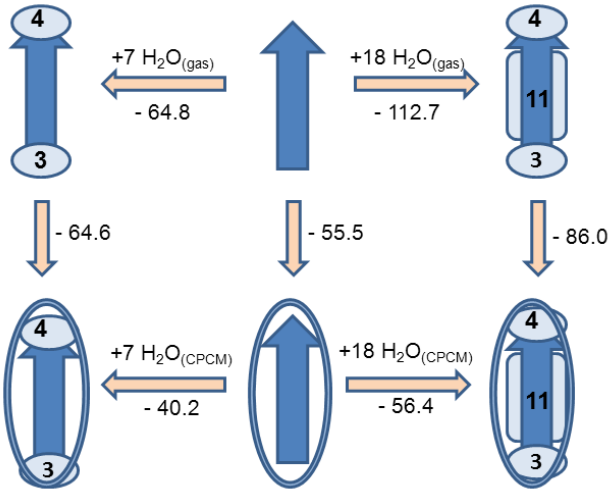


Figure 4.6: Thermodynamic properties of the hydrated α-helix

Table 4.6: Enthalpies of Hydration by Single Water in Gas and Solvent Models, Free Energies of Tranfer of Different Species (in kcal/mol) and Dipole Moments (Debyes)

	ΔH_{inter}			$\Sigma\Delta G_{solv}$		ΔG_{solv}(complex)		μ
Species	Gas	CPCM	SM5.2	CPCM	SM5.2	CPCM	SM5.2	CPCM
α-helix				-55.5	-91.7			
Single Water				-4.8	-6.3			
Water Dimer	-4.3	-3.3	+9.1	-9.6	-12.6	-7.8	+0.7	
50 Water Cluster				-240.0	-315.0	-43.3	+2206.2	
1C	-8.0	-4.6	-2.7	-60.3	-98.0	-56.8	-92.7	88.1
1N	-7.7	-3.9	-4.1	-60.3	-98.0	-56.4	-94.4	87.5
1B	-4.7	-2.2	+3.5	-60.3	-98.0	-57.7	-89.8	84.5
4C+3N	-64.8	-40.2	+131.5	-89.2	-136.0	-64.6	+60.4	104.8
6C+3N	-112.7	-55.9	+207.2	-142.3	-205.5	-85.5	+114.4	105.2
11C+6N	-77.7	-48.5	+147.3	-98.8	-148.6	-69.7	+76.4	111.9
4C+3N+11B	-163.2	-100.8	+381.5	-137.4	-199.2	-75.1	+345.4	103.6

agreement [110]. The solvation energy of the bare α-helix predicted by SM5.2 is -36.5 kcal/mol more negative than than that predicted by CPCM. However, the water clusters solvated by SM5.2 are unstable, with ΔG=+0.7 kcal/mol for the water dimer and 2206 kcal/mol for 50 waters (see table 4.6). The latter cluster has ΔS=-148 cal/degree per mol per water which is far beyond the entropy of a water in the liquid water (estimated to be between -10 to -17 kcal/mol in 298K [86]).

From the above I do not find the SM5.2 results for water clusters at the helix reliable. On the other hand, the solvation of the single water seems to give plausible results. The C- and N-termini hydrated helices are 2.7 and 4.1 kcal/mol stabilized more, while the backbone position is 3.5 kcal/mol less, than the sum of the individual free energies of water and the helix.

Discussion

We can divide the individually hydrated complexes into two classes:

1. Complexes with terminal water which contain the cooperative HB's between the water and the helix in the manner similar to that presented by Karplus [127]. These waters increase the dipole moment of the hydrate above the sum of the dipoles of the individual molecules and experience stronger stabilization than those in the complexes with NMA [134].
2. Complexes with the backbone water which lower the overall dipole moment (from 0.9 to 1.9D) which is consistent with expected behavior of the bulk dielectric medium to oppose

solute's dipole moment as suggested by Onsager [130]. These waters are also bound more weakly and probably experience more dynamics than the terminal waters.

In more general view, both continuous solvation models (CSM's: CPCM and SM5.2), and the explicit terminal water clusters increase the dipole moment of the solute (CSM's) and the aggregate over that calculated for the gas phase. However, increasing size of the cluster has a diminishing effect, which is apparent from the transition from 6C to 11C or 3N to 6N. In both cases the molecular dipole moment decreases for bigger clusters which suggest that water clusters experience more bulk-like properties.

The second group – backbone waters – decrease the overall dipole moment of the solute in a cumulative way. As shown in table 4.5, 11 explicit backbone waters (as a single point at the 4C+3N+11B geometry, because the optimization with only 11 backbone waters disturbs the helix) lower the dipole moment by 10.8D for the bare helix and by 16.5D for the helix hydrated at the end. Likewise, the addition of 11 backbone waters to 4C+3N helix in the CPCM solvent decrease the overall dipole moment only by 0.4D. However, this behavior becomes understandable if one thinks of the CPCM as an infinite reservoir of the opposing dipole moment. In this case it is energetically more favorable to rotate backbone waters (as shown in figure 4.5) to increase the solute's (helix+18 waters) dipole moment which is the source of interaction with the dielectric medium.

The table 4.7 presents the length of HB's within the helix in the gas phase and its hydrated forms. The discrete waters placed at the termini significantly decrease the length of HB's, especially close to the ends. However, the additional 11 backbone waters increase the length of the HB's to the value comparable with the bare helix in the gas phase (except for terminal HB's). Placing the helix in PCM leads to even larger decrease in HB's length than just by water clusters and the explicitly added waters affect only the terminal HB's. Nonetheless the addition of the backbone waters increase the length to that observed for the helix in the gas phase. The foregoing suggest that terminal hydration of the α-helix strengthen the helical HB's due to cooperative effect, similarly to that in formamide chains [128] or in hydration of NMA [134].

Since waters solvating ends of α-helix play very different roles in stabilizing the structure from those near the backbone, one might expect that the different positions in the helix would be stabilized by different kinds of hydration. Crystal structures of proteins that contain α-helices commonly contain ordered waters near the ends of the helices. Some examples are the following

Table 4.7: Length of Helical Hydrogen Bonds (Å) in Hydrated Helices (Numbering from the N-End)

	Gas Phase			CPCM		
H-Bonds	Helix	4C+3N	4C+3N+11B	Helix	4C+3N	4C+3N+11B
1	2.498	2.142	2.106	2.044	1.965	1.964
2	2.057	1.965	1.988	1.926	1.906	1.990
3	2.083	2.066	2.089	2.021	1.999	2.039
4	2.045	1.994	2.040	1.943	1.934	2.018
5	2.026	1.998	2.033	1.947	1.937	1.982
6	2.017	1.986	2.019	1.936	1.931	2.032
7	2.026	1.987	2.032	1.954	1.935	2.004
8	2.021	1.969	2.030	1.934	1.920	2.001
9	2.024	1.998	2.051	1.944	1.923	2.035
10	2.051	1.972	2.031	1.945	1.924	2.075
11	2.178	2.098	2.382	2.007	1.922	2.099
12	2.007	1.927	1.962	1.937	2.000	2.013
13	2.221	2.048	2.085	1.961		

pdb structures: 1AKI (Hen eggwhite lysozyme)[136] 3G4P (OXA-24 β-lactamase) [137] and 1Z98 (Aquaporin)[138] the last of which is a trans-membrane protein. While trans-membrane proteins need to have hydrophobic residues that interact with the lipids in the membrane, favorable interaction with aqueous phases on either side of the membrane must also be important. One such protein, the trans-membrane domain of virus protein “u” (Vpu) from HIV-1, has been studied by NMR in a lipid bilayer[139], and reported to have its hydrophobic backbones traversing the lipid and its hydrophillic termini in the aqueous medium on either side of the bilayer, consistent with the present report of solvation by discrete waters.

Conclusions

I present the data that suggest that solvation of the alanine α-helix occurs via two very different mechanisms: first, water hydrates the termini through cooperative HB’s as suggested by Karplus [127], and then the aggregate interacts with the bulk water as it is suggested by Onsager in his solvent reaction field theory (also implemented in CPCM)[130].

In essence the individual waters and the small water clusters placed at the termini tend to increase overall dipole moment more than the simple sum of entities involved in formation of the molecular complex. However waters that hydrate the helix backbone, which are less stable than

those at the termini, decrease the overall dipole moment. The shortening of helical HB's caused by terminal hydration by explicit waters and CPCM suggest that helical bonds become stronger upon aqueous solvation.

4.5 Aqueous Solvation of β-sheets

Introduction

The β-sheets, widely abundant motif, might serve many purposes – be a hydrophobic surface deeply buried in the protein core or large, solvent exposed β-sheets, like big tertiary motif of transmembrane β-barrels (composed of parallel sheets, [140]) or (mostly) antiparallel silk protein with many alanines and serines on a sheet's surface [84]. Moreover, the amyloids driven diseases, like Alzheimer's [141] and Parkinson's are thought to be caused by the precipitation and the aggregation of parallel sheets from the (physiological) aqueous solution. There have been many theoretical studies aiming the stability of the β-sheets, few of them published by Dannenberg's group using DFT theory [142], [143]. Although so many theoretical studies were performed on the β-sheets in the gas phase, only little is known about the influence of aqueous solution on their structural properties and even the magnitude of cooperative hydrogen bonding (or if it exists at all) similar to that observed for helices ([79], see section 4.4). It is not understood if the observed twist in the β-sheets (please look into table 3.11) is their intrinsic property (as observed in the ONIOM calculations or in crystal structures of globular proteins) or the effect of the environment. Ireta[144] in his paper used DFT to investigate the influence of the water clusters on the twist of two-stranded β-sheets and concluded that water is responsible for the observed change in the dihedral angle, while others used explicit waters to investigate the stability of smaller motives [145].

