

POTENTIAL FUNCTIONS FOR HYDROGEN BONDS IN PROTEIN STRUCTURE PREDICTION AND DESIGN

By ALEXANDRE V. MOROZOV* AND TANJA KORTEMME†

*Center for Studies in Physics and Biology, Rockefeller University, New York, New York 10021
†Department of Biopharmaceutical Sciences and California Institute for Quantitative Biomedical
Research, University of California, San Francisco, San Francisco, California 94142

I. Introduction	2
II. Physical Mechanism of Hydrogen Bond Formation	5
III. Main Approaches to Modeling Hydrogen Bonds in Biomolecular Simulations	6
A. Potentials Derived from Hydrogen Bonding Geometries Observed in Crystal Structures	6
B. Molecular Mechanics: Comparison with the Structure-Derived, Orientation-Dependent Potential	9
C. Quantum Mechanics: Comparison with Molecular Mechanics and the Structure-Derived Potential	12
IV. Applications of Hydrogen Bonding Potentials	20
A. Protein Structure Prediction and Refinement	20
B. Prediction of Structures and Energetics of Protein-Protein Interfaces .	24
C. Protein Design	27
V. Conclusions and Perspectives	30
References	32

ABSTRACT

Hydrogen bonds are an important contributor to free energies of biological macromolecules and macromolecular complexes, and hence an accurate description of these interactions is important for progress in biomolecular modeling. A simple description of the hydrogen bond is based on an electrostatic dipole-dipole interaction involving hydrogen-donor and acceptor-acceptor base dipoles, but the physical nature of hydrogen bond formation is more complex. At the most fundamental level, hydrogen bonding is a quantum mechanical phenomenon with contributions from covalent effects, polarization, and charge transfer. Recent experiments and theoretical calculations suggest that both electrostatic and covalent components determine the properties of hydrogen bonds. Likely, the level of rigor required to describe hydrogen bonding will depend on the problem posed. Current models include knowledge-based descriptions based on surveys of hydrogen bond geometries in structural databases of proteins and small molecules, empirical molecular

mechanics models, and quantum mechanics-based electronic structure calculations. *Ab initio* calculations of hydrogen bonding energies and geometries accurately reproduce energy landscapes obtained from the distributions of hydrogen bond geometries observed in protein structures. Orientation-dependent hydrogen bonding potentials were found to improve the quality of protein structure prediction and refinement, protein-protein docking, and protein design.

I. INTRODUCTION

Accurate modeling of hydrogen bonding interactions is critical for progress in protein structure prediction, protein-protein docking, and protein design. While the large number of hydrogen bonds in proteins and protein interfaces underlines their importance, there may be no net gain in free energy for hydrogen bond formation in protein folding and binding; the formation of hydrogen bonds between protein atoms results in the loss of hydrogen bonds made with water. However, most polar groups in the nonsurface accessible interior of proteins form hydrogen bonds to satisfy their hydrogen bonding potential (Baker and Hubbard, 1984; McDonald and Thornton, 1994). These requirements result in considerable energetic and structural constraints and are in part responsible for the regular backbone-backbone hydrogen bonding patterns of α -helix and β -sheet regular secondary structure elements (Pauling and Corey, 1951). Similarly, hydrogen bonds, particularly side chain-side chain hydrogen bonds, are thought to play important roles in the specificity of macromolecular interactions (Lumb and Kim, 1995; Petrey and Honig, 2000) and need to be taken into account in the prediction of protein interaction preferences. Hydrogen bonds may be crucial for enabling a unique three-dimensional protein conformation or binding mode in protein design applications (Looger *et al.*, 2003; Lumb and Kim, 1995).

What is needed for an accurate description of hydrogen bonding interactions within and between proteins? The physical nature of hydrogen bonds is complex, and calculation of electrostatics, polarization, exchange repulsion, charge-transfer, and coupling contributions to hydrogen bonding energetics (Kollman, 1977; Morokuma, 1971; Singh and Kollman, 1985; Umeyana and Morokuma, 1977) from first principles is not straightforward for biological macromolecules. Likely, the level of rigor required to explain certain molecular properties in question will depend on the problem posed. Which simplifications can be made in which context? An example discussed in detail in this chapter is the orientation dependence of hydrogen bonds, which has been a subject of considerable debate. An

electrostatic dipole–dipole model of a hydrogen bond would predict a linear arrangement of the donor and acceptor dipoles. However, a “lone pair” concept would imply directionality of the hydrogen bond (Fig. 1a). What are the structural and energetic characteristics of hydrogen bonds in protein structures and how can a model be devised that reproduces them?

Any simplified description of hydrogen bonds in biological molecules needs to be tested by comparing its predictions against a large body of

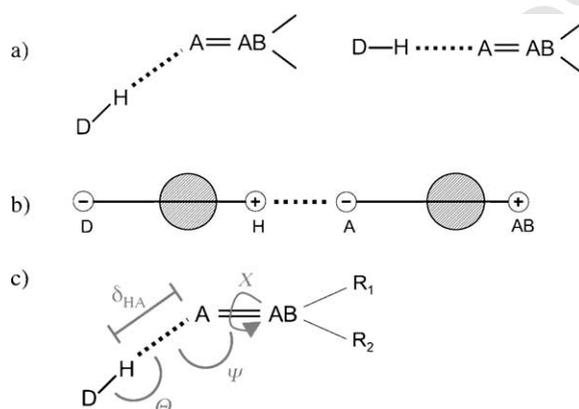


FIG. 1. Mechanism and orientation dependence of hydrogen bond formation. A, acceptor; D, donor; H, hydrogen; AB, acceptor base. (a) Orientation dependence of hydrogen bond formation. Hydrogen bond formation along lone-pair directions would predict hydrogen bonding geometries such as the one shown on the left, whereas an electrostatically dominated mechanism based on a dipole–dipole interaction (see b) would favor the arrangement on the right. (b) Simple description of hydrogen bonding interactions as the interaction of two dipoles with atom-centered partial point charges. Shaded spheres represent electron density shifted along the H–D and AB–A covalent bonds toward more electronegative atoms, resulting in the appearance of partial charges on all four atoms. (c) Schematic representation of hydrogen bond geometry. D, donor atom; H, hydrogen atom; A, acceptor atom; AB, acceptor base; R_1 , R_2 , atoms bound to the acceptor base. Geometric parameters describing the hydrogen bond are as follow: δ_{HA} (Å), distance between hydrogen and acceptor atoms; ψ (degree), angle at the acceptor atom; θ (degree), angle at the hydrogen atom; X (degree), dihedral angle around the A–AB axis. As hydrogen atoms are generally not included in the coordinates derived from X-ray crystallographic data, polar hydrogen atoms were added in cases where the position of the hydrogen atom was given by the chemistry of the donor group (backbone amide protons and side chains donor groups of tryptophan, histidine, asparagine, and glutamine residues). For the derivation of hydrogen bonding statistics, histidine, asparagine, and glutamine residues were taken in their crystallographic conformations; similarly, polar hydrogens associated with a rotatable bond were not considered, as they could not be placed without making assumptions about the hydrogen bonding geometry.

experimental data, preferably obtained from macromolecules. A direct comparison of predicted and observed hydrogen bonding energies in biological macromolecules is not straightforward because the individual components of the free energy cannot readily be measured independently in experiments. More feasible but less direct strategies rely on the vast information available on protein sequences and structures and use concepts from computational protein design, protein structure prediction, and protein–protein docking. The structure prediction and docking tests measure the discrimination of misfolded from native or near-native protein structures and the identification of correct relative orientations of protein partners in protein–protein complexes, respectively (Kortemme *et al.*, 2003; Morozov *et al.*, 2003). The tests are based on the assumption that native protein structures and protein–protein interfaces are lower in free energy than the vast majority of nonnative conformations. While it is not necessary for every individual contribution to the free energy (such as the hydrogen bonding component) to favor the native structure, it is plausible that given several alternative models of a certain energetic contribution, the one that most favors the native sequence and structure is the most accurate (Morozov *et al.*, 2003). In the protein design test, different potentials are evaluated by their ability to reproduce native amino acid sequences (Kuhlman and Baker, 2000). Given the constraints on protein structure imposed by hydrogen bonding requirements and the presumed role of side chain–side chain hydrogen bonds in protein interaction specificity, this test can be expected to be sensitive to different models of hydrogen bonding.

This chapter reviews approaches to describing hydrogen bonding interactions in biomolecular simulations and the applications of these methods to protein structure prediction, protein–protein docking, and design of proteins and protein-mediated interactions. We start with a brief description of the physical mechanism of hydrogen bond formation. We then discuss various simplifications made in modeling hydrogen bonding interactions using knowledge-based potentials, force field methods, and electronic structure calculations. Comparing these approaches, we have found a remarkable agreement between the orientation dependence of hydrogen bonds observed in protein structures and electronic structure calculations (Morozov *et al.*, 2004). We illustrate how a simple orientation-dependent hydrogen bonding potential derived from the geometric characteristics of hydrogen bonds in high-resolution structures can be tested in protein structure prediction, protein–protein docking, and protein design applications. Combining the generality and *ab initio* nature of quantum mechanical electronic structure calculations with more computationally efficient empirical models may help to approach remaining

important challenges in modeling protein hydrogen bonds, such as polarization effects causing nonadditivity in hydrogen bonding energies and proton transfer in biological catalysis.

