J|A|C|S
ARTICLES
Published on Web 12/03/2003

β-Sheet Preferences from First Principles

Jan Rossmeisl, Iben Kristensen, Misha Gregersen, Karsten W. Jacobsen, and Jens K. Nørskov*

Contribution from the Center for Atomic-scale Materials Physics, Technical University of Denmark, DK-2800 Lyngby, Denmark

Received May 6, 2003; E-mail: norskov@fysik.dtu.dk

Abstract: The natural amino acids have different preferences of occurring in specific types of secondary protein structure. Simulations are performed on periodic model β-sheets of 14 different amino acids, at the level of density functional theory, employing the generalized gradient approximation. We find that the statistically observed β-sheet propensities correlate very well with the calculated binding energies. Analysis of the calculations shows that the β-sheet propensities are determined by the local flexibility of the individual polypeptide strands.

Introduction

The protein sequence in principle contains all the information necessary for the protein to fold into its three-dimensional structure. It is, however, enormously complicated to predict the three-dimensional structure from the sequence. Ultimately, the structure must be determined by the interactions between different parts of the polypeptide, but it is difficult to reliably measure or calculate interactions, and the link between the chemical interactions and the structure is complicated by the competition between energetic and entropic effects.

We focus in the present paper on a description of the interactions leading to the formation of β-sheets. The different amino acids have been observed to have considerably different propensities. Several authors have sampled β-sheet propensities,[1−4] and a number of attempts have been made to understand them. Experimental studies of changes in protein stability as specific amino acids are systematically replaced have shown a correlation between stability and propensity of different amino acids,[5−7] although later experiments have indicated that the stability is dependent on the model proteins and the specific site of the substitution.[8] Many physical models have been invoked to explain the differences in propensities. It has been suggested that they are related to the free-energy cost of confining the polypeptide strands to the β-sheet region of the Ramachandran plot,[9] the ability of the side chains to interfere with the hydrogen bonds between the solvent and the backbone,[10] and the side chains modulating electrostatic screening by the solvents of the backbone.[11,12] There is, however, no rigorous description of the bonding allowing an unambiguous understanding of the phenomenon. Quantum mechanical simulations on model α-helixes and β-sheets have previously been performed investigating properties of the hydrogen bonds (see, for example, ref 13) but the focus has not been on the propensities.

In the present paper, we calculate interaction energies between different amino acid residues in polypeptides using first principles density functional theory (DFT). We show that they correlate directly with statistic β-sheet propensities derived from experimental structures.[1−4] We establish in this way a link between interaction energies based on quantum mechanics and protein structure. We use the calculations to identify the origin of the differences in interaction energies for different residues, thus supporting the notion that propensities are directly given by interaction energies. The first principles calculations also allow for an in-depth understanding of the origin of the variations in bond strength. We will show that the main factor is the flexibility of the individual polypeptide strands.

Methods

The calculations are performed at the level of DFT,[14] where exchange-correlation effects are described using a nonlocal generalized gradient approximation (GGA) functional.[15] DFT at this level does not rigorously include the long-range dispersion interactions, but the exchange-correlation functional employed in the present work has been shown to give a good description of various types of hydrogen bonds.[16,17] We use the plane wave, pseudopotential code *Dacapo*,[18] with