Methods

All calculations were performed in Gaussian 09 computer programs package [50] using X3LYP[54] functional with d95(d,p) basis set. The six alanine capped $Ac(Ala)_6NHMe$ peptide was used as the sheet's building block. I investigate up to six strands in both parallel and antiparallel conformations in the gas phase and the continuous solvent. I also construct the first hydration shell using up

to 12 water molecules to properly hydrate sheets' termini. Calculations with explicit waters were performed both in the gas phase and with the bulk solvation using sheets with up to four strands. I chose CPCM[42] as implemented in Gaussian 09 [50], with the defaulft UFF[40] and the Pauling[38] radii for the definition of the cavity around the solute. First model uses radii from the common force field universally parameterized for all atoms, while the latter uses van der Waals radii which, as suggested in the literature [32], are better suited for calculation in polarizable medium (for more details please see section 2.3 and discussion in 4.3). Preliminary calculations on the free energy of transfer of trans-N-methylacetamide (ΔG_{exp}=-10.1 kcal/mol, the only relevant reference point [105] for amides), showed the best agreement for Bondi radii model (estimated free energy -10.2 kcal/mol), followed by Pauling model (-11.5 kcal/mol). However, it is not clear if the Ben-Naim's correction [146] of 1.89 kcal/mol is included in Gaussian 09 algorithm for calculating the free energy of a solute in the continuous medium, thus I decided to choose Pauling model which overestimates the free energy but provides the better fit if the correction is not included.

Harmonic vibrations were calculated in order to confirm the existence of the local minima (all frequencies were positive) and to evaluate the enthalpies at 298 K. I use enthalpies instead of energies that are recommended for use with continuous solvent[147] to calculate the free energy of transfer since the solvent models does not contain larger molecules in their training sets and the vibrational correction might influence calculated ΔG's. Some of the water – sheets complexes in the CPCM needed to be reoptimized using denser grid to remove low lying imaginary frequencies, similar to those observed in section 4.3.

The counterpoise (CP) correction was added as a single point on the non-CP optimized potential energy surface (PES) due to the excessive time necessary to fully reoptimize the complex on the CP-corrected surface. The basis set superposition error (BSSE) was divided into three parts: the BSSE within the sheet, between the sheet and water and within the water clusters, and added as a one ad posteriori correction.

Results and Discussion

The results for calculations on formation of the bare β-sheets and energies of hydration by explicit waters in the gas phase and the CPCM are collected in a table 4.8, while the sheet formation enthalpies corrected by explicit water molecules are collected in table 4.9. Each table corresponds

to specific hydration cycle (similar for those reported in [37]) showed in the figure 4.9. For calculations of ΔH_{solv} I use ΔH_{gas} plus a correction for a change in the solubility ($\Delta\Delta G$, the difference between solvation free energies between products and the substrates) as showed in the equations in the relevant places in the tables.

First, let me discuss the results for the simplest case of formation of the β-sheets in continuous solvent presented in table 4.8. We know that the polyalanine peptide does not aggregate in water solution but rather assume a polyproline II (PP2, [114]) or other extended strand conformations[148]. Therefore we should expect that the aggregation is not favorable and the strands would prefer to remain individually solvated in the aqueous solution.

As seen from the table 4.8, the ΔH_{gas} of adding the second strand in antiparallel coformation is -25.56 kcal/mol (or -3.65 kcal/mol per HB) and decreases with the size of the sheet as expected from slightly cooperative hydrogen bonding in the antiparallel β-sheets ([142], [143]) reaching -32.03 kcal/mol for addition of the sixth strand. We also observe that the decrease is not monotonic but oscillates between four, five and six strands, which agrees with the repeating unit of two strands for antiparallel sheets. $\Delta\Delta G$ is defined as the change in the solubility between the final sheet composed of n strands, and the single strand (in Fully Extended Strand for UFF and PP2 conformations for Pauling model) and the sheet smaller by one unit. The $\Delta\Delta G$ can be thought of the desolvation penalty associated with removing the strand from the solution and extending the sheet. However, in UFF model we observe that the term is smaller than a half of the absolute value of ΔH_{gas} and as the result the overall ΔH_{sol} (defined as Σ in table 4.8) is strongly negative which would imply the favorable (and probably irreversible) aggregation of the sheets. Although the desolvation penalty for Pauling model is much larger, up to 28.71 kcal/mol for the sixth strand, the overall enthalpy in solution is still negative, from -0.77 kcal/mol for the dimer to -3.32 kcal/mol for the sixth strand. Pauling model still predicts aggregation of the β-sheets but the overall effect is in the order of hydrogen bonding enthalpy.

The results for the parallel conformer follow the similar trend. The enthalpy of adding another stands decreases with the size of the β-sheet in the gas phase from -23.94 kcal/mol for the second strand to -28.21 kcal/mol for the sixth strand. It is worth to note that the parallel sheets in the gas phase contain two extra C_7 HB's at the edges of the sheet and a dimer contain one more such HB at the one of the termini. The UFF radii model predicts the incremental solubility change to be

Table 4.8: The enthalpy in the gas phase (ΔH_{gas}), the desolvation penalty ($\Delta\Delta G$), and their sum (Σ) for the hydrated parallel and antiparallel β-sheets using different cavity models

Size	Antiparallel					Parallel				
	ΔH_{gas}	UFF		Pauling		ΔH_{gas}	UFF		Pauling	
		$\Delta\Delta$G	Σ	$\Delta\Delta$G	Σ		$\Delta\Delta$G	Σ	$\Delta\Delta$G	Σ
Bare β-sheets										
$\Sigma = \Delta H_{gas} + \Delta G_{\beta-sheet(n)} - \Delta G_{strand} - \Delta G_{\beta-sheet(n-1)} = \Delta H_{gas} + \Delta\Delta G$										
1→ 2	-26.56	13.57	-12.99	25.78	-0.77	-23.94	14.44	-9.50	24.37	0.44
2→ 3	-30.09	14.92	-15.17	27.46	-2.63	-26.44	13.84	-12.60	25.83	-0.62
3→ 4	-31.64	15.45	-16.20	28.41	-3.23	-27.90	13.26	-14.64	26.59	-1.31
4→ 5	-31.49	15.31	-16.18	28.14	-3.36	-28.34	14.05	-14.29	26.66	-1.68
5→ 6	-32.03	15.41	-16.62	28.71	-3.32	-28.21	14.27	-13.93	26.26	-1.95
6 waters at C_5 position										
$\Sigma = \Delta H_{gas} + \Delta G_{water\sim\beta-sheet} - \Delta G_{\beta-sheet} - (6 \vee 12) \cdot \Delta G_{water} = \Delta H_{gas} + \Delta\Delta G$										
1	-44.98	20.88	-24.10	35.54	-9.45	-44.98	20.80	-24.10	35.54	-9.45
2	-52.48	24.65	-27.83	39.36	-13.11	-49.85	20.36	-29.48	36.66	-13.19
3	-52.18	23.46	-28.72	38.79	-13.39	-53.56	22.39	-31.17	39.00	-14.56
4	-52.62	23.74	-28.88	39.72	-12.89	-53.80	23.00	-30.80	39.60	-14.21
6 waters at C_7 position										
1	-52.65	22.70	-29.95	39.45	-13.20	-43.29	13.34	-29.50	30.09	-13.20
2	-55.25	24.27	-30.98	37.44	-17.81	-49.65	15.87	-33.79	32.95	-16.70
3	-52.41	21.85	-30.56	36.36	-16.05	-51.51	17.55	-33.97	34.68	-16.84
4	-53.86	24.36	-29.50	38.62	-15.25	-51.94	18.73	-33.21	34.70	-17.15
12 waters at C_5 position										
1	-114.68	51.77	-62.91	-	-	-114.68	51.77	-62.91	-	-
2	-111.78	48.45	-63.34	-	-	-113.86	46.22	-67.64	-	-
3	-109.85	47.74	-62.12	-	-	-114.24	47.20	-67.03	-	-
4	-109.07	47.57	-61.51	-	-	-114.34	48.15	-66.20	-	-
12 waters at C_7 position										
1	-109.36	47.22	-62.14	77.91	-31.45	-109.36	47.22	-62.14	77.91	-31.45
2	-114.22	49.06	-65.16	80.46	-33.76	-113.32	46.60	-66.72	78.42	-34.89
3	-115.02	49.01	-66.01	80.66	-34.36	-116.05	48.64	-67.40	80.02	-36.02
4	-115.15	48.92	-66.23	80.07	-36.08	-116.88	49.55	-67.32	81.34	-35.54
Single water on the surface										
$\Sigma = \Delta H_{gas} + \Delta G_{water\sim\beta-sheet} - \Delta G_{\beta-sheet} - \Delta G_{water} = \Delta H_{gas} + \Delta\Delta G$										
SR, O	-5.22	2.94	-2.28	4.04	-1.17	-4.57	-2.70	-1.86	3.14	-1.43
LR, O	-2.52	1.19	-1.33	1.25	-1.27					
SR, Me	-0.83	2.62	1.79	0.61*	-0.22	-1.10	0.05	-1.05	-1.02*	-2.11
LR, Me	-1.39	2.64	1.25	1.76	0.37					

less than a half of the gas enthalpy and the ΔH_{sol} to be strongly negative while the Pauling model predicts it to be positive (+0.44 kcal/mol) for the dimer, but negative (-1.95 kcal/mol) for the sixth strand due to the much larger $\Delta\Delta G$. Similarly to antiparallel sheets, the enthalpy of adding another strand in the solution decreases with the size of the sheet suggesting that larger sheets would be even more likely to accommodate extra strands and form aggregates

For both cavity models, the predicted trends would suggest the aggregation of the single strands in order to form larger suluble β-sheets, while in the real solution they become insoluble and precitipate from the solution. Since the desolvation penalty (or change in the solvation energy, $\Delta\Delta G$) is not big enough to overcome the stabilizing hydrogen bonds, they are the driving force for the observed β-sheets formation. Probably some corrections using explicit water molecules are necessary, as it is widely used in the literature [36], [34], [37].