II. PHYSICAL MECHANISM OF HYDROGEN BOND FORMATION

A simple description of the hydrogen bond is based on an electrostatic dipole–dipole interaction involving hydrogen-donor and acceptor–acceptor base dipoles (Fig. 1b), where the dipole moments depend only on the intrinsic electronegativity of the donor and acceptor atoms (i.e., on the affinity of those atoms for electron density). In this approximation, dipoles are represented by atom-centered partial charges, and more complex context-dependent effects are assumed to be negligible. For the hydrogen bond dipoles in Fig. 1b, the most favorable orientation is head to tail, resulting in linear hydrogen bonds with Ψ at 180° (Ψ is the angle at the acceptor atom; see Fig. 1c).

The simple dipole–dipole interaction picture does not account for cases in which optimal hydrogen bonding geometry deviates significantly from linearity. Such deviations from linearity could be rationalized on the basis of the electronic structure of the acceptor and hydrogen atoms, which reveals that representing hydrogen–acceptor interactions with point charges of fixed magnitude is a significant simplification. In a more detailed description, the spatially distributed charge density of the hydrogen atom interacts with the valence electron cloud on the acceptor and is thus sensitive to the number of valence electrons and their charge density distribution. For example, the “lone pairs” of a sp^2 hybridized oxygen atom are at positions corresponding to an angle Ψ at the acceptor atom of 120° , and hence hydrogen bonds with $\Psi = 120^\circ$ should be more favorable than hydrogen bonds with $\Psi = 180^\circ$. Likewise, sp^2 hybridized acceptors should exhibit variations in hydrogen bonding energies when the dihedral angle X around the acceptor–acceptor base bond is changed.

Concepts of hybridization states and lone pairs are themselves simplifications (McGuire *et al.*, 1972). The overall shape of valence electron orbitals on the hydrogen and acceptor atoms can change in a variety of ways. Formation of the hydrogen bond itself leads to altered electron density; in fact, polarization of the electron cloud around the hydrogen atom is one of the reasons for short hydrogen–acceptor distances. The electron density around the hydrogen atom is redistributed, minimizing the distance between the acceptor orbitals and the proton. This phenomenon imparts a partially covalent character to the hydrogen bond due to mixing of hydrogen and acceptor orbitals. Polarization of hydrogen bonding orbitals can also be affected by the presence of nearby charged groups

and hydrogen bonds and by an external electric field from the remainder of the molecule. The quantum mechanical description of hydrogen bond formation detailed later thus implies that some properties of hydrogen bonds cannot be represented accurately by electrostatic descriptions relying on a point charge model. The question remains, however, whether varying approximations of the physics of hydrogen bond formation can lead to useful biological predictions and which level of theory is required for which application. The following sections illustrate approaches to modeling hydrogen bonds in biomolecular simulations and their application to protein structure prediction, protein-protein docking, and protein design.

III. MAIN APPROACHES TO MODELING HYDROGEN BONDS IN BIOMOLECULAR SIMULATIONS

A. Potentials Derived from Hydrogen Bonding Geometries Observed in Crystal Structures

Structure-derived potential functions (for a review, see Jernigan and Bahar, 1996) have been popular in protein modeling, in part due to their simplicity and computational speed compared to more sophisticated models of detailed balances between physical forces. While most of these structure-derived approaches classify interactions based on identities of pairs of amino acid side chains or types of atoms, specific potentials describing hydrogen bonding interactions (Fabiola *et al.*, 2002; Gavezzotti and Filippini, 1994; Grishaev and Bax, 2004; Grzybowski *et al.*, 2000; Kortemme *et al.*, 2003) have also been developed.

In general, derivation of such “knowledge-based potentials” involves conversion of experimentally observed frequency distributions of certain features in the protein structure database (i.e., spatial proximity of positively and negatively charged side chains) into pseudo-energies by assuming a Boltzmann distribution over those features. Caveats inherent in this approach have been pointed out. First, Boltzmann statistics apply to a single closed system at a fixed temperature that can populate different energy levels, but sets of unrelated small molecule or protein structures solved under different experimental conditions are not in thermodynamic equilibrium with each other. Nonetheless, statistically derived energies can correlate with experiment, for example, for exterior-interior partition energies that match experimental water-octanol transfer energies (Thomas and Dill, 1996). Second, the assumption that different statistical terms derived for pairwise interactions are independent from each other

may not be valid (Thomas and Dill, 1996). Third, commonly used potentials differ in their “reference state” (in which interactions are assumed to be absent). The particular reference state chosen can significantly influence predictions; for example, nonphysical properties of knowledge-based potentials such as long-range repulsions between hydrophobic residues may be eliminated by the choice of the reference state (Zhang *et al.*, 2005).

Grzybowski *et al.* (2000) argued that inversion of frequency distributions to obtain pseudo-energies is justified theoretically for a set of molecules frozen in low energy states, where the total energy is the sum of many independent contributions that are functions of some parameter p ; in such ensembles, the negative logarithm of the observed frequency of the value of p is proportional to the interaction energy for that value of p , even though the set of molecules is not in thermal equilibrium (Grzybowski *et al.*, 2000). The short-range geometric features of hydrogen bonds in crystal structures may be a reasonably good example of such a dataset (Grishaev and Bax, 2004; Grzybowski *et al.*, 2000; Kortemme *et al.*, 2003).

Another potential shortcoming of database-derived potentials is the question of transferability if relevant parameters of a potential are derived for a specific system. While the physical principles governing interactions should be the same for all classes of molecules, the details may be different. For these reasons, different energy functions based on hydrogen bond geometrical parameters were derived from sets discriminated by the chemical characteristics of hydrogen bonded groups (e.g., ester versus amide), by the type of system (small molecule or protein), or by structural criteria (backbone–backbone hydrogen bonds in α helices and β sheets of proteins versus side chain–side chain hydrogen bonds) (Fabiola *et al.*, 2002; Gavezzotti and Filippini, 1994; Grishaev and Bax, 2004; Grzybowski *et al.*, 2000; Kortemme *et al.*, 2003). Potentials also differ in the choice of geometric parameters used to describe hydrogen bonds (e.g., see Fig. 1c) and whether multidimensional potentials were extracted that take parameter dependencies into account (Grishaev and Bax, 2004; Grzybowski *et al.*, 2000).

Figure 2 shows the distributions of four geometrical parameters of hydrogen bonds observed for side chain–side chain interactions involving sp^2 hybridized acceptor groups in high-resolution protein crystal structures (Kortemme *et al.*, 2003): (a) the distance δ_{HA} between the hydrogen atom and the acceptor atom, (b) the angle Ψ at the acceptor atom, (c) the angle θ at the hydrogen atom, and (d) the dihedral angle X corresponding to rotation around the acceptor–acceptor base bond (Fig. 1c depicts geometrical degrees of freedom used to describe a hydrogen bond). The protein hydrogen bonding geometry distributions shown in Fig. 2 are similar to those compiled on a smaller protein set in the classic paper by Baker and

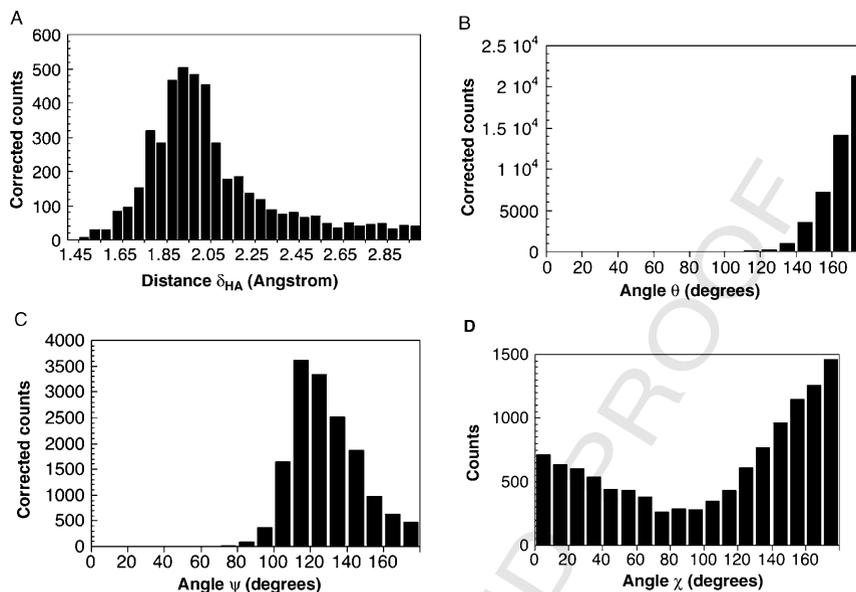


FIG. 2. Distributions of hydrogen bonding geometric parameters obtained from 698 protein crystal structures for side chain–side chain hydrogen bonds involving an sp^2 hybridized acceptor (Kortemme *et al.*, 2003). Corrected counts take into account the different volume elements encompassed by the angular bins for psi and theta using a $\sin(\text{angle})$ angular and a $(\text{distance})^2$ distance correction. Hydrogen bond geometric parameters are defined in Fig. 1c.