(1) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13*, 211.
(2) Chou, P. Y.; Fasman, G. D. *Adv. Enzymol.* **1978**, *47*, 45.
(3) Chakrabarti, P.; Pal, D. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1.
(4) Fasman, G. D. In *Prediction of Protein Structure and the Principles of Protein Conformation*; Fasman, G. D., Ed.; Plenum Publishing: New York, 1989.
(5) Kim, C. A.; Berg, J. M. *Nature* **1994**, *362*, 267.
(6) Minor, D. L.; Kim, P. S. *Nature* **1994**, *367*, 660.
(7) Smith, C. K.; Withka, J. M.; Regan, L. A. *Biochemistry* **1994**, *33*, 5510.
(8) Minor, D. L.; Kim, P. S. *Nature* **1994**, *371*, 264.
(9) Street, A. G.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9074.
(10) Bai, Y.; Englander, S. W. *Proteins* **1994**, *18*, 262.
(11) Avbelj, F. *J. Mol. Biol.* **2000**, *300*, 1335.
(12) Avbelj, F.; Moult, J. *Biochemistry* **1995**, *34*, 755.
(13) (a) Ireta, J.; Neugebauer, J.; Scheffler, M.; Rojo, A.; Galvàn, M. *J. Phys. Chem. B* **2003**, *107*, 1432. (b) Wu, Y. D.; Zhao, Y. L. *J. Am. Chem. Soc.* **2001**, *123*, 5313. (c) Zhao, Y. L.; Wu, Y. D. *J. Am. Chem. Soc.* **2002**, *124*, 1570.
(14) Kohn, W. *Rev. Mod. Phys.* **1999**, *71*, 1253.
(15) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671.
(16) Tsuzuki, S.; Lüthi, H. P. *J. Chem. Phys.* **2000**, *114*, 3949.
(17) Hamann, D. R. *Phys. Rev. B* **1997**, *55*, 10157.
(18) The DFT calculations were performed with the Dacapo pseudopotential code, which is available at http://www.fysik.dtu.dk/campos.

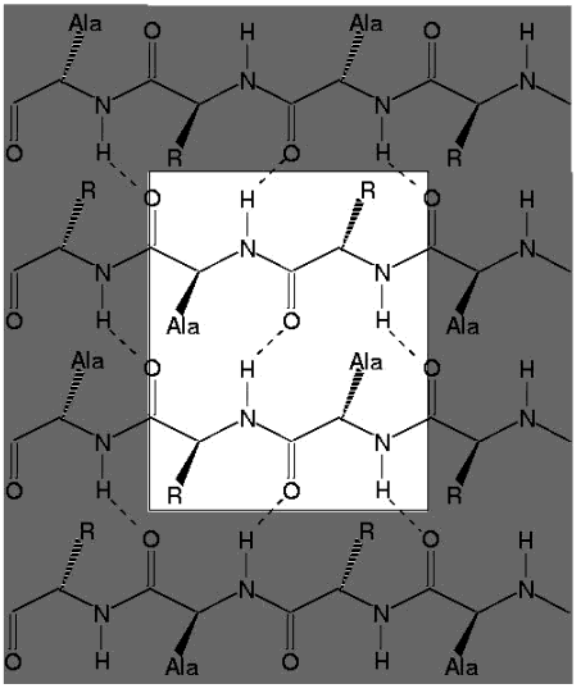

Figure 1. Schematic drawing of the model β-sheet. The strands are parallel and contain alternating alanine and R side chains in both directions of the sheet. Periodic boundary conditions are applied, and the white box indicates the simulation unit cell. The length of the unit cell is fixed at 15 Å perpendicular to the sheet, and the two remaining dimensions of the unit cell and all atomic coordinates are relaxed in the simulations. Hydrogen bonds are shown as dashed lines. R is one of the residues: Gly, Ser, Ala, Asn, Thr, Cys, Met, Gln, Tyr, Phe, Trp, Leu, Ile, Val.

periodic boundary conditions. The advantage of the plane wave basis set is that convergence is controlled by a single parameter, the cutoff energy. By setting the cutoff energy sufficiently high, basis set convergence can be ensured. In this work, the cutoff energy is 340 eV. The density is described on a grid corresponding to an energy cutoff at 1000 eV. The calculations employ ultra-soft pseudopotentials,[19] a Fermi smearing of 0.001 eV, and Pulay mixing is used to obtain the self-consistent electron density.[20] For sampling the k-space, we use 2 k-points in each direction of the sheet and one k-point in the direction perpendicular to the sheet. All atomic structures are relaxed using a conjugated gradient algorithm until the square root of the sum of the absolute forces is less than 0.05 eV/Å.