In the next part of the table 4.8 I examine the explicit solvation of 1-4 stranded β-sheets using 6 and 12 waters in the C_5 and C_7 hydration models as showed in figure 4.7. The numbers indicate the structure of waters at the sheet's termini, constructed as a bridge between potential C_5 or C_7 HB's. In general, six waters at C_7 position are more stable than C_5 for all cases except the parallel sheets in the gas phase. The 12 waters are more stable in C_5 position for one and two strands in the parallel conformation and only one strand in the antiparallel probably due to an extra water-water HB not present in the larger aggregates or C_7 structures. The change in the solvation energy is defined as the difference between the solvation energy of the sheet hydrated by the $6\backslash 12$ explicit waters, and of the sheet and $6\backslash 12$ waters separately. As I argued in two previous sections, if the implicit solvents were perfect, there should be no energetic penalty associated with moving the waters from the bulk water to the first hydration shell. However in all cases we observe that these waters are gaining on the energy when moved to the sheet's termini indicating it is preferable for them to localize in the first hydration shell.

For the six waters hydrating the sheet in the antiparallel conformation I observe that they experience almost none of the cooperative bonding, in oppose to what I observed for the α-helices. The enthalpy of bringing the waters to the first solvation shell at the termini in the gas phase is essentially constant for all size of the sheets, and the only exception - change from one to two strands in the C_5 hydration is caused by a reorganization in the single strand because the C_5 binding is not very effective. On the other hand the parallel sheets experience little of cooperativity, as

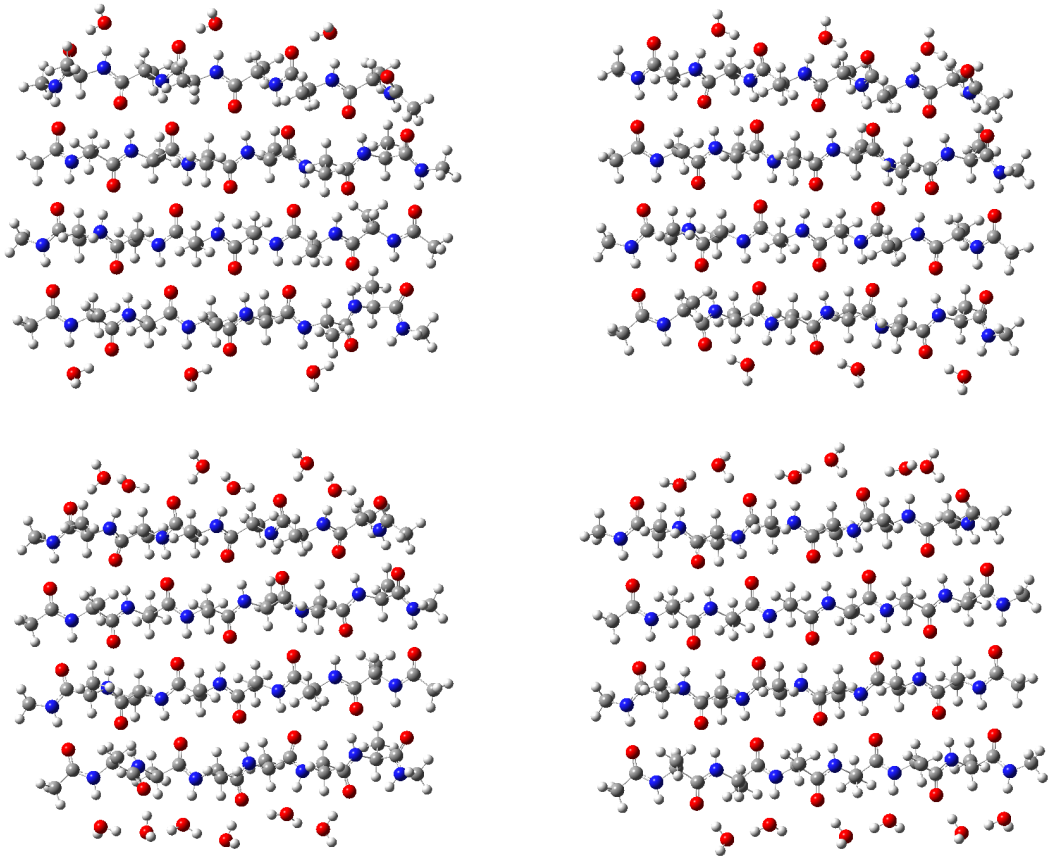


Figure 4.7: Different models ofβ-sheet hydration by discrete water molecules

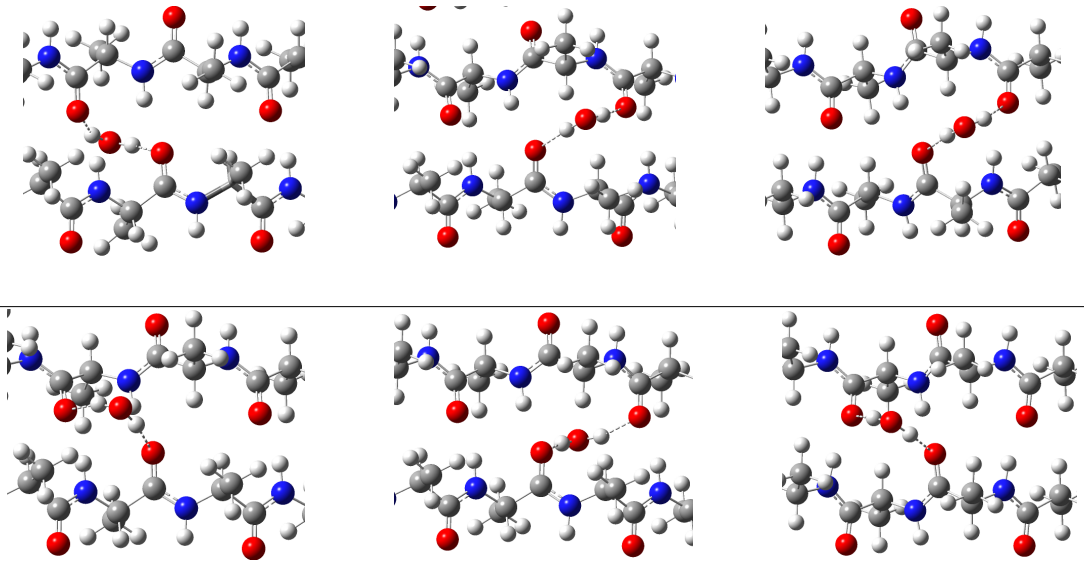


Figure 4.8: Different discrete waters at the β-sheet's surface. The top row shows from waters from the methyls side (Me), the bottom shows from the opposite side (O) at Small and Large Rings in the atniparallel β-sheet and at the ring in the parallel β-sheet

Table 4.9: The enthalpy in the gas phase (ΔH_{gas}), desolvation penalty ($\Delta\Delta G$, defined in the table), and the sum of those two contributions (Σ) for the formation of explicitly hydrated β-sheets in the solution (all values in kcal/mol).

Size	Antiparallel					Parallel				
	ΔH_{gas}	UFF		Pauling		ΔH_{gas}	UFF		Pauling	
		$\Delta\Delta$G	Σ	$\Delta\Delta$G	Σ		$\Delta\Delta$G	Σ	$\Delta\Delta$G	Σ
Cycle for 6 waters at C_5 position										
$\Sigma = \Delta H_{gas} + \Delta G_{w\sim\beta-sheet(n)} - \Delta G_{w\sim\beta-sheet(n-1)} - \Delta G_{w\sim strand} + (6 \vee 12) \cdot \Delta G_w = \Delta H_{gas} + \Delta\Delta G$										
1→ 2	10.93	-3.54	7.39	-5.93	5.00	16.18	-6.96	9.22	-10.04	6.14
2→ 3	15.20	-7.16	8.04	-8.65	6.55	14.83	-5.01	9.81	-7.37	7.46
3→ 4	12.90	-5.15	7.75	-6.19	6.71	16.84	-7.01	9.83	-8.35	8.49
Cycle for 6 waters at C_7 position										
1→ 2	23.49	-7.56	15.93	-15.68	7.81	12.99	3.62	16.61	-2.85	10.14
2→ 3	25.41	-10.21	15.20	-13.08	12.34	14.98	2.18	17.17	-2.54	12.44
3→ 4	19.55	-4.74	14.81	-8.78	10.77	14.96	1.10	16.06	-3.38	11.58
Cycle for 12 waters at C_7 position										
1→ 2	77.94	-31.80	46.14	-49.57	28.37	81.47	-33.40	48.07	-53.02	28.45
2→ 3	78.48	-32.36	46.12	-50.25	28.23	80.19	-31.33	48.86	-50.48	29.71
3→ 4	77.58	-31.85	45.73	-50.08	27.50	80.63	-33.05	47.58	-50.00	30.63