Hubbard (1984). For side chain–side chain hydrogen bonds with sp^2 hybridized acceptor atoms, the angle at the hydrogen is largely linear (after proper correction for the geometric bias at angles close to linearity). Distribution of the angle at the acceptor atom peaks around 120° , as expected based on simple considerations of lone pair geometries at the acceptor atom. Similar dependencies are observed for hydrogen bonds to sp^3 hybridized acceptor atoms, with a slightly sharper distribution for the Ψ -angle acceptor (but not shifted to significantly lower angles). The geometries of backbone–backbone hydrogen bonds differ significantly from those of side chain–side chain hydrogen bonds: the angle at the hydrogen is not predominantly linear and the angle at the acceptor is shifted from 120° to larger angles. Distributions also vary between different secondary structures (Kortemme *et al.*, 2003). A likely explanation is that the formation of regular secondary structures imposes steric constraints, causing hydrogen bond geometries to deviate from ideal values. Deviations of the

geometries of backbone–backbone hydrogen bonds from those of less sterically constrained hydrogen bonds may have consequences for hydrogen bonding energetics in α helices and β sheets.

We derived an empirical orientation-dependent hydrogen bonding potential from the negative logarithm of the observed hydrogen bonding distributions (Kortemme *et al.*, 2003). We assumed independence of the different geometric parameters to generate one-dimensional landscapes relating the distance and angle distributions to energetic variations. An exception is the distance dependence of the angular terms for side chain–side chain hydrogen bonds (Kortemme *et al.*, 2003) that was also noticed for small molecules (Grzybowski *et al.*, 2000) and backbone–backbone hydrogen bonds in proteins (Lipsitz *et al.*, 2002). Moreover, the relative scaling of structure-based statistical potentials may be complicated. Thomas and Dill (1996) pointed out that the temperature parameter of the Boltzmann distribution is not necessarily a single parameter for all substructures in proteins, which would affect the relative strength of hydrogen bonding interactions.

B. Molecular Mechanics: Comparison with the Structure-Derived, Orientation-Dependent Potential

Most modern molecular mechanics (MM) force fields (for a recent account of force fields, see Ponder and Case, 2003) rely on the combination of Coulomb and Lennard–Jones interactions to model hydrogen bonds implicitly, using a relationship similar to Eq. (1) to describe all nonbonded interactions:

$$V(r) = \sum_{\text{nonbonded}} \left\{ \frac{q_i q_j}{\epsilon r_{ij}} + d_{ij} \left[\left(\frac{R_{\min, ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{\min, ij}}{r_{ij}} \right)^6 \right] \right\}, \quad (1)$$

where r_{ij} is the distance between atoms i and j , q_i and q_j are atomic partial charges, ϵ is the effective dielectric constant (which may be distance dependent in some force fields), and d_{ij} and $R_{\min, ij}$ are functions of the well depth and the distance at the minimum of the Lennard–Jones 6–12 empirical potential, respectively (see, e.g., MacKerrell *et al.*, 1998). The sum in Eq. (1) includes all nonbonded atoms. The first term in Eq. (1) describes electrostatic interactions, whereas the second and the third terms provide an empirical description of exchange repulsion and attractive van der Waals interactions. Even though the functional form of the potential energy is quite simple, it depends on a large number of empirical parameters, which must be obtained from *ab initio* electronic structure calculations on small molecules and/or experimental data such as densities and

enthalpies of vaporization of organic liquids. A set of parameters together with the functional form of the potential constitute a specific force field. Differences in parameterization strategies and input experimental data lead to alternative sets of force field parameters that are similar but not mutually transferable. Furthermore, *a priori* emphasis on specific aspects of simulations, such as the ability to reproduce certain gas phase or condensed phase properties, means that each force field has a range of molecular systems and molecular properties for which its application is suitable and thoroughly tested.

The simple model of a hydrogen bond described by the combination of Coulomb and electrostatic interactions [Eq. (1)] without an explicit hydrogen bonding description goes back to early work on MM simulations (Hagler and Lifson, 1974; Hagler *et al.*, 1974). Because each new term in the MM potential function requires additional empirical parameters, it is quite appealing to keep the functional form of the potential function as simple as possible. While most widely used current force fields such as AMBER, OPLS, and CHARMM (Cornell *et al.*, 1995; Jorgensen *et al.*, 1996Z; MacKerrell *et al.*, 1998) do not employ explicit hydrogen bonding terms, this was not always the case. For example, the original AMBER potential function published in 1984 (Weiner *et al.*, 1984) included a Lennard–Jones-like 10–12 function for the description of hydrogen bonding energies:

$$\sum_{H-bonds} \left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right]. \quad (2)$$

This potential does not have any terms describing angular dependencies of hydrogen bonds and is similar to the 10–12 hydrogen bonding potential originally proposed by McGuire *et al.* (1972). They found that hydrogen bonding energies were represented adequately by a sum of Lennard–Jones and electrostatic interactions plus the 10–12 hydrogen bonding term with empirical constants adjusted according to the hydrogen bond type. This notion was supported by CNDO/2 *ab initio* calculations [an approximation to Hartree–Fock (HF) theory] on hydrogen bonded dimers of small molecules. Explicit orientation dependence of hydrogen bonding energies was omitted on the grounds that it can be reasonably well reproduced as a sum of *all* distance-dependent interatomic interactions in hydrogen bonded dimers. Thus orientation dependence would be enforced by “nonlocal” interactions involving atoms other than the donor-hydrogen and acceptor–acceptor base dipoles (see Fig. 1). Because the functional form of such a hydrogen bonding term was very close to the Lennard–Jones component of the force field, the second-generation AMBER force field omitted it

altogether (Cornell *et al.*, 1995), relying instead on the combination of Lennard–Jones and Coulomb interactions to model hydrogen bonded complexes.

This point of view was supported in a study by No and co-workers (1995), who argued on the basis of *ab initio* molecular orbital calculations with the 6–31G** basis set that the angular dependence is important in an empirical hydrogen bonding function, but can be modeled with a 6–12 type potential involving 1–3 atomic pairs.

Similarly, the widely used OPLS force field does not contain an explicit hydrogen bonding term: the emphasis of OPLS parameterization is on reproducing thermodynamic properties of organic liquids such as enthalpies of vaporization, densities, and free energies of hydration (Jorgensen and Tirado-Rives, 1988; Jorgensen *et al.*, 1996). No special hydrogen bonding functions were found to be necessary to describe these properties in molecular simulations. The original CHARMM potential function (Brooks *et al.*, 1983) had a dedicated hydrogen bonding term in which a Lennard–Jones-like potential between donor and acceptor atoms was modulated by a $\cos^m(\theta)\cos^n(\Psi)$ function, where θ is the angle at the hydrogen atom and Ψ is the angle at the acceptor atom (Fig. 1c). The exponent m was determined by the donor atom type, and the exponent n was determined by the acceptor atom type. This explicit hydrogen bonding term was subsequently dropped (Neria *et al.*, 1996; MacKerrell *et al.*, 1998) based on *ab initio* calculations carried out in the Karplus group (Reiher, 1985). Reiher compared CHARMM energies with empirically scaled HF energies and concluded that (1) hydrogen bonding was described adequately by the sum of Coulomb and Lennard–Jones interactions with refined van der Waals parameters and atomic partial charges and (2) the explicit hydrogen bonding term with the cosine-based angular factor was no longer necessary to describe hydrogen bonding energetics.

The challenge of reproducing the directional character of hydrogen bonds with empirical potentials was taken into account by the developers of the MM3 force field (Allinger, 1989; Lii and Allinger, 1994, 1998). The MM3 force field contains an orientation-dependent hydrogen bonding term motivated by the molecular orbital picture of hydrogen–acceptor interactions. Including this term into the potential function resulted in the substantial improvement of MM3 predictions of energies and geometries of hydrogen bonded complexes with respect to the results from *ab initio* calculations. The *ab initio* calculations of small molecule hydrogen bonded complexes were carried out at the 6–31G** MP2 level, and MM3 hydrogen bonding parameters were subsequently reoptimized to fit the *ab initio* results.

Likewise, the DREIDING force field developed by Mayo *et al.* (1990) includes an explicit orientation-dependent hydrogen bonding term modeled as a product of a 10–12 Lennard–Jones-like potential with a $\cos^4(\theta)$ angle-dependent factor. Because this potential depends only on the angle at the hydrogen atom, it can lead to nonphysical hydrogen bonding geometries, as all angles at the acceptor atom are equally allowed. The authors recognized this problem and modified the hydrogen bonding potential, making it a function of additional angles (including the angle at the acceptor atom) in a hybridization-dependent manner (Gordon *et al.*, 1999). The modified hydrogen bonding potential became a part of a force field developed explicitly for computational protein design.

In general, the assignment of single point partial charges to hydrogen bonded atoms leads to intrinsic preference for linear hydrogen bonds because of the interaction between the donor-hydrogen and the acceptor–acceptor base dipoles. The idea that orientation dependence of hydrogen bonds could be rescued by adjusting empirical parameters of both hydrogen bonded atoms and neighboring covalently bound atoms suggests a more nonlocal picture of hydrogen bond formation (Buck and Karplus, 2001). The orientation dependence of hydrogen bonding energies modeled as a function of empirical parameters in the nonlocal picture (i.e., involving atoms other than donor, hydrogen, acceptor, and acceptor base) may be distorted when these parameters are refitted for the next generation of the force field.