The theoretical modeling offers the possibility of constructing systems where specific effects can be singled out. In the present case, we would like to study trends in interaction energies between different amino acids situated in a β-sheet keeping the context the same. We do that by studying an idealized periodic β-sheet consisting of only two kinds of amino acids at a time as shown in Figure 1. Each strand contains one alanine residue followed by the residue R in an infinite sequence −Ala-R−Ala-R-. The strands are placed parallel to each other such that residue R in the sheet is surrounded by alanine residues on both sides. We now keep the alanine amino acids and systematically change the other. To test that the results do not depend on the "spacer" amino acid, we have exchanged alanine by valine and find that the main results are unchanged.

The model has the additional nice feature that it allows for periodic boundary conditions and thus for the use of efficient algorithms for solving the Schrödinger equation. The simulations are only performed on one period of the sheet in this work restricted to be two residues long (Figure 1). Because of the periodic boundaries, the β-sheets are

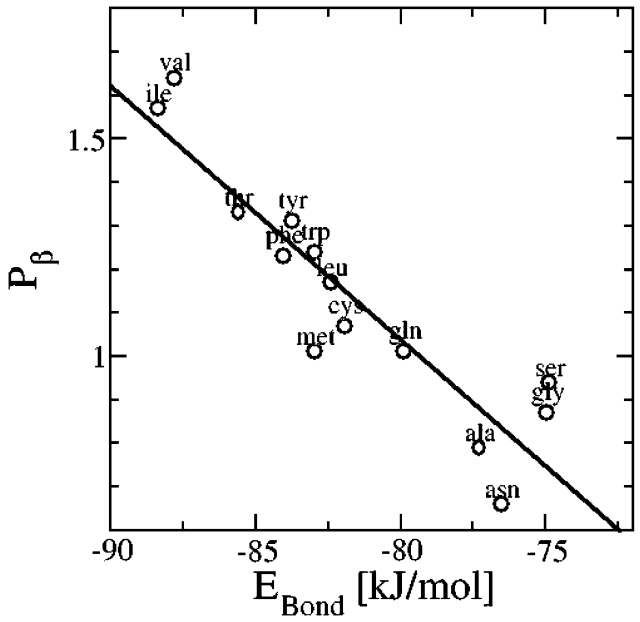

Figure 2. Measured β-sheet propensity, P_β,[4] plotted as a function of the calculated binding energy ΔE. ΔE is the binding energy per unit cell; each cell contains four peptide hydrogen bonds (Figure 1).

infinitely long and broad. This means that effects due to terminations do not influence the results. The length and width of the unit cell in the plane of the sheet are varied until the equilibrium size is obtained. The size is slightly dependent on the side chain; however, in most cases a length of the unit cell along the strands is 6.8−7.0 Å and the width is 9.6−9.8 Å. The periodic boundary conditions make it possible to perform simulations using an accurate description of the interatomic interaction on a realistic system, the only drawback being that the periodicity imposes constraints on the structure. Within the periodicity of two residues no long-range structural effects such as twisting can be described; however, the effect of twisting on the binding energy has been shown to be small.[21]

We include 14 different amino acids, excluding only the ionic amino acids and proline. The ionic side chains are excluded because we do not include the solvent in the present set of calculations. It is in principle possible to also include solvent molecules, but at present it is too computationally demanding. Proline is excluded because it cannot form interstrand hydrogen bonds. The binding energy is calculated as $\Delta E = E_{\text{Sheet}} - 2E_{\text{Strand}}$, where E_{Sheet} is the total energy per calculation cell of the relaxed sheet and E_{Strand} is the energy of the reference system of the relaxed single strand. The relaxations of both the single strands and the sheets are complicated by many local minima. To search the conformational space, the calculations are started in different initial conformations and the most stable conformations found are used. The most stable rotamer in the simulations of the sheets is t except for valine, isoleucine, and phenylalanine, where it is a $g+$ conformation. These rotamers are also frequently observed in real parallel β-sheets.[3] The different rotamers of the single strands are, in general, close in energy.

Results and Discussion

We find that our calculated binding energies correlate very well with the statistically observed β-sheet propensities, P_β[4] (Figure 2). The calculations are *first principles* in the sense that the main approximation is a general one about the treatment of electronic exchange and correlation effects and no input is made about the systems we are describing. The calculations are thus completely unbiased, and the good correlation in Figure 2 points

(19) Vanderbilt, D. *Phys. Rev. B* **1990**, *41*, 7892.
(20) Kresse, G.; Furthmüller, J. *J. Comput. Mater. Sci.* **1996**, *6*, 15.
(21) Rossmeisl, J.; Hinnemann, B.; Jacobsen, K. W.; Nørskov, J. K.; Olsen, O. H.; Pedersen, J. T. *J. Chem. Phys.* **2003**, *118*, 9783.