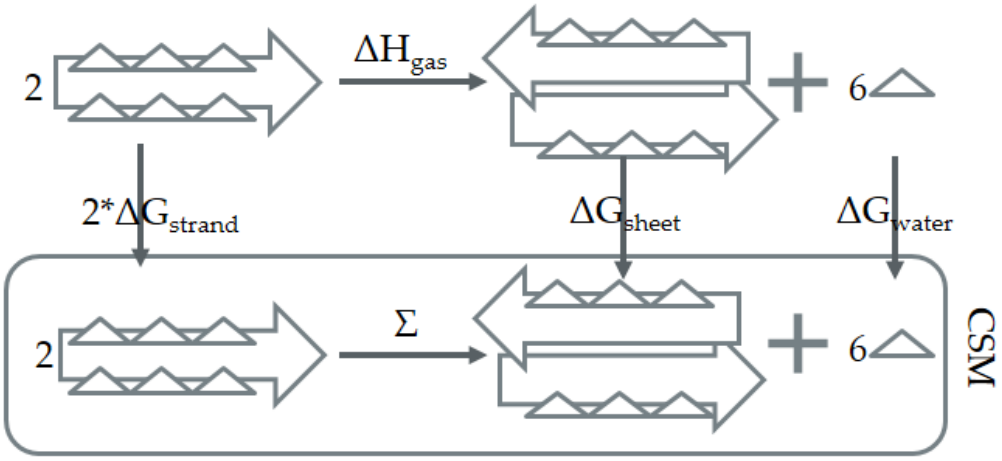


Figure 4.9: Example of the thermodynamic cycle for aggregation of the β-sheets capped with explicit water molecules.

observed by the decrease in hydration enthalpy from two to four strands. Also, one can see that the difference in the hydration energy in the gas phase between the antiparallel and parallel sheets is rather small and would favor antiparallel sheets due to their intrinsic stability. In the bulk solvent the C_7 hydration is more favorable than the C_5. However in the case of the antiparallel β-sheets we see that enthalpy os interaction increases with the size of the sheet in the Pauling model. The parallel sheets retain some cooperative effect but it is even smaller than in the gas phase. For the 12 waters models the observation remain the same. In the antiparallel sheets, the C_7 model is more stable for 2-4 strands, while the single strand is more stable in C_5 hydration, again due to the extra water-water hydrogen bonding not present in another cluster. In the parallel conformer, the C_5 model is more stable for one and two strands and C_7 dominates the larger aggregates. Likewise for of the six waters, they experience none (C_5) or just a little (C_7) of cooperative effect which is more pronounced in the gas phase than in the solution.

In table 4.9 I present the incremental ΔH's of the aggregation of the explicitly hydrated by the six and 12 C_7 waters β-sheets. Since both the sheet and the single strand are hydrated by six (or 12) waters and the product is hydrated only at the termini by six (or 12) waters, the remaining waters are released to the vacuum or bulk water. The stabilization by the bulk water is included in $\Delta\Delta$G term, as shown in the equation in the table. Since the number of hydrogen bonds between the solute and the waters decreases, the energetic contribution to the process of the β-sheet formation is different than these for bare β-sheets. As one can see, the positive enthalpy of adding an extra strand in the gas phase reflects the fact that 6 (or 12) waters are released, however if the waters are released in solution, they are stabilized by bulk water which affects the $\Delta\Delta$G term. In case of the bare β-sheets I observed that the formation of the sheet was associated with some desolvation penalty due to removing the strand from the solution. If we explicitly hydrate the termini, the released waters will compensate for the change in the solvation energy decreasing the $\Delta\Delta$G's, however, the overall process of the aggregation will be unfavorable due to break of the strong hydrogen bonds between the solute and the first solvation shell. The incremental enthalpies for sheets with six C_7 discrete waters are in the range of 15 kcal/mol for UFF model and from 7.8 kcal/mol for forming a dimer to 12.34 kcal/mol for a trimer for Pauling model, while the enthalpies for parallel sheets are slightly more positive. The model of the 12 explicit C_7 water predicts incremental enthalpy of 45-46 kcal/mol for UFF model and 28-27 kcal for Pauling model for

antiparallel sheets. This calculations do not take into account strong water-water hydrogen bond present in the explicitly hydrated sheets since we move waters into the bulk individually, while the enthalpy of water – water HB in the UFF model is -4.8 kcal/mol. However, if I correct the incremental enthalpy by 6·(-4.8)=-28.8 kcal/mol, the estimated incremental enthalpy is around 15-16 kcal/mol, in agreement with the model of six waters. Therefore the addition of explicit waters at the termini seem to correct the erratic behavior of continuous solvation models.

The remaining question is what happens if waters do not move to the bulk solution, but find another stabilizing spot at the newly formed sheet. The potential places are on the surface of the sheet as showed in figure 4.8. In the bottom of the table 4.8 I present data on binding of a single water molecule to the small and big rings (please see figure 3.3 for antiparallel sheets, in parallel sheet both rings are equivalent) in the middle of 4-stranded β sheet from the side of methyl groups (Me) and opposite to them (O). For the antiparallel sheet in Pauling model the hydration on both rings opposite to the methyl group is negative, providing around stabilization of -1 kcal/mol per ring. For parallel sheets, the waters binding opposite to the methyls provide stabilization of -1.4 kcal/mol per ring, while water binding from the methyls side inserts to the middle of the sheet, breaking one of the hydrogen bonds. However, the structure of the complex with inserted water might be the artefact of chosen starting geometry since the water molecule first molecule has to squeeze between two methyl groups which sterically screen the hydration of the ring. what is enthalpically very unfavorable when compared to the isolated sheet and water. Therefore I would assume stabilization energy of -1.43 kcal/mol per ring as predicted for the “opposite“ geometries. Assuming that the stabilization due to the ring hydration is additive and there are six spots at the sheet’s surface (3 small and 3 big rings for antiparallel and 6 rings for parallel) per interaction between two strands I obtain extra stabilization of -7.4 kcal/mol and -9 kcal/mol per six waters for antiparallel and parallel conformation respectively which correction is too small to reverse the trends observed in the table 4.9.

Conclusions

In this section I show that continuous solvation models predict aggregations of $Ac(Ala)_6NHM$ e hexapeptides into the stable β-sheets in aqueous solution. The calculations shows that waters have strong favorable interactions with the strand/sheet edge which experience only little of the

cooperative effect. These specific water molecules seem to be essential to prohibit grow of the β-sheets in the solution. Using them as a correction for the solvent models leads to positive enthalpies of formations and as a result they maintain polyalanines peptide as single strands. I also find that there are few position more favorable than bulk water on the surface of the sheet, however they do not reverse the observed trend for terminal solvation.

4.6 Concluding remarks

In the foregoing chapter I presented data regarding aqueous solvation of protein secondary structures. While not directly encoded in the amino acid sequence, one must remember that proteins evolved in the water environment, hence, despite indirectly, water (its ability to form hydrogen bonds, very high dipole moment, high poralizability) plays an active role in any living organism. My major goal was to investigate the influence of the specific waters on the various protein secondary motives described at the high level DFT theory.

First I focused only on the water itself. It has been observed that the hydrophobic effect behaves differently for the rare gases and larger non-polar molecules. In my work I distinguish between the hydrophobic effect as observed for proteins and larger molecules with conformational flexibility and hydrophobic solvation which can be explained without invoking the entropy. I find that the positive free energy of transfer from the gas phase to aqueous solution correlate with the solvent CED, i.e. the energy required to make a hole in large water cluster big enough to accommodate the solute is not negligible (around 7 kcal/mol) and correlates with the internal pressure (22,700 atmospheres) of water. Therefore the work is done not against the atmospheric pressure but the solvent's CED and if one corrects the enthalpy of transfer by this factor, the entropy loses its significance.

In the section 3.3 I revise the Venkatachalam's [4] classification of β-turns. I used a model six L-alanine capped polypeptide for a model β-turn. I found that the C_{10} hydrogen bonds is very weak in the gas phase and it tends to open and form the C_7 hydrogen bond with theoretically non-bonding peptide group between i+1 and i+2 positions. I collectively call turns having the C_{10} HB intact the A-turns while turns which have other HB between residues i and i+3 are named the B-type. The most stable turn, B_I, with one C_7 hydrogen bond was found to be more stable than

the A_{III} turn (related to the type I in the old classification) by 7.2 kcal/mol. Other A-type turns, A_I and A_{II} (correlating with best with the type I' and type II', respectively) were found to be less stable by 6.1 and 3.8 kcal/mole. I see that not only the A_{III} (type I) turn occurred to be less stable than B_I turn but also other A_{II} and A_I turns which are supposed to exist only for the DL and DD amino acid are more stable than the A_{III} turn.

This results were the motivation for the further studies of influence of the solvent on the turns' stability. Turns in proteins are usually exposed to the solvent hence hydration might influence its energy landscape. I use explicit and implicit hydration models to investigate the problem on various levels. I found that implicit models reduces differences between the A- and B-type turns but leaves their relative order unchanged with the B_1 turn being the most stable. I also use 12 water molecules to simulate the first hydration shell and I find that while some turns - A_{II} and newly identified A_{VI} turns (which correlates with the type II turn) – are more stabilized than the B-type of turns, the A_{III} turn is in fact very poorly hydrated and lies higher in energy than the most stable B_I (the implicit solvation) and A_{II} (the explicit solvation) turns.