C. Quantum Mechanics: Comparison with Molecular Mechanics and the Structure-Derived Potential

Quantum mechanical (QM) electronic structure calculations can, in principle, provide the most fundamental way of describing the subtle physical phenomena associated with hydrogen bonding interactions in macromolecular systems. Because high-level QM description of hydrogen bonds in their biomolecular setting is currently impossible due to computational limitations on the size of the molecules, model hydrogen bonded systems based on small molecule analogs have to be studied instead. Despite their limitations, these studies may serve to address the question of the relative importance of local versus nonlocal effects in hydrogen bonding geometries and energetics. Electronic structure calculations on simple hydrogen bonded model systems will describe local interactions in the absence of the complex context dependence found in macromolecular structures. Comparison with hydrogen bonding properties in experimentally determined structures of proteins may then shed light on the macromolecular context dependence, as described later. Moreover,

extension of simple model systems to include higher order cooperative effects can also be carried out using electronic structure methods.

There is a wide variety of quantum chemistry methods available for computing energies of hydrogen bonded systems. One standard approach is based on predicting molecular energies and geometries with the Hartree–Fock (HF) self-consistent field method (Szabo and Ostlund, 1982). By construction, HF theory neglects explicit electron–electron correlations but includes exact exchange interactions. In some cases, the HF approximation leads to quantitatively and even qualitatively inaccurate predictions that can be improved upon using multiconfigurational wave functions, pair and coupled-pair theories, or the many-body Moller–Plesset perturbation theory (Szabo and Ostlund, 1982). The many-body Moller–Plesset perturbation theory (commonly abbreviated as MPX, where X is the order of the highest perturbative correction) is the most computationally efficient, especially if the perturbative expansion is truncated at the lowest MP2 level. The accuracy of MP energies depends primarily on the quality of the basis set used in the calculation and on the order at which the perturbative series is truncated.

An alternative approach to computing energies and geometries of hydrogen bonded systems is based on density functional theory (DFT) (Parr and Yang, 1989). DFT takes advantage of the fact that all ground-state properties of a molecular system, particularly its energy, are a function of only the electron density with 3 degrees of freedom rather than the full many-body wave function with $3N$ degrees of freedom for N electrons. This allows for a formulation of the theory, which is no more computationally demanding than HF, but can, in principle, take all electron–electron correlations and exchange into account. However, the exact form of the “correlation and exchange” contribution to the energy density functional is unknown, and various approximate exchange-correlation functionals have to be constructed. The accuracy of such functionals has to be verified explicitly using diverse sets of molecules, and predicted energies have to be compared with MP and other molecular orbital methods and experimental measurements. In the case of hydrogen bonds, testing against other *ab initio* methods and experimental data shows that DFT methods are capable of reproducing hydrogen bonding energies with reasonable accuracy (Kaschner and Hohl, 1998; Topol *et al.*, 1995; Tuma *et al.*, 1999). In particular, Topol *et al.* (1995) carried out DFT energy calculations for six hydrogen bonded dimers and demonstrated that experimental dimerization enthalpies were, in most cases, reproduced with discrepancy of about 1 kcal/mol. Kaschner and Hohl (1998) concluded by comparison with experiment and post-HF molecular orbital calculations that DFT with gradient-corrected (nonlocal) exchange-correlation functionals was

a reliable method for calculating relative energies and geometries of isomers of glycine and alanine and their oligopeptides.

Hydrogen bonds observed in proteins and other biomolecules are characterized by a range of different orientations of donor and acceptor groups. Therefore, in order to model hydrogen bonds typically found in proteins it is not sufficient to sample the vicinity of one or several energy minima of a given hydrogen bonded model. Rather, a hydrogen bonding energy landscape needs to be constructed in which all geometric degrees of freedom are consistently varied and sampled. Moreover, the structural environment in which hydrogen bonds are formed, as well as competing contributions from other interactions, may influence hydrogen bonding energies and geometries. Even neglecting potential higher order context effects in proteins, the sampling problem becomes formidable when several degrees of freedom are involved. For example, sampling a full four-dimensional energy landscape of a single hydrogen bond for the four geometric parameters described in Fig. 1c would require at least 10^4 – 10^5 separate QM calculations. In order to simplify the problem, we have chosen a more practical approach based on creating one-dimensional projections of the complete multidimensional energy landscape. Each landscape projection corresponds to varying only one degree of freedom at a time. The initial dimer conformation for each projection is the minimum energy one, obtained by optimizing geometric positions of all nuclei on the energy landscape computed with the selected QM method (Fig. 3).

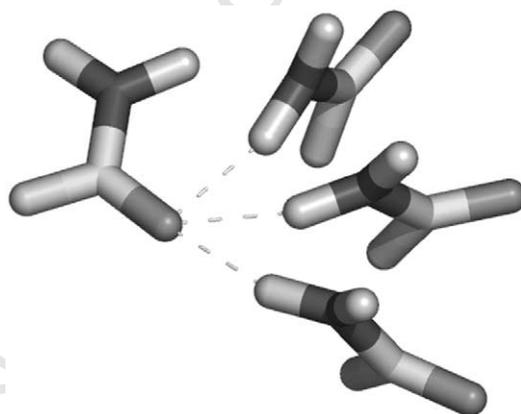


FIG. 3. Three representative conformations of a hydrogen bonded formamide dimer. The angle at the acceptor (Ψ) is varied to create a one-dimensional projection of the full hydrogen bonding energy landscape. Shown are conformations with $\Psi = 95$, 135 , and 175° . All other degrees of freedom are taken from the out-of-plane formamide dimer, which was optimized geometrically using DFT.

Different model systems have been used for quantum mechanical calculations of hydrogen bonding energy landscapes. Morozov *et al.* (2004) carried out a QM analysis using an out-of-plane formamide dimer as a model system for side chain–side chain hydrogen bonds. The formamide molecule can be viewed as the hydrogen bonded moiety of asparagine or glutamine truncated at $C\beta$ ($C\gamma$ for glutamine) and capped by a hydrogen atom. Using a methyl group instead of the hydrogen atom to cap the truncation site suggests acetamide as an alternative small molecule model of the hydrogen bonded side chain. Acetamide retains more of the asparagine side chain by replacing $C\alpha$ rather than $C\beta$ with a hydrogen; however, including the group may restrict the range of conformations an acetamide dimer can adopt, altering the energy landscape. Conformational restriction due to excluded volume plays an important role in *N*-methylacetamide (NMA), which forms almost linear hydrogen bonds in NMA dimers and NMA–formamide dimers, likely due to the repulsion caused by NMA methyl groups (Buck and Karplus, 2001; Guo and Karplus, 1992; Qian *et al.*, 1999). For example, using the CHARMM22 empirical energy function, Buck and Karplus found that the deviation from linearity at the acceptor angle is 18° for antiparallel and 0° for parallel NMA dimer configurations, in agreement with *ab initio* quantum mechanical results and other empirical calculations (Gao and Freindorf, 1997; Guo and Karplus, 1992, 1994; Torii *et al.*, 1998; Watson and Hirst, 2002). The NMA dimer may be a reasonable model for main chain hydrogen bonds, but side chain hydrogen bonds are probably better modeled with the formamide or acetamide dimer. A number of low-energy formamide dimer arrangements (parallel, antiparallel, and out of plane) have been described in the literature (Vargas *et al.*, 2001; Watson and Hirst, 2002). For comparison with protein side chain statistics, the out-of-plane formamide dimer with a single hydrogen bond (Fig. 3) is an ideal system: while the cyclic dimer conformation with two $N-H\cdots O=C$ hydrogen bonds is the global energy minimum of the formamide dimer (Vargas *et al.*, 2001; Watson and Hirst, 2002), it is less relevant to studies of single side chain–side chain hydrogen bonds typically found in proteins. However, side chains making multiple hydrogen bonds and main chain hydrogen bonds in secondary structure elements require different small molecule models that are more suitable for studies of the relevant physical phenomena, such as hydrogen bonding cooperativity in α helices and β sheets.

We used DFT, HF, and MP2 methods in our electronic structure calculations to make sure that the resulting energies do not depend strongly on the chosen quantum chemistry method. We used the NWChem 4.1 (Harrison *et al.*, 2002) quantum chemistry software package [other standard software packages are Gaussian (<http://www.gaussian.com>) and

Jaguar (<http://www.schrodinger.com>)]. The *aug-cc-pVDZ* basis set was employed, with all dimerization (hydrogen bonding) energies counterpoise (CP) corrected (Boys and Bernardi, 1970) to account for the basis set superposition errors caused by using finite basis sets. For DFT calculations, we used the Perdew, Burke, and Ernzerhof gradient-corrected exchange-correlation functional (PBE96) (Perdew *et al.*, 1996), which reproduces results obtained using alternative nonlocal density functionals and MP2 perturbative hydrogen bonding calculations with reasonable accuracy (Ireta *et al.*, 2003; Kaschner *et al.*, 1998; Tuma *et al.*, 1999). In the case of MP2 calculations, absolute dimerization energies of hydrogen bonded water dimers in the gas phase computed using CP-corrected MP2 with the *aug-cc-pVDZ* basis set are within a few tenths of kcal/mol of the experimentally observed values (Feller, 1992). Furthermore, the difference in dimerization energies between two alternative dimer conformations should be more accurate than the absolute energy values because of the partial cancellation of errors related to finite basis sets.