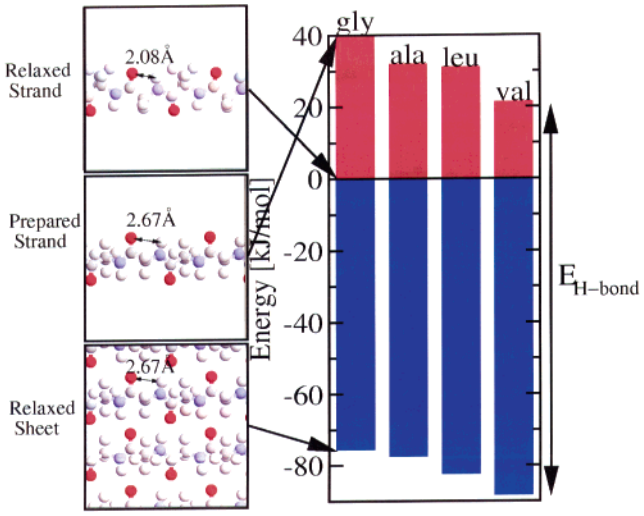

Figure 3. Binding energy, ΔE, indicated with the blue bars and the conformational energy penalty, E_{Conf}, indicated with the red bars for the amino acids glycine, alanine, leucine, and valine. $E_{\mathrm{H-bond}}$ for valine is shown with the arrow to the right. The structures for glycine are shown to the left with the O_i-H_i distance sketched; this local distance is changed significantly from the structure of single strand to the sheet structure, reflecting the change in the conformational energy. The structures are from the top: the relaxed structure of the single strand, the conformation of the strand prepared in the structure of a strand situated in the sheet, and the relaxed structure of the parallel sheet.

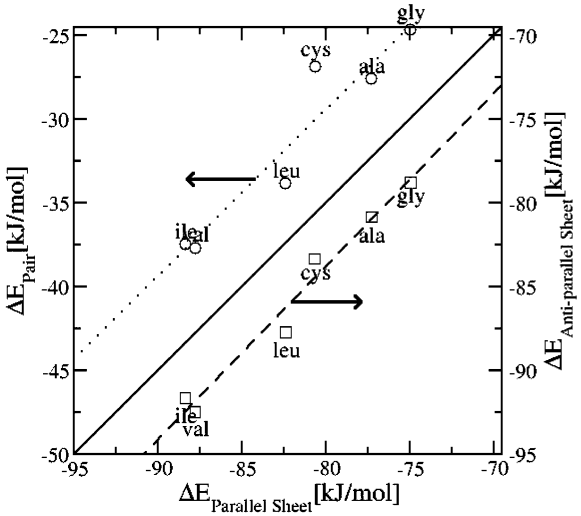

Figure 4. The binding energy of the interacting pair of strands (the dotted line is the best linear fit to the circular points) and the binding energy of antiparallel β-sheet (the dashed line and square points) as functions of the binding energy in parallel sheets. The orientations of the side chains are similar for all three model sheets (g^+ for leucine and cysteine and t for isoleucine and valine). Calculations have only been performed on a subset of the amino acids in Figure 2, but the extremes are included.

directly at the difference in binding energies as the main driving force in the determination of the β-sheet propensities.

The construction we have made deals with the uncertainty about the influence of surroundings (context). In the simulations, the context is always alanine residues, which is often used as a reference amino acid. The unbiased context makes our simulations suitable for comparison with the observed statistical propensities.

We note that the correlation coefficient, r, between the calculated binding energy and the statistical propensities is essentially independent of the basis set of proteins used to obtain the statistical propensities: ($r = -0.91$, slope $= -0.06$ (kJ/mol)$^{-1}$ [4], $r = -0.91$, slope $= -0.08$(kJ/mol)$^{-1}$ [3], $r = -0.93$, slope $= -0.08$(kJ/mol)$^{-1}$ [2]).