α-helix is the next investigated motif. The aqueous solvation of α-helices, especially those that do not contain hydrophilic residues, occurs via two very different mechanisms. Water hydrates the termini through cooperative H-bonding, as suggested by Karplus, then hydrates the helix/water aggregate (partially hydrated at the termini) as would normally be expected from a solvent reaction field such as that incorporated in the CPCM method.

As the last project I investigate the hydration of the β-sheets. I found that continuous solvation models incorrectly predict that single strands should aggregate in the solution into big soluble β-sheets which stays in contradiction with the reality where insoluble silk (alanine based antiparallel β-sheets) and formation of parallel β-sheets in Alzheimer disease might serve as the simplest examples. I used models of six alanines strands in parallel and antiparallel sheets conformations. I observed that water molecules are more stable at the sheet's termini than in the bulk water hence there must be a driving force bringing them to the sheet's first hydration shell. Therefore well hydrated strands do not form sheets since it requires the dehydration of the two sides of the strand which is more than the energy gained from the formed β-sheet. This penalty is underestimated in the CPCM continuous solvent which manifest in the negative enthalpy of bringing two single strands together.

Chapter 5

Conclusions

I would like to briefly summarize my research in few following points:

1. I compared the dispersion and induction interactions for noble gas dimers and for neon, methane and 2-butyne with HF and LiF using a variety of functionals with ab initio methods CCSD(T) and MP2. I observed that inductive interactions tend to enhance dispersion and may be accompanied by charge-transfer. I found that the functionals parameterized to treat dispersion interactions often overestimate these interactions, sometimes by quite a lot, when compared to higher level calculations and functionals which do not describe pure dispersion interactions. They also appear to describe dispersion mixed with induction about as accurately as those parameterized to treat dispersion.

2. I reevaluated the the interaction of pyrimidine and p-benzoquinone using functionals designed to treat dispersion, traditional functionals and ab initio CCSD(T) and MP2 methods. The functionals designed to treat dispersion behaved erratically as the predictions of the most stable structure vary considerably, predicting dispersion-based structures to be more stable than the experimentally observed hydrogen bonded structure. The erroneous results of dispersion-corrected functionals affect their reliability in the evaluation of possible π-stacking in biologically important systems.

3. I reinvestigated the energetic and structural properties of fully optimized α-helical and antiparallel β-sheets polyalanines using functionals designed to treat dispersion and other traditional functionals. The functionals developed to treat dispersion significantly overesti-

mated interaction enthalpies of folding for the α-helix and predict unreasonable structures that contain Ramachandran ϕ and ψ angles that are out of the bounds of databases compiled β-sheets.

4. I used DFT calculation of three noble gases (Neon, Argon and Krypton) in clusters of 50 waters to investigate the origins of the hydrophobic effect. Vibrational analyses showed no substantial decreases in the vibrational entropies of the waters. The observed positive free energies of transfer from the gas phase to water appear to be due to the work needed to make a suitable hole in the aquoues solution.

5. I presented a new classification of β-turns specific to antiparallel β-sheets based upon the topology of hydrogen bond formation. I found the most stable structures in the gas phase and CPCM solvent model have C_7 cyclic hydrogen bond in place of the C_{10} interactions specified in the gas phase. The explicit waters find the C_{10} turn to be the most stable, however, its mean unsigned error for ϕ and ψ Ramachandran angles correlates best with the type II', not type I' as expected from the theory.

6. I investigated the effects of discrete water molecules and the CPCM and SM5.2 solvent continuum model on the solvation of polyalanine α-helices in water. I found that individual water molecules cooperatively hydrogen bond to both the C- and N- termini of the helix, which result in increases in the dipole moment of the helix-water complex more than the vector sum of their individual dipole moments. On the other hand individual water molecules that interact with the backbone lower the dipole moment of the helix/water complex to below that of the helix itself. The results imply that solvation of α-helix undergo two different mechanisms, first the specific terminal hydration as suggested by Karplus, and then the bulk solvation as defined by Onsager.

7. I studied the formation of the antiparallel and parallel β-sheets in aqueous solution using discrete waters and continuum solvation model using different size of cavities. I found that the solvation models erratically predict aggregation of β-sheets in the solution and correction for explicit waters hydrating the termini of the strands and sheets.

Bibliography

[1] http://www.hunter.cuny.edu/chemistry/faculty/Joe/Joe.

[2] J. Pople, "Nobel lecture: Quantum chemical models". nobelprize.org."

[3] J. Klimeš and A. Michaelides, "Perspective: Advances and challenges in treating van der waals dispersion forces in density functional theory," *The Journal of Chemical Physics*, vol. 137, no. 12, p. 120901, 2012.

[4] C. M. Venkatachalam, "Stereochemical criteria for polypeptides and proteins. v. conformation of a system of three linked peptide units," *Biopolymers*, vol. 6, no. 10, pp. 1425--1436, 1968.

[5] www.xkcd.com/1095.

[6] K. Burke, "Perspective on density functional theory," *The Journal of Chemical Physics*, vol. 136, no. 15, p. 150901, 2012.

[7] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Challenges for density functional theory," *Chemical Reviews*, vol. 112, no. 1, pp. 289--320, 2012.

[8] C. J. Cramer and D. G. Truhlar, "Implicit solvation models: equilibria, structure, spectra, and dynamics," *Chemical Reviews*, vol. 99, no. 8, pp. 2161--2200, 1999.

[9] J. Tomasi, B. Mennucci, and R. Cammi, "Quantum mechanical continuum solvation models," *Chemical Reviews*, vol. 105, no. 8, pp. 2999--3094, 2005.

[10] J. Tomasi and M. Persico, "Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent," *Chemical Reviews*, vol. 94, no. 7, pp. 2027--2094, 1994.

[11] J. Tomasi, E. Cances, C. S. Pomelli, M. Caricato, G. Scalmani, M. J. Frisch, R. Cammi, M. V. Basilevsky, G. N. Chuev, and B. Mennucci, *Modern Theories of Continuum Models*. John Wiley & Sons, Ltd, 2007.

[12] K. E. Riley and P. Hobza, "On the importance and origin of aromatic interactions in chemistry and biodisciplines," *Accounts of Chemical Research*, vol. 0, no. 0, p. null, 0.

[13] T. M. Parker, E. G. Hohenstein, R. M. Parrish, N. V. Hud, and C. D. Sherrill, "Quantum-mechanical analysis of the energetic contributions to π stacking in nucleic acids versus rise, twist, and slide," *Journal of the American Chemical Society*, vol. 135, no. 4, pp. 1306--1316, 2013.

[14] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.*, vol. 136, pp. B864--B871, Nov 1964.

[15] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Phys. Rev.*, vol. 140, pp. A1133--A1138, Nov 1965.

[16] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "General performance of density functionals," *The Journal of Physical Chemistry A*, vol. 111, no. 42, pp. 10439--10452, 2007.

[17] L. A. Burns, Álvaro Vázquez-Mayagoitia, B. G. Sumpter, and C. D. Sherrill, "Density-functional approaches to noncovalent interactions: A comparison of dispersion corrections (dft-d), exchange-hole dipole moment (xdm) theory, and specialized functionals," *The Journal of Chemical Physics*, vol. 134, no. 8, p. 084107, 2011.

[18] K. S. Thanthiriwatte, E. G. Hohenstein, L. A. Burns, and C. D. Sherrill, "Assessment of the performance of dft and dft-d methods for describing distance dependence of hydrogen-bonded interactions," *Journal of Chemical Theory and Computation*, vol. 7, no. 1, pp. 88--96, 2011.

[19] O. A. Vydrov and T. Van Voorhis, "Benchmark assessment of the accuracy of several van der waals density functionals," *Journal of Chemical Theory and Computation*, vol. 8, no. 6, pp. 1929--1934, 2012.

[20] M. S. Marshall, L. A. Burns, and C. D. Sherrill, "Basis set convergence of the coupled-cluster correction, $\delta^{MP2}_{CCSD(T)}$: Best practices for benchmarking non-covalent interactions and the attendant revision of the s22, nbc10, hbc6, and hsg databases," *The Journal of Chemical Physics*, vol. 135, no. 19, p. 194102, 2011.

[21] http://en.wikipedia.org/wiki/List_of_most_popular_websites.

[22] J. P. Perdew, A. Ruzsinszky, L. A. Constantin, J. Sun, and G. I. Csonka, "Some fundamental issues in ground-state density functional theory: A guide for the perplexed," *Journal of Chemical Theory and Computation*, vol. 5, no. 4, pp. 902--908, 2009.

[23] E. Engel and R. Dreizler, *Density Functional Theory: An Advanced Course*. Theoretical and mathematical physics, Springer Berlin Heidelberg, 2011.

[24] A. D. Becke, "Density-functional thermochemistry. iii. the role of exact exchange," *The Journal of Chemical Physics*, vol. 98, no. 7, pp. 5648--5652, 1993.

[25] H. Eshuis, J. Bates, and F. Furche, "Electron correlation methods based on the random phase approximation," *Theoretical Chemistry Accounts*, vol. 131, no. 1, pp. 1--18, 2012.

[26] S. Grimme, "Semiempirical hybrid density functional with perturbative second-order correlation," *The Journal of Chemical Physics*, vol. 124, no. 3, p. 034108, 2006.