Hydrogen bonding energies of formamide dimers as a function of δ_{HA} , Ψ , θ , and X are plotted in Fig. 4 using DFT, HF, and MP2 methods. There are pronounced minima in the δ_{HA} , Ψ , and X energy dependences and a shallower minimum in the θ energy dependence. DFT and MP2 calculations produce essentially identical results (compare green solid curves and blue dashed curves in Fig. 4), whereas HF calculations exhibit substantial differences, especially in the location and magnitude of the dimerization energy minima as a function of δ_{HA} and Ψ (red curves with dashes and dots in Fig. 4). The HF approach neglects explicit electron–electron correlations, which are known to be important for accurate estimates of hydrogen bonding energies and geometries (Scheiner, 1997). Indeed, when the electron–electron correlation energy is subtracted from the total DFT dimerization energy, the shape of the energy surface becomes closer to that computed using HF theory, with minima positions shifted and dimerization energies underestimated (black solid curves in Fig. 4) as in the case of HF calculations. The starting point for each series of calculations shown in Fig. 4 was the formamide dimer optimized geometrically with the corresponding *ab initio* method. Comparison of geometrical parameters of the optimized dimers shows that DFT and MP2 minimum energy conformations are virtually identical (e.g., $\delta_{\text{HA}} = 1.94 \text{ \AA}$ for DFT and 1.97 \AA for MP2), whereas with HF theory the hydrogen bond is longer ($\delta_{\text{HA}} = 2.10 \text{ \AA}$), and the angle at the acceptor atom Ψ is more linear than in either of the other QM methods ($\Psi = 112.91^\circ$ for DFT, 110.49° for MP2, and 138.16° for HF).

It is important to investigate whether our method of creating four series of dimer geometries in which one degree of freedom is varied at

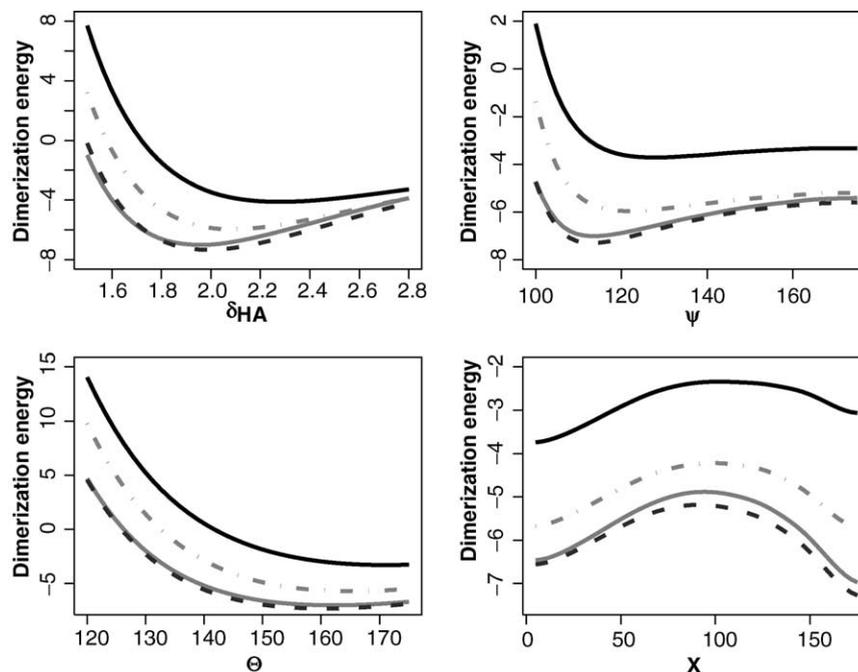


FIG. 4. Formamide dimer hydrogen bonding energies (kcal/mol) versus δ_{HA} (\AA), Ψ (degree), θ (degree), and X (degree). Green (solid lines), DFT; blue (dashes), MP2; red (dashes and dots), HF self consistent field quantum chemistry methods; black (solid lines), DFT minus the correlation energy component. The hydrogen bonding energy is equal to the dimerization energy of a given dimer conformation.

a time while all the others stay fixed at their minimum energy values leads to a significant distortion of the hydrogen bonding energy landscape. In a more realistic description, all degrees of freedom should be allowed to adjust in order to better accommodate a fixed value of the hydrogen bond geometric parameter. We address this issue with an additional DFT calculation in which a constrained geometric optimization is carried out: the geometric parameter, which is varied to create a given projection of the dimerization energy landscape, is kept fixed, but all other degrees of freedom in the dimer are allowed to relax. The resulting landscapes are shown as black curves in Fig. 5 (the green DFT curves without the extra optimization from Fig. 4 are also shown in Fig. 5 for comparison). Both approaches produce similar energies and qualitative behavior.

Comparison of the DFT calculations with the empirical orientation-dependent hydrogen bonding potential described earlier shows a remarkably

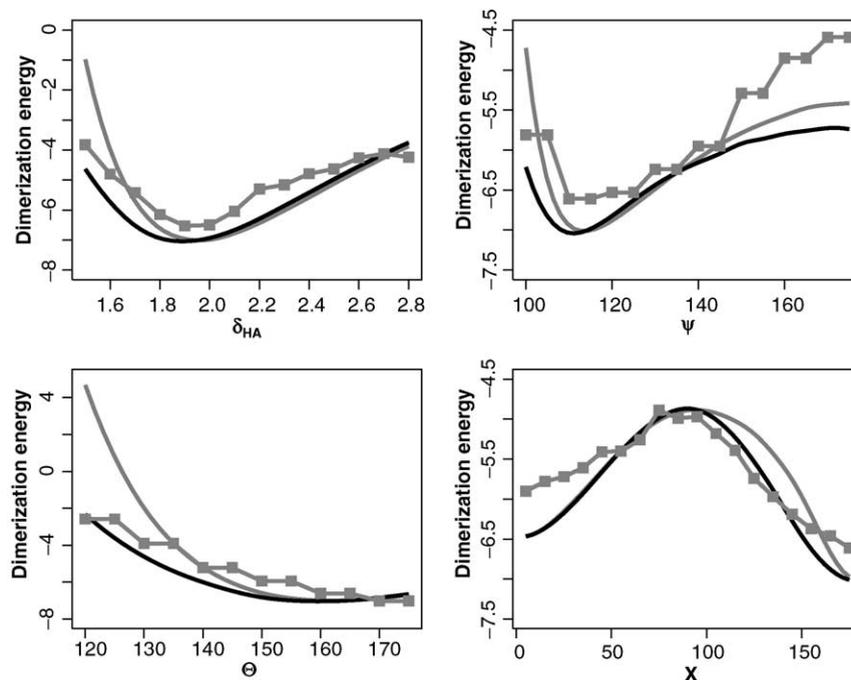


FIG. 5. Formamide dimer hydrogen bonding energies (kcal/mol) versus δ_{HA} (Å), ψ (degree), θ (degree), and X (degree). Green, DFT (same as in Fig. 4); black, DFT with constrained geometry optimization over all degrees of freedom other than the one plotted; red, knowledge-based hydrogen bonding potential (negative logarithm of frequency distributions for side chain–side chain sp^2 hydrogen bonds in proteins). Filled squares correspond to the middle of frequency bins.

close agreement (Fig. 5). This correspondence is especially striking if we recall that derivation of the structure-based potential involves averaging over solvent degrees of freedom and different protein environments. The similarity between *ab initio* energies and hydrogen bond geometry distributions observed in proteins suggests that the DFT and MP2 calculations on the small molecule models capture the essential features of hydrogen bonding interactions between amino acid side chains in protein structures, perhaps because the short range and the partially covalent nature of the hydrogen bond make it relatively insensitive to the nonlocal macromolecular context.