Having established that different amino acids have different binding energies and that this is directly observable in the propensities, we now turn to the question of the origin of these differences. The side chains affect the bond strength between amino acids along the polypeptide chain, even if the side chains are not directly taking part in the interstrand interaction. To understand this, we split the interaction energy, ΔE, into two terms: the energy, E_{Conf}, which is required to change the conformation of the strands into the one it assumes when it is situated in the sheets and the hydrogen bond energy, $E_{\mathrm{H-bond}}$:

$$\Delta E = E_{\mathrm{Conf}} + E_{\mathrm{H-bond}}$$

We start by splitting the sheet into the individual strands keeping the ionic positions fixed. The total energy of this structure is higher than the total energy of the relaxed single strand by the amount E_{Conf}. We find that trends in ΔE are reflected in variations of E_{Conf}, while $E_{\mathrm{H-bond}}$ is essentially independent of the system (Figure 3). A high value of E_{Conf} results in a weak bond. E_{Conf} is larger for the strands containing small side chains compared to the strands with larger side chains because the large side chains prevent the single strands from relaxing to the extent possible with smaller side chains. Our full first principles calculations thus provide strong evidence to the notion first introduced by Street and Mayo on the basis of a simple, semiempirical model[9] that it is the local interaction between the side chains and the backbone which is responsible for the variations in the ability of different amino acids to form β-sheets.

We have focused until now on infinitely broad, parallel β-sheets. On the other hand, the observed propensities to which we compare include both broad and narrow and parallel and antiparallel sheets. We have therefore investigated whether the trends in the interaction energies depend on the number of strands and on the orientation. We find that this is not the case (Figure 4). To this end, we perform simulations on a two-strand parallel β-sheet model and on an infinitely broad antiparallel sheet. The interacting pair of strands shows the same relative stability as the infinite sheet although the number of hydrogen bonds per unit cell is only half of what it is in an infinite sheet. The absolute binding energies are more than a factor of 2 weaker, though, reflecting the cooperative nature of β-sheets: the binding energy for an additional strand is larger for broad sheets compared to narrow sheets.[22−26] Similarly, the trends are the same for antiparallel and parallel sheets except that the former are generally about 0.04 eV more stable than the latter. This is in agreement with the fact that antiparallel sheets are observed more frequently than parallel ones.[27] The above examples illustrate that the trends in the relative propensity are conserved for different hydrogen bond geometries.

The simulations presented here do not include effects of solvation and entropy. Both effects have previously been proposed as the origin of β-sheet propensities.[9−11] These effects

(22) Kortemme, T.; Ramirez-Alvarado, M.; Serrano, L. *Science* **1998**, *281*, 253.
(23) Sharmann, G. J.; Searle, M. S. *J. Am. Chem. Soc.* **1998**, *120*, 5291.
(24) Griffiths-Jones, S. R.; Searle, M. S. *J. Am. Chem. Soc.* **2000**, *122*, 8350.
(25) Schenck, H. L.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4869.
(26) Rossmeisl, J.; Nørskov, J. K.; Jacobsen, K. W., submitted for publication.
(27) Richardson, J. S. *Adv. Protein Chem.* **1981**, *34*, 167−339.

can possibly be important for the propensity, but the variations in binding energies calculated in this paper are an inherent feature of the different amino acids related to the conformational energy of the strands situated in sheets, which is not removed due to entropy or solvent effects.

Conclusion

We conclude that we can describe interactions in β-sheets directly from first principles density functional calculations. We have shown that the variations in the calculated bond strengths from one residue to the next can account for the observed propensities in β-sheets, and we have shown that the bond energy variations can be traced back to differences in the local rigidity of the individual strands making up the β-sheet.

Acknowledgment. We thank Jan Torleif Pedersen for critically reading the manuscript and for his many constructive comments. We acknowledge support from the Danish Center for Scientific Computing through Grant No. HDW-1101-05. CAMP is sponsored by the Danish National Research Foundation.

JA0359658