[27] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, "Development and use of quantum mechanical molecular models. 76. am1: a new general purpose quantum mechanical molecular model," *Journal of the American Chemical Society*, vol. 107, no. 13, pp. 3902--3909, 1985.

[28] C. Lee, W. Yang, and R. G. Parr, "Development of the colle-salvetti correlation-energy formula into a functional of the electron density," *Phys. Rev. B*, vol. 37, pp. 785--789, Jan 1988.

[29] A. D. Becke, "Density-functional thermochemistry. iv. a new dynamical correlation functional and implications for exact-exchange mixing," *The Journal of Chemical Physics*, vol. 104, no. 3, pp. 1040--1046, 1996.

[30] Y. Zhao and D. G. Truhlar, "Density functionals with broad applicability in chemistry," *Accounts of Chemical Research*, vol. 41, no. 2, pp. 157--167, 2008. PMID: 18186612.

[31] C. J. Cramer and D. G. Truhlar, "A universal approach to solvation modeling," *Accounts of Chemical Research*, vol. 41, no. 6, pp. 760--768, 2008.

[32] A. Klamt, "Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena," *The Journal of Physical Chemistry*, vol. 99, no. 7, pp. 2224--2235, 1995.

[33] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, no. 2, pp. 926--935, 1983.

[34] J. R. Pliego and J. M. Riveros, "Theoretical calculation of pka using the cluster-continuum model," *The Journal of Physical Chemistry A*, vol. 106, no. 32, pp. 7434--7439, 2002.

[35] K.-H. Cho, K. T. No, and H. A. Scheraga, "Ion pair interactions in aqueous solution: self-consistent reaction field (scrf) calculations with some explicit water molecules," *The Journal of Physical Chemistry A*, vol. 104, no. 27, pp. 6505--6509, 2000.

[36] C. P. Kelly, C. J. Cramer, and D. G. Truhlar, "Adding explicit solvent molecules to continuum solvent calculations for the calculation of aqueous acid dissociation constants," *The Journal of Physical Chemistry A*, vol. 110, no. 7, pp. 2493--2499, 2006.

[37] V. S. Bryantsev, M. S. Diallo, and W. A. Goddard III, "Calculation of solvation free energies of charged solutes using mixed cluster/continuum models," *The Journal of Physical Chemistry B*, vol. 112, no. 32, pp. 9709--9719, 2008. PMID: 18646800.

[38] R. S. Rowland and R. Taylor, "Intermolecular nonbonded contact distances in organic crystal structures: Comparison with distances expected from van der waals radii," *The Journal of Physical Chemistry*, vol. 100, no. 18, pp. 7384--7391, 1996.

[39] A. Bondi, "van der waals volumes and radii," *The Journal of Physical Chemistry*, vol. 68, no. 3, pp. 441--451, 1964.

[40] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, "Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations," *Journal of the American Chemical Society*, vol. 114, no. 25, pp. 10024--10035, 1992.

[41] M. Bachs, F. J. Luque, and M. Orozco, "Optimization of solute cavities and van der waals parameters in ab initio mst-scrf calculations of neutral molecules," *Journal of Computational Chemistry*, vol. 15, no. 4, pp. 446--454, 1994.

[42] D. M. York and M. Karplus, "A smooth solvation potential based on the conductor-like screening model," *The Journal of Physical Chemistry A*, vol. 103, no. 50, pp. 11060--11079, 1999.

[43] M. Cossi, N. Rega, G. Scalmani, and V. Barone, "Energies, structures, and electronic properties of molecules in solution with the c-pcm solvation model," *Journal of Computational Chemistry*, vol. 24, no. 6, pp. 669--681, 2003.

[44] R. A. Pierotti, "A scaled particle theory of aqueous and nonaqueous solutions," *Chemical Reviews*, vol. 76, no. 6, pp. 717--726, 1976.

[45] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, "Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions," *The Journal of Physical Chemistry B*, vol. 113, no. 18, pp. 6378--6396, 2009.

[46] S. Lem, *The cyberiad : fables for the cybernetic age*. San Diego: Harcourt Brace Jovanovich, 1985.

[47] Y. Andersson, D. C. Langreth, and B. I. Lundqvist, "van der waals interactions in density-functional theory," *Phys. Rev. Lett.*, vol. 76, pp. 102--105, Jan 1996.

[48] J. F. Dobson and B. P. Dinte, "Constraint satisfaction in local and gradient susceptibility approximations: Application to a van der waals density functional," *Phys. Rev. Lett.*, vol. 76, pp. 1780--1783, Mar 1996.

[49] S. Grimme, "Accurate description of van der waals complexes by density functional theory including empirical corrections," *Journal of Computational Chemistry*, vol. 25, no. 12, pp. 1463--1473, 2004.

[50] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, 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, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, "Gaussian 09 revision a.1." Gaussian Inc. Wallingford CT 2009.

[51] S. Grimme, "Semiempirical gga-type density functional constructed with a long-range dispersion correction," *Journal of Computational Chemistry*, vol. 27, no. 15, pp. 1787--1799, 2006.

[52] J.-D. Chai and M. Head-Gordon, "Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections," *Phys. Chem. Chem. Phys.*, vol. 10, pp. 6615--6620, 2008.

[53] T. Schwabe and S. Grimme, "Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability," *Phys. Chem. Chem. Phys.*, vol. 9, pp. 3397--3406, 2007.

[54] X. Xu and W. A. Goddard, "The x3lyp extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 9, pp. 2673--2677, 2004.

[55] C. Adamo and V. Barone, "Toward reliable density functional methods without adjustable parameters: The pbe0 model," *The Journal of Chemical Physics*, vol. 110, no. 13, pp. 6158--6170, 1999.

[56] J. A. Plumley and J. J. Dannenberg, "A comparison of the behavior of functional/basis set combinations for hydrogen-bonding in the water dimer with emphasis on basis set superposition error," *Journal of Computational Chemistry*, vol. 32, no. 8, pp. 1519--1527, 2011.

[57] R. Wieczorek and J. J. Dannenberg, "Comparison of fully optimized a- and 310-helices with extended b-strands. an oniom density functional theory study," *Journal of the American Chemical Society*, vol. 126, no. 43, pp. 14198--14205, 2004. PMID: 15506786.

[58] R. Wieczorek and J. J. Dannenberg, "H-bonding cooperativity and energetics of α-helix formation of five 17-amino acid peptides," *Journal of the American Chemical Society*, vol. 125, no. 27, pp. 8124--8129, 2003. PMID: 12837081.

[59] J. Dobson, "Dispersion (van der waals) forces and tddft," in *Fundamentals of Time-Dependent Density Functional Theory* (M. A. Marques, N. T. Maitra, F. M. Nogueira, E. Gross, and A. Rubio, eds.), vol. 837 of *Lecture Notes in Physics*, pp. 417--441, Springer Berlin Heidelberg, 2012.

[60] E. R. Johnson, A. D. Becke, C. D. Sherrill, and G. A. DiLabio, "Oscillations in meta-generalized-gradient approximation potential energy surfaces for dispersion-bound complexes," *The Journal of Chemical Physics*, vol. 131, no. 3, p. 034111, 2009.

[61] J. J. A. Montgomery, M. J. Frisch, J. W. Ochterski, and G. A. Petersson, "A complete basis set model chemistry. vii. use of the minimum population localization method," *The Journal of Chemical Physics*, vol. 112, no. 15, pp. 6532--6542, 2000.

[62] M. O. Sinnokrot and C. D. Sherrill, "High-accuracy quantum mechanical studies of pi-pi interactions in benzene dimers," *The Journal of Physical Chemistry A*, vol. 110, no. 37, pp. 10656--10668, 2006. PMID: 16970354.

[63] R. L. Baldwin, C. Frieden, and G. D. Rose, "Dry molten globule intermediates and the mechanism of protein unfolding," *Proteins: Structure, Function, and Bioinformatics*, vol. 78, no. 13, pp. 2725--2737, 2010.

[64] M. Purugganan, C. Kumar, N. Turro, and J. Barton, "Accelerated electron transfer between metal complexes mediated by dna," *Science*, vol. 241, no. 4873, pp. 1645--1649, 1988.

[65] W. McCarthy, A. M. Plokhotnichenko, E. D. Radchenko, J. Smets, D. M. A. Smith, S. G. Stepanian, and L. Adamowicz, "H-bonded and stacked dimers of pyrimidine and p-benzoquinone. a combined matrix isolation infrared and theoretical ab initio study," *The Journal of Physical Chemistry A*, vol. 101, no. 39, pp. 7208--7216, 1997.

[66] D. Moran, A. C. Simmonett, F. E. Leach, W. D. Allen, P. v. R. Schleyer, and H. F. Schaefer, "Popular theoretical methods predict benzene and arenes to be nonplanar," *Journal of the American Chemical Society*, vol. 128, no. 29, pp. 9342--9343, 2006.

[67] P. Salvador, R. Wieczorek, and J. J. Dannenberg, "Direct calculation of trans-hydrogen-bond 13c-15n 3-bond j-couplings in entire polyalanine α-helices. a density functional theory study," *The Journal of Physical Chemistry B*, vol. 111, no. 9, pp. 2398--2403, 2007. PMID: 17295533.

[68] M. Marianski, A. Asensio, and J. J. Dannenberg, "Comparison of some dispersion-corrected and traditional functionals as applied to peptides and conformations of cyclohexane derivatives," *The Journal of Chemical Physics*, vol. 137, no. 4, p. 044109, 2012.