The quantum chemistry methods discussed so far are designed specifically to predict detailed electronic density distributions and model covalent bonding. Thus they can be expected to produce a quite accurate

description of orientation and distance dependence of hydrogen bonding energies. As discussed in the previous section, current molecular mechanics force fields used widely in biomolecular simulations essentially model hydrogen bonding as an electrostatic interaction: positive partial point charges are placed on the proton and the acceptor base, and negative partial point charges are placed on the acceptor and donor atoms. The energy of two dipoles is at a minimum when all four atoms are collinear, favoring linear hydrogen bonds. The nonspherical distribution of electron density, particularly the lone pairs of the sp^2 hybridized oxygen atom positioned at 120° are not captured by the simple point charge model. Therefore, MM calculations are likely to result in dimerization energy landscapes that are not as close to the experimentally observed hydrogen bond geometry distributions as DFT or MP2 calculations (Morozov *et al.*, 2004). Figure 6 shows a comparison of the hydrogen bonding landscapes based on the knowledge-based potential with landscapes obtained from

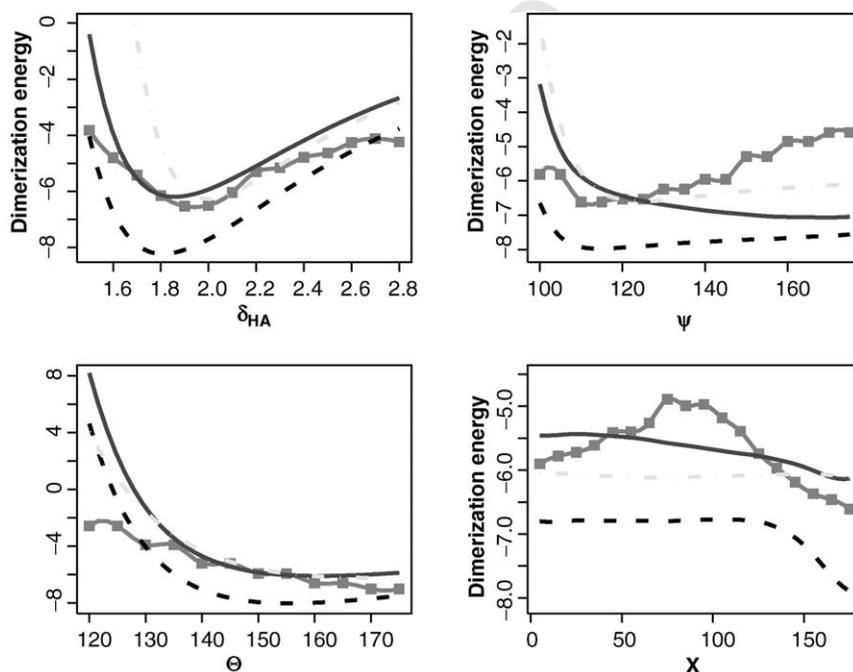


FIG. 6. Formamide dimer hydrogen bonding energies (kcal/mol) versus δ_{HA} (Å), ψ (degree), θ (degree), and X (degree). Red (solid lines with filled squares), knowledge-based hydrogen bonding potential (same as in Fig. 5); blue (solid lines), CHARMM27; black (dashes), OPLS-AA; cyan (dashes and dots), MM3-2000.

three popular MM force fields: MM3-2000 (Allinger, 1989; Lii and Allinger, 1994, 1998), OPLS-AA (Jorgensen *et al.*, 1996), and CHARMM27 (MacKerrell *et al.*, 1998) for DFT optimized formamide dimer geometries. It has been shown that the MM description of hydrogen bonding is improved if partial charges are placed at the lone pair sites and if molecular polarizability is taken into account (Ma *et al.*, 2000). The development of polarizable force fields allowing for more accurate descriptions of hydrogen bonding and electrostatic effects is an active area of research (Halgren and Damm, 2001).

IV. APPLICATIONS OF HYDROGEN BONDING POTENTIALS

A. Protein Structure Prediction and Refinement

Hydrogen bonding potentials have found numerous applications in the prediction of the three-dimensional structure of proteins from their sequence and the refinement of protein models built using experimental constraints from nuclear magnetic resonance (NMR) and X-ray crystallography. Protein structure refinement is used routinely as the final step in constructing macromolecular models from experimental data. Because protein hydrogen bonds are orientation dependent, including an explicit hydrogen bonding term into the effective energy, functions should prove beneficial to the quality of refined structures. Fabiola *et al.* (2002) found that the quality of medium-resolution structures is indeed improved if a hydrogen bonding potential is added to the MM-like effective energy function, which includes Lennard-Jones and Coulomb nonbonded interactions, as well as distance constraints based on X-ray diffraction data. The improvement is evident from the decrease in R^{free} values for a set of 10 medium-resolution crystal structures compared to refinement without hydrogen bonding restraints. Lipsitz *et al.* (2002) studied hydrogen bonds in a set of high-resolution protein crystal structures and discovered a strong correlation between the hydrogen-acceptor distance and the angle at the hydrogen atom. The correlation was substantiated with *ab initio* electronic structure calculations on an alanine-acetamide model system and used in evaluation of the quality of protein structures and NMR structure refinement. The authors found considerable improvement in hydrogen bond geometries after refinement of NMR-derived structural models of Bax, a 192 residue α -helical protein from the Bcl-2 family. Experimental energy terms such as NOE, dihedral, and residual dipolar couplings remained essentially the same, showing that the refinement was consistent with the rest of experimental data. Grishaev and Bax (2004) used an empirical

AU2

backbone–backbone hydrogen bonding potential for NMR structure determination and validation. They found a pronounced improvement in structural quality of NMR models, including a considerable decrease in backbone root mean square deviation (RMSD) relative to the X-ray structures and improvement in the Ramachandran map statistics (Grishaev and Bax, 2004). With respect to biomolecular simulations, Hassan *et al.* (2000) found during development of a continuum solvent model for the CHARMM22 force field that the directional character of hydrogen bonds had to be taken into account for accurate folding predictions of small peptides.

Detailed atom level descriptions of atomic interactions such as the orientation-dependent hydrogen bonding potential should be also useful in the prediction of the three-dimensional structure of a protein from its amino acid sequence. Computational algorithms for protein structure prediction typically consist of three main components: (a) a scoring function, which defines the protein folding landscape; (b) a conformational sampling strategy [Monte Carlo search for energy minima or molecular dynamics (MD) simulations, which employ Newton's laws to construct dynamical molecular trajectories (Hansson *et al.*, 2002)]; and (c) postprocessing of an ensemble of computationally predicted models (decoys) occupying local minima on the energy landscape (Hardin *et al.*, 2002). Protein folding can be visualized using the concept of a multidimensional free energy landscape on which the native conformation occupies a global minimum at the bottom of a folding funnel (Fig. 7). If the free energy function was accurate enough to produce a folding funnel leading to the native state, the postprocessing step would simply consist of sorting decoys by energy. In practice the computed energy landscape is often flat or dominated by misfolded minima in the vicinity of the native conformation. Hence clustering of decoys with similar conformations and additional refinement with more sophisticated sampling and high-resolution scoring functions are employed to pick native-like structures from the decoy ensemble. The computational costs of refining and rescoreing large ensembles of decoys exclude quantum mechanical approaches and limit available choices to empirical computational models. One standard approach to protein structure prediction is to use MD simulations and state-of-the-art force fields that have led to successful discrimination of near-native and misfolded decoys (Lazaridis and Karplus, 1998, 2000; Lee *et al.*, 2001; Vorobjev and Hermans, 1999; Vorobjev *et al.*, 1998) and in folding small proteins using extensive worldwide distributed computer power (Zagrovic, 2002). An alternative approach is to construct an empirical scoring function that can have both statistics- and physics-based terms. Statistics-based terms utilize experimental structural information in an average way: similar

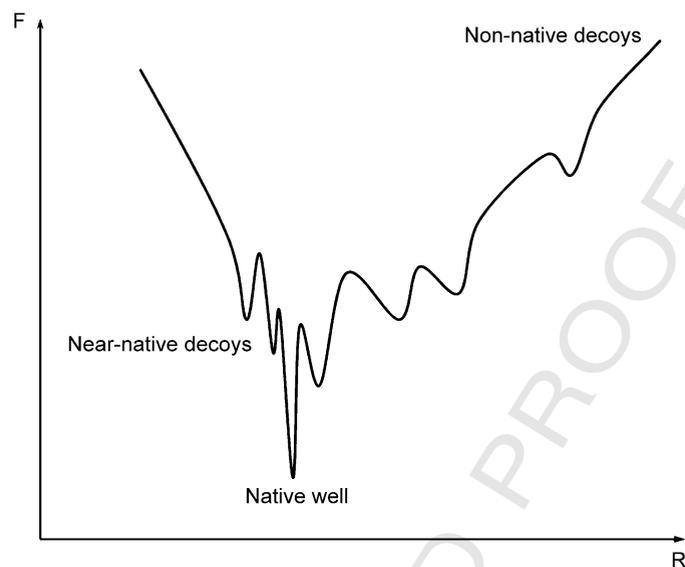


FIG. 7. Schematic representation of the one-dimensional free energy landscape for protein folding and protein-protein binding. F is the free energy, and R is an arbitrary reaction coordinate. The free energy landscape is defined as the free energy of the protein as a function of a set of reaction coordinates (e.g., its conformational degrees of freedom). The native structure occupies the global minimum on the free energy landscape (native well). There is a folding (or binding) funnel in the vicinity of the global minimum. Near-native decoys occupy local minima close to the native well, whereas nonnative decoys occupy more distant local minima.

to the structure-derived hydrogen bonding potential, experimentally observed frequency distributions of geometric features in the protein database are converted into effective energies by assuming a Boltzmann distribution. Different statistical terms can be constructed depending on the requirements of the model, taking into account such features as the degree of residue burial, amino acid-dependent probabilities of being in a secondary structure element, probabilities of side chain and main chain dihedral angles, and so on. These terms could be used to describe protein energetics together with physics-based terms responsible for electrostatics, salvation, and van der Waals interactions.