[69] P. Salvador, A. Asensio, and J. J. Dannenberg, "The effect of aqueous solvation upon α-helix formation for polyalanines," *The Journal of Physical Chemistry B*, vol. 111, no. 25, pp. 7462--7466, 2007. PMID: 17552560.

[70] T. Ooi and M. Oobatake, "Prediction of the thermodynamics of protein unfolding: the helix-coil transition of poly(l-alanine).," *Proceedings of the National Academy of Sciences*, vol. 88, no. 7, pp. 2859--2863, 1991.

[71] G. E. Job, R. J. Kennedy, B. Heitmann, J. S. Miller, S. M. Walker, and D. S. Kemp, "Temperature- and length-dependent energetics of formation for polyalanine helices in water: assignment of wala(n,t) and temperature-dependent cd ellipticity standards," *Journal of the American Chemical Society*, vol. 128, no. 25, pp. 8227--8233, 2006. PMID: 16787087.

[72] J. M. Scholtz, H. Qian, E. J. York, J. M. Stewart, and R. L. Baldwin, "Parameters of helix–coil transition theory for alanine-based peptides of varying chain lengths in water," *Biopolymers*, vol. 31, no. 13, pp. 1463--1470, 1991.

[73] J. M. Richardson, M. M. Lopez, and G. I. Makhatadze, "Enthalpy of helix–coil transition: Missing link in rationalizing the thermodynamics of helix-forming propensities of the amino acid residues," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 5, pp. 1413--1418, 2005.

[74] G. Rialdi and J. Hermans, "Calorimetric heat of the helix-coil transition of poly-l-glutamic acid1a," *Journal of the American Chemical Society*, vol. 88, no. 24, pp. 5719--5720, 1966.

[75] P. Y. Chou and H. A. Scheraga, "Calorimetric measurement of enthalpy change in the isothermal helix–coil transition of poly-l-lysine in aqueous solution," *Biopolymers*, vol. 10, no. 4, pp. 657--680, 1971.

[76] D. Roy and J. Dannenberg, "The effects of regularly spaced glutamine substitutions on alpha-helical peptide structures: A dft/oniom study," *Chemical Physics Letters*, vol. 512, no. 46, pp. 255 -- 257, 2011.

[77] E. Baker and R. Hubbard, "Hydrogen bonding in globular proteins," *Progress in Biophysics and Molecular Biology*, vol. 44, no. 2, pp. 97 -- 179, 1984.

[78] O. Koch, M. Bocola, and G. Klebe, "Cooperative effects in hydrogen-bonding of protein secondary structure elements: A systematic analysis of crystal data using secbase," *Proteins: Structure, Function, and Bioinformatics*, vol. 61, no. 2, pp. 310--317, 2005.

[79] M. Marianski and J. J. Dannenberg, "Aqueous solvation of polyalanine α-helices with specific water molecules and with the cpcm and sm5.2 aqueous continuum models using density functional theory," *The Journal of Physical Chemistry B*, vol. 116, no. 4, pp. 1437--1445, 2012.

[80] M. P. Waller, H. Kruse, C. Muck-Lichtenfeld, and S. Grimme, "Investigating inclusion complexes using quantum chemical methods," *Chem. Soc. Rev.*, vol. 41, pp. 3119--3128, 2012.

[81] D. Roy, M. Marianski, N. T. Maitra, and J. J. Dannenberg, "Comparison of some dispersion-corrected and traditional functionals with ccsd(t) and mp2 ab initio methods: Dispersion, induction, and basis set superposition error," *The Journal of Chemical Physics*, vol. 137, no. 13, p. 134109, 2012.

[82] T. Creighton, *Proteins: Structures and Molecular Principles*. International student edition, Freeman, 1983.

[83] S. Spera and A. Bax, "Empirical correlation between protein backbone conformation and c.alpha. and c.beta. 13c nuclear magnetic resonance chemical shifts," *Journal of the American Chemical Society*, vol. 113, no. 14, pp. 5490--5492, 1991.

[84] J. D. van Beek, S. Hess, F. Vollrath, and B. H. Meier, "The molecular structure of spider dragline silk: Folding and orientation of the protein backbone," *Proceedings of the National Academy of Sciences*, vol. 99, no. 16, pp. 10266--10271, 2002.

[85] L. Carroll, *Alice in Wonderland*. Classics illustrated, Rylee Limited, 1949.

[86] P. Ball, "More than a bystander," *Nature*, vol. 478, pp. 467--468, 2011.

[87] D. Lucent, C. D. Snow, C. E. Aitken, and V. S. Pande, "Non-bulk-like solvent behavior in the ribosome exit tunnel," *PLoS Comput Biol*, vol. 6, p. e1000963, 10 2010.

[88] H. S. Frank and M. W. Evans, "Free volume and entropy in condensed systems iii. entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes," *The Journal of Chemical Physics*, vol. 13, no. 11, pp. 507--532, 1945.

[89] W. Kauzmann, "Some factors in the interpretation of protein denaturation," vol. 14 of *Advances in Protein Chemistry*, pp. 1 -- 63, Academic Press, 1959.

[90] M. H. Abraham, "Free energies of solution of rare gases and alkanes in water and nonaqueous solvents. a quantitative assessment of the hydrophobic effect," *Journal of the American Chemical Society*, vol. 101, no. 19, pp. 5477--5484, 1979.

[91] T. Asano and W. J. Le Noble, "Activation and reaction volumes in solution," *Chemical Reviews*, vol. 78, no. 4, pp. 407--489, 1978.

[92] J. J. Gajewski, "A semitheoretical multiparameter approach to correlate solvent effects on reactions and equilibria," *The Journal of Organic Chemistry*, vol. 57, no. 20, pp. 5500--5506, 1992.

[93] T. Urbic, V. Vlachy, Y. V. Kalyuzhnyi, and K. A. Dill, "Theory for the solvation of nonpolar solutes in water," *The Journal of Chemical Physics*, vol. 127, no. 17, p. 174505, 2007.

[94] A. Bymaster, A. Dominik, and W. G. Chapman, "Hydration structure and interfacial properties of water near a hydrophobic solute from a fundamental measure density functional theory," *The Journal of Physical Chemistry C*, vol. 111, no. 43, pp. 15823--15831, 2007.

[95] R. Godawat, S. N. Jamadagni, and S. Garde, "Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations," *Proceedings of the National Academy of Sciences*, vol. 106, no. 36, pp. 15119--15124, 2009.

[96] B. Lee, "Solvent reorganization contribution to the transfer thermodynamics of small nonpolar molecules," *Biopolymers*, vol. 31, no. 8, pp. 993--1008, 1991.

[97] R. Breslow and T. Guo, "Diels-alder reactions in nonaqueous polar solvents. kinetic effects of chaotropic and antichaotropic agents and of β-cyclodextrin," *Journal of the American Chemical Society*, vol. 110, no. 17, pp. 5613--5617, 1988.

[98] M. K. Cambell and S. O. Farrel, *Biochemistry*. Brooks Cole, 7th ed., 2011.

[99] J. Ali-Torres and J. J. Dannenberg, "The folding of acetyl(ala)28nh2 and ac(ala)40nh2 extended strand peptides into antiparallel b-sheets. a density functional theory study of b-sheets with b-turns," *The Journal of Physical Chemistry B*, vol. 116, no. 48, pp. 14017--14022, 2012.

[100] G. D. Rose, L. M. Glerasch, and J. A. Smith, "Turns in peptides and proteins," vol. 37, pp. 1 -- 109, 1985.

[101] P. N. Lewis, F. A. Momany, and H. A. Scheraga, "Folding of polypeptide chains in proteins: A proposed mechanism for folding," *Proceedings of the National Academy of Sciences*, vol. 68, no. 9, pp. 2293--2297, 1971.

[102] B. L. Sibanda, T. L. Blundell, and J. M. Thornton, "Conformation of Î²-hairpins in protein structures: A systematic classification with applications to modelling by homology, electron density fitting and protein engineering," *Journal of Molecular Biology*, vol. 206, no. 4, pp. 759 -- 777, 1989.

[103] I. L. Karle, S. K. Awasthi, and P. Balaram, "A designed beta-hairpin peptide in crystals," *Proceedings of the National Academy of Sciences*, vol. 93, no. 16, pp. 8189--8193, 1996.

[104] M. C. Etter, "Encoding and decoding hydrogen-bond patterns of organic compounds," *Accounts of Chemical Research*, vol. 23, no. 4, pp. 120--126, 1990.

[105] R. Wolfenden, "Interaction of the peptide bond with solvent water: a vapor phase analysis," *Biochemistry*, vol. 17, no. 1, pp. 201--204, 1978.

[106] F. Avbelj and R. L. Baldwin, "Limited validity of group additivity for the folding energetics of the peptide group," *Proteins: Structure, Function, and Bioinformatics*, vol. 63, no. 2, pp. 283--289, 2006.

[107] F. Avbelj and R. L. Baldwin, "Origin of the change in solvation enthalpy of the peptide group when neighboring peptide groups are added," *Proceedings of the National Academy of Sciences*, vol. 106, no. 9, pp. 3137--3141, 2009.

[108] L. Pauling, "Atomic radii and interatomic distances in metals," *Journal of the American Chemical Society*, vol. 69, no. 3, pp. 542--553, 1947.

[109] Y. Takano and K. N. Houk, "Benchmarking the conductor-like polarizable continuum model (cpcm) for aqueous solvation free energies of neutral and ionic organic molecules," *Journal of Chemical Theory and Computation*, vol. 1, no. 1, pp. 70--77, 2005.