Kortemme *et al.* (2003) and Morozov *et al.* (2003) studied how the empirical orientation-dependent hydrogen bonding potential affects discrimination of native structures and near-native decoys from incorrect protein conformations. This approach is based on the assumption that

conformations with unrealistic hydrogen bond geometries should be penalized relative to those with more native-like hydrogen bonds. We used a diverse protein decoy set comprising 41 single domain proteins with less than 90 amino acids in length. For each protein about 2000 decoys were generated using the ROSETTA method for *ab initio* structure prediction (Rohl *et al.*, 2004). This decoy set was split into two subsets: 25 proteins with high-resolution crystal structures (the high-resolution subset) and 23 proteins for which ROSETTA was able to produce sufficiently many native-like decoys, as evaluated by the RMSD of decoy C α coordinates relative to the native structure (the low RMSD subset). Note that some proteins appear in both subsets. In addition, for the latter subset, 300 native-like decoys were created for each protein, starting from the native conformation, in order to better sample the free energy landscape in the vicinity of the native free energy minimum.

We used Z-score analysis to quantify the signal-to-noise ratio on this data set. A Z score of a structure is the average energy of all decoys made for a given protein minus the energy of that structure, divided by the standard deviation of all decoy energies. Z scores serve as a convenient measure of the discriminatory power of various terms and combinations of terms in the empirical scoring function. The high-resolution decoy subset was used to compute Z scores with respect to the native and native repacked structures [in which all side chains were modeled (“repacked”) with the same rotamer-based protocol as that employed in creation of decoys so that information about native side chains is lost but the backbone stays in its native conformation]. In the low RMSD subset, all decoys were divided into low RMSD (native-like) and nonnative classes and Z scores are computed with respect to the average energy of the low RMSD decoys. Native-like decoys were defined as being in the lowest 5% of the RMSD histogram, which leads to an average cutoff of 2.84 and 2.33 Å (if decoys created by perturbing the native structure were included). Native and native repacked Z scores are used to assess the degree of similarity between native structures and decoys, whereas low RMSD Z scores are constructed for a more stringent test, which evaluates whether native-like decoys can be distinguished from their nonnative counterparts. A structure or a group of structures was defined to be discriminated successfully if the corresponding Z score was greater than 1.0.

The hydrogen bonding potential (including contributions from main chain–main chain, side chain–side chain and side chain–main chain hydrogen bonds) was found to successfully discriminate 22 out of 25 native structures in the high-resolution decoy set (discrimination is defined as successful if the corresponding Z score is greater than 1.0). Overall, there is a large energy gap between native structures and average decoys: the

average Z score is 4.03. The Z score drops to 3.34 when native side chains are repacked, but 23 out of 25 native repacked structures are still discriminated successfully. Evidently, for some proteins the rotamer repacking procedure does not reproduce native side chain conformations accurately, either due to limitations in rotamer sampling or errors in the scoring function used for repacking side chains. In contrast to native and native repacked results, discrimination is poor for the low RMSD decoy set—only 12 proteins out of 23 have Z scores greater than 1.0. Some of the difficulty is rooted in the inability of the ROSETTA method to create a sufficient number of native-like decoys starting from the unfolded conformation. Indeed, when perturbed native decoys are removed from the set, low RMSD Z scores are greater than 1.0 in only 4 out of 23 cases. This is not surprising given that the hydrogen bonding potential is relatively short ranged, and thus if there are few structures in the decoy set that are close enough to the native energy minimum to have native-like hydrogen bonds, discrimination is expected to be poor. In other words, the width of the hydrogen bonding folding funnel is fairly narrow on the scale of our decoy sets. If hydrogen bonds are grouped into side chain–side chain, side chain–backbone and backbone–backbone classes and their Z scores are considered separately, backbone hydrogen bonds provide most discrimination. Perhaps in less densely packed decoys the freedom of side chain orientations is sufficient to locally optimize hydrogen bonds to the extent seen in native structures.

Z -score analysis of native structures and low RMSD decoys with empirical scoring functions is not directly related to the question of the net energetic contribution of hydrogen bonding to protein stability. The energy component analysis of the type carried out earlier is based on the assumption that protein structures are optimized on average in terms of their electrostatic and hydrogen bonding properties when compared to alternative compact conformations. This assumption does not necessarily imply that hydrogen bonding interactions in native structures are stronger than hydrogen bonds made between protein chemical groups and water in the unfolded state.

B. Prediction of Structures and Energetics of Protein–Protein Interfaces

Computational modeling of protein–protein interactions attracted much attention in recent years, motivated by the central role of protein interactions in cellular processes and the impracticality of determining high-resolution structures experimentally for the vast numbers of protein interactions observed in proteomic studies. Several excellent

reviews give a detailed account of the state of the art in this area (Camacho and Vajda, 2002; Halperin *et al.*, 2002; Smith and Sternberg, 2002; Vajda and Camacho, 2004). Computational protein docking often uses a two-step procedure: in the first step, a large number of docked conformations is generated using rigid body search and a scoring function that models shape and chemical complementarity and other biochemical constraints. In the second step, the models of protein-protein complexes generated in the first step are rescored using MM force fields or mixed scoring functions combining van der Waals, electrostatics, and solvation interactions with statistical terms. Differences in the descriptions of hydrogen bonding and electrostatic interactions are expected to have an impact on this latter rescoring step (Morozov *et al.*, 2003). Consistent with this idea, a recent evaluation of protein-ligand docking identified the lack of explicit treatment of hydrogen bonds as one of the sources of failure of computational docking algorithms (Perez and Ortiz, 2001).

Just as in protein structure prediction, a computational search for the global minimum on the binding free energy landscape (which has six dimensions if binding partners are treated as rigid bodies) relies on the assumption that native conformations have lower free energies than docking “decoys” (alternative docked conformations). Then the concepts of the binding free energy landscape and the free energy funnel can be used by analogy with protein folding (Tsai *et al.*, 1999). In the process of searching the initial ensemble of docked conformations for native-like protein-protein complexes, protein flexibility has to be taken into account. While in many cases structural rearrangement at the binding interface is limited to a few side chains, changing their conformations in the active site (Najmanovitch *et al.*, 2000), large-scale structural changes such as hinge bending, domain, or loop movement have also been observed (Ramakrishnan and Qasba, 2001). It is conceivable that there is a range of protein conformations with similar energies, and different conformations are chosen by different ligands in the process of binding. In most docking methods, protein flexibility is limited to modeling changes in side chain conformations at the binding interface. For example, repacking interface side chains is an integral part of the docking protocol developed by Gray and co-workers (2003a,b).

We created a set of docking decoys in order to evaluate the ability of our structure-derived hydrogen bonding potential to discriminate native and near native from incorrectly docked conformations (Kortemme *et al.*, 2003). The set included 18 antibody-antigen and 13 enzyme-enzyme inhibitor and other complexes. Antibody-antigen complexes were considered separately because they are known to exhibit systematic differences

from other protein–protein structures, including poorer shape complementarity (Lawrence and Colman, 1993). For each structure, a decoy ensemble with 400 to 2000 docked models was created by rigid body perturbation of the relative orientation of the two partners in the protein–protein complex. Thus the test of the hydrogen bonding potential was limited to the “bound” docking problem in which the polypeptide backbone conformations of the protein partners in the complex were known. Protein flexibility was modeled by repacking all side chain conformations at the binding interface. *Z*-score analysis analogous to that used for monomeric decoys was carried out. The orientation-dependent hydrogen bonding potential alone was sufficient to successfully discriminate native docked conformations in 23 out of 31 protein–protein complexes studied (as for monomeric proteins, discrimination is defined as successful if the corresponding *Z* score was greater than 1.0). The mean native *Z* score is 3.12 for antibody–antigen and 5.72 for other complexes. When native structures were rescored after repacking interface side chains, the *Z* scores did not become considerably lower: 3.38 for antibody–antigen and 5.89 for other complexes; 26 out of 31 structures were discriminated successfully in the native repacked case.

In contrast to the results obtained for single domain proteins, reasonably good discrimination between near-native and high RMSD decoys is achieved for protein–protein complexes. Even though the low RMSD *Z* scores are considerably lower on average (the mean low RMSD *Z* score is just 1.29 for antibody–antigen and 2.70 for other complexes), they are still capable of discriminating 22 out of 31 structures. The correlation between hydrogen bonding energies and RMSD to the native structure starts to play a role in the RMSD range of 2–3 Å, consistent with the width of the folding funnel deduced using the monomeric decoy set. This observation shows that the hydrogen bonding potential is quite useful in the protein docking problem if decoys populating near-native conformations are present in the initial ensemble.

In a related test, we evaluated the ability of a simple energy function dominated by packing interactions, salvation, and hydrogen bonding to account for the change in binding free energy brought about by alanine mutations at protein interfaces (Kortemme and Baker, 2002). The orientation-dependent treatment of hydrogen bonding contributed significantly to the predictions and yielded a better agreement with experimental data than a description of polar interactions using a distance-dependent Coulomb model. Guerois and Serrano (2002) obtained remarkable agreement with experimental data on more than 1000 mutations in proteins and protein–protein complexes, explicitly modeling hydrogen bonds between protein atoms and with water molecules.