[110] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, "Universal quantum mechanical model for solvation free energies based on gas-phase geometries," *The Journal of Physical Chemistry B*, vol. 102, no. 17, pp. 3257--3271, 1998.

[111] A. V. Marenich, R. M. Olson, C. P. Kelly, C. J. Cramer, and D. G. Truhlar, "Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges," *Journal of Chemical Theory and Computation*, vol. 3, no. 6, pp. 2011--2033, 2007.

[112] I. Soteras, C. Curutchet, A. Bidon-Chanal, M. Orozco, and F. J. Luque, "Extension of the mst model to the ief formalism: Hf and b3lyp parametrizations," *Journal of Molecular Structure: THEOCHEM*, vol. 727, pp. 29 -- 40, 2005.

[113] J. Ireta, "Microsolvation effects on the stability of polyalanine in extended and polyproline ii conformation," *International Journal of Quantum Chemistry*, vol. 112, no. 22, pp. 3612--3617, 2012.

[114] Z. Shi, C. A. Olson, G. D. Rose, R. L. Baldwin, and N. R. Kallenbach, "Polyproline ii structure in a sequence of seven alanine residues," *Proceedings of the National Academy of Sciences*, vol. 99, no. 14, pp. 9190--9195, 2002.

[115] L. L. Porter and G. D. Rose, "Redrawing the ramachandran plot after inclusion of hydrogen-bonding constraints," *Proceedings of the National Academy of Sciences*, vol. 108, no. 1, pp. 109--113, 2011.

[116] S. Marqusee, V. H. Robbins, and R. L. Baldwin, "Unusually stable helix formation in short alanine-based peptides," *Proceedings of the National Academy of Sciences*, vol. 86, no. 14, pp. 5286--5290, 1989.

[117] M. M. Lopez, D.-H. Chin, R. L. Baldwin, and G. I. Makhatadze, "The enthalpy of the alanine peptide helix measured by isothermal titration calorimetry using metal-binding to induce helix formation," *Proceedings of the National Academy of Sciences*, vol. 99, no. 3, pp. 1298--1302, 2002.

[118] R. Wieczorek and J. J. Dannenberg, "Enthalpies of hydrogen-bonds in α-helical peptides. an oniom dft/am1 study," *Journal of the American Chemical Society*, vol. 127, no. 42, pp. 14534--14535, 2005. PMID: 16231881.

[119] R. Wieczorek and J. J. Dannenberg, "The energetic and structural effects of single amino acid substitutions upon capped α-helical peptides containing 17 amino acid residues. an oniom dft/am1 study," *Journal of the American Chemical Society*, vol. 127, no. 49, pp. 17216--17223, 2005. PMID: 16332068.

[120] M. Elstner, K. Jalkanen, M. Knapp-Mohammady, T. Frauenheim, and S. Suhai, "Dft studies on helix formation in n-acetyl-(l-alanyl)-n-methylamide for n=1-20," *Chemical Physics*, vol. 256, no. 1, pp. 15 -- 27, 2000.

[121] J. Ireta, J. Neugebauer, M. Scheffler, A. Rojo, and M. Galván, "Density functional theory study of the cooperativity of hydrogen bonds in finite and infinite α-helices," *The Journal of Physical Chemistry B*, vol. 107, no. 6, pp. 1432--1437, 2003.

[122] M. V. Vener, A. N. Egorova, D. P. Fomin, and V. G. Tsirelson, "Hierarchy of the non-covalent interactions in the alanine-based secondary structures. dft study of the frequency shifts and electron-density features," *Journal of Physical Organic Chemistry*, vol. 22, no. 3, pp. 177--185, 2009.

[123] H. Nymeyer and A. E. García, "Simulation of the folding equilibrium of α-helical peptides: A comparison of the generalized born approximation with explicit solvent," *Proceedings of the National Academy of Sciences*, vol. 100, no. 24, pp. 13934--13939, 2003.

[124] A. E. García and K. Y. Sanbonmatsu, "α-helical stabilization by side chain shielding of backbone hydrogen bonds," *Proceedings of the National Academy of Sciences*, vol. 99, no. 5, pp. 2782--2787, 2002.

[125] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations," *Journal of Computational Chemistry*, vol. 24, no. 16, pp. 1999--2012, 2003.

[126] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *Journal of the American Chemical Society*, vol. 117, no. 19, pp. 5179--5197, 1995.

[127] H. Guo and M. Karplus, "Solvent influence on the stability of the peptide hydrogen bond: A supramolecular cooperative effect," *The Journal of Physical Chemistry*, vol. 98, no. 29, pp. 7104--7105, 1994.

[128] N. Kobko, L. Paraskevas, E. del Rio, and J. J. Dannenberg, "Cooperativity in amide hydrogen bonding chains: implications for protein-folding models," *Journal of the American Chemical Society*, vol. 123, no. 18, pp. 4348--4349, 2001. PMID: 11457207.

[129] A. Masunov and J. J. Dannenberg, "Theoretical study of urea and thiourea. 2. chains and ribbons," *The Journal of Physical Chemistry B*, vol. 104, no. 4, pp. 806--810, 2000.

[130] L. Onsager, "Electric moments of molecules in liquids," *Journal of the American Chemical Society*, vol. 58, no. 8, pp. 1486--1493, 1936.

[131] J. T. Su, X. Xu, and W. A. Goddard, "Accurate energies and structures for large water clusters using the x3lyp hybrid density functional," *The Journal of Physical Chemistry A*, vol. 108, no. 47, pp. 10518--10526, 2004.

[132] S. Simon, M. Duran, and J. J. Dannenberg, "How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers?," *The Journal of Chemical Physics*, vol. 105, no. 24, pp. 11024--11031, 1996.

[133] "Ampac, 8.16 ed.; semichem, inc.: Shawnee, ks.."

[134] J. J. Dannenberg, "Enthalpies of hydration of n-methylacetamide by one, two, and three waters and the effect upon the co stretching frequency. an ab initio dft study," *The Journal of Physical Chemistry A*, vol. 110, no. 17, pp. 5798--5802, 2006.

[135] R. M. Shields, B. Temelso, K. A. Archer, T. E. Morrell, and G. C. Shields, "Accurate predictions of water cluster formation, (h2o)n=2−10," *The Journal of Physical Chemistry A*, vol. 114, no. 43, pp. 11725--11737, 2010.

[136] P. J. Artymiuk, C. C. F. Blake, D. W. Rice, and K. S. Wilson, "The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme at 6 å resolution," *Acta Crystallographica Section B*, vol. 38, no. 3, pp. 778--783, 1982.

[137] C. P. Garnham, R. L. Campbell, and P. L. Davies, "Anchored clathrate waters bind antifreeze proteins to ice," *Proceedings of the National Academy of Sciences*, vol. 108, no. 18, pp. 7363--7367, 2011.

[138] P. e. a. Kjellbom, "Structural mechanism of plant aquaporin gating," *Nature*, vol. 439, p. 688.

[139] S. H. Park, A. A. Mrse, A. A. Nevzorov, M. F. Mesleh, M. Oblatt-Montal, M. Montal, and S. J. Opella, "Three-dimensional structure of the channel-forming trans-membrane domain of virus protein (vpu) from hiv-1," *Journal of Molecular Biology*, vol. 333, no. 2, pp. 409 -- 424, 2003.

[140] T. Lazaridis, "Structural determinants of transmembrane β-barrels," *Journal of Chemical Theory and Computation*, vol. 1, no. 4, pp. 716--722, 2005.

[141] J. C. Stroud, C. Liu, P. K. Teng, and D. Eisenberg, "Toxic fibrillar oligomers of amyloid-β have cross-β structure," *Proceedings of the National Academy of Sciences*, vol. 109, no. 20, pp. 7717--7722, 2012.

[142] R. Viswanathan, A. Asensio, and J. J. Dannenberg, "Cooperative hydrogen-bonding in models of antiparallel β-sheets," *The Journal of Physical Chemistry A*, vol. 108, no. 42, pp. 9205--9212, 2004.

[143] J. A. Plumley and J. J. Dannenberg, "The importance of hydrogen bonding between the glutamine side chains to the formation of amyloid vqivyk parallel β-sheets: An oniom dft/am1 study," *Journal of the American Chemical Society*, vol. 132, no. 6, pp. 1758--1759, 2010. PMID: 20088582.

[144] J. Ireta, "Microsolvation effect on the twist of β-sheets," *Journal of Chemical Theory and Computation*, vol. 7, no. 8, pp. 2630--2637, 2011.

[145] J. M. Mullin and M. S. Gordon, "Alanine: Then there was water," *The Journal of Physical Chemistry B*, vol. 113, no. 25, pp. 8657--8669, 2009. PMID: 19485320.

[146] A. Ben-Naim, "Standard thermodynamics of transfer. uses and misuses," *The Journal of Physical Chemistry*, vol. 82, no. 7, pp. 792--803, 1978.

[147] J. Ho, A. Klamt, and M. L. Coote, "Comment on the correct use of continuum solvent models," *The Journal of Physical Chemistry A*, vol. 114, no. 51, pp. 13442--13444, 2010.

[148] J. Makowska, S. Rodziewicz-Motowidlo, K. Baginska, M. Makowski, J. A. Vila, A. Liwo, L. Chmurzynski, and H. A. Scheraga, "Further evidence for the absence of polyproline ii stretch in the xao peptide," *Biophysical journal*, vol. 92, pp. 2904--2917, 2007.