C. Protein Design

Many scoring functions for protein design contain explicit hydrogen bonding potentials (Dahiyat and Mayo, 1997; Gordon *et al.*, 1999; Kortemme *et al.*, 2003; Looger *et al.*, 2003). Computational protein design methods seek to identify low energy amino acid sequences for a specified, in most cases fixed, backbone target structure (for a review of protein design methods, see Pokala and Handel, 2001). Experimental characterization of designed proteins provides a stringent test of our understanding of the physicochemical principles underlying protein structure and stability and reflected in the computational models. Correspondingly, successful engineering of hydrogen bonds by computational protein design tests the accuracy of the hydrogen bonding representation in the model.

Before evaluating the structure-derived, orientation-dependent hydrogen bonding potential described earlier in experimental protein design applications, we used a computational test of the hydrogen bonding potential related to protein design. This test is based on the assumption that, on average, substitution of the sequences of proteins with nonnative amino acids is unfavorable compared to the naturally occurring sequence. Thus, a protein design energy function can be evaluated by the extent to which it reproduces the sequences of native proteins (Kuhlman and Baker, 2000). Using this metric, the orientation-dependent hydrogen bonding potential is superior to a pure electrostatic description of hydrogen bonding using a Coulomb model in monomeric proteins, protein-protein complexes (Kortemme *et al.*, 2003; Morozov *et al.*, 2003) and protein-RNA interfaces (Chen *et al.*, 2004). Figure 8 shows an example of the protein design test: for a set of 50 crystal structures of heterodimeric protein-protein complexes, amino acid side chains at each position in the protein interface were substituted one by one by all amino acids in different side chain conformations (rotamers). For each sequence position, the energy of all rotamers of all amino acids was determined using a protein design energy function, and the lowest energy amino acid was selected. The design energy function is dominated by van der Waals packing interactions, solvation, and hydrogen bonding modeled either using our orientation-dependent hydrogen bonding function or a Coulomb term with a linear distance-dependent dielectric constant (Kortemme *et al.*, 2003). Figure 8 shows that the native amino acid is picked more frequently for the polar residues in interfaces using the orientation-dependent hydrogen model, whereas predictions for nonpolar amino acids were essentially unaffected.

There are now many examples of experimental validation of computational protein design methods (for reviews, see Kortemme and Baker,

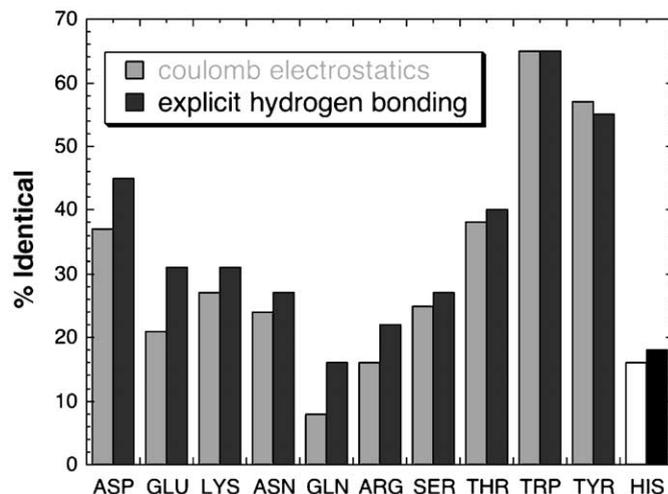


FIG. 8. Recovery of native sequences in protein-protein interfaces. For all sequence positions containing a polar amino acid, bars show how often each native amino acid type is found to be energetically most favorable using different energy functions: dark gray bars represent results from the complete energy function as described in Kortemme *et al.* (2003), including the orientation-dependent hydrogen bonding potential; light gray bars use the same energy function without the hydrogen bonding term, but with a Coulomb term with a linear distance-dependent dielectric constant, scaled to be of a similar magnitude.

2004; Pokala and Handel, 2001). Hydrogen bonding rules were applied to increase the stability of thioredoxin, by eliminating polar residues in the protein core that are not involved in a minimum number of hydrogen bonds generally observed in native proteins (Bolon *et al.*, 2003). Methods to engineer buried polar interactions are especially challenging, but have been applied successfully to the design of specificity in coiled-coil interfaces and protein-peptide interactions (Havranek and Harbury, 2003; Reina *et al.*, 2002). A term ensuring that potential hydrogen bonding donors and acceptors in a protein-ligand interface are satisfied was found to be crucial in the design of novel receptor and sensor proteins (Looger *et al.*, 2003). This strategy was then extended in a landmark study to the design of a ribose binding protein variant exhibiting triose phosphate isomerase activity (Dwyer *et al.*, 2004).

We have structurally characterized computationally designed protein-protein interfaces containing buried hydrogen bonds (Chevalier *et al.*,

2002; Kortemme *et al.*, 2004). A redesigned interface between two distantly related protein domains in an engineered homing endonuclease contains several hydrogen bonds involving mutated amino acids (Chevalier *et al.*, 2002). Using a “computational second site suppressor” strategy (Kortemme *et al.*, 2004), we aimed to alter the specificity of a DNase-inhibitor protein complex. The computational method identifies amino acid changes in one complex partner that would destabilize the interaction, but can be compensated for by corresponding mutations in the other interface partner. This strategy predicted a new buried tyrosine-glutamine side chain-side chain hydrogen bond across the interface in the redesigned protein-protein complex. The new protein pair was found to be functional and specific *in vitro* and *in vivo*. Although this is just a single example, it is still encouraging that the tyrosine-glutamine hydrogen bonding geometry in the X-ray structure of the redesigned complex was very close to what was predicted computationally (Fig. 9).

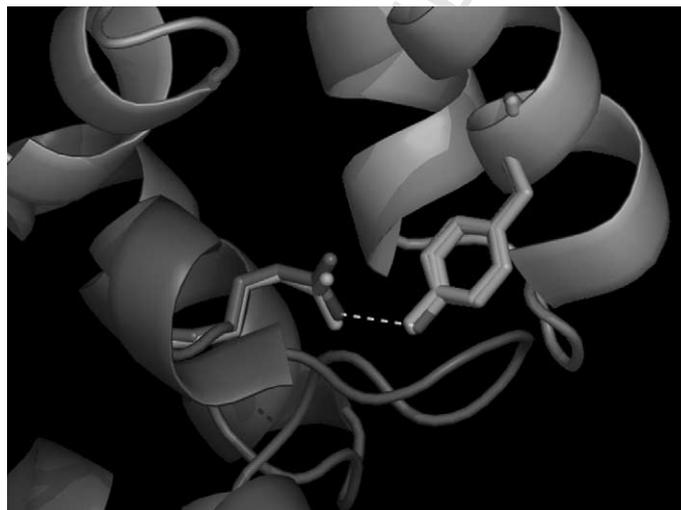


FIG. 9. Prediction and experimental validation of a buried hydrogen bond in a designed protein-protein interface between a colicin E7 DNase variant (K528Q, T539R) and an Im7 inhibitor protein variant (D35Y). The DNase backbone is shown in magenta, the inhibitor protein backbone in yellow. Overlay of the model from computational protein design (green side chains) with the experimentally determined structure (magenta/yellow side chains) shows a buried hydrogen bond between residues Y35 and Q528 that formed as predicted and conferred binding specificity (Kortemme *et al.*, 2004).

V. CONCLUSIONS AND PERSPECTIVES

Hydrogen bonding is an orientation-dependent interaction caused by proton sharing between donor and acceptor atoms. At the most fundamental level, it is a quantum mechanical phenomenon with contributions from polarization and charge transfer. The orientation dependence of hydrogen bonds observed in structures of proteins and small molecules cannot be described accurately using a simple electrostatics model based on dipole-dipole interactions with fixed atomic charges. A more accurate empirical description of hydrogen bonds would have to take their orientation dependence into account, possibly by introducing off-atom partial charges and polarization into the hydrogen bonding model. The orientation dependence of hydrogen bonding energies can be explained by the partially covalent character of hydrogen bonds. In addition, the charge density at the atoms forming a hydrogen bond may be polarized further due to interactions with nearby atomic and molecular groups.

The relative importance of covalent and electrostatic contributions to hydrogen bonding has been somewhat controversial, despite the availability of high-level electronic structure calculations for hydrogen bonded complexes. A major problem was lack of direct experimental evidence supporting the partially covalent character of hydrogen bonding. Such an experiment was carried out for hydrogen bonds in ice using inelastic X-ray (Compton) scattering (Isaacs *et al.*, 1999). Compton scattering can be used to probe the ground state electronic wave function; anisotropies in the Compton scattering profile are sensitive to covalence between neighboring molecules. Experimental oscillations in the anisotropic part of the Compton scattering profile are well reproduced by a full quantum mechanical model, but cannot be captured with a simple electrostatic description, which neglects mixing of electron orbitals upon hydrogen bond formation. Even though the covalent character of hydrogen bonds was only demonstrated in ice crystals, the mechanism will likely be similar for $\text{N}-\text{H}\cdots\text{O}=\text{C}$ and $\text{O}-\text{H}\cdots\text{O}=\text{C}$ hydrogen bonds that play a major role in biological macromolecules.

Another phenomenon that is partially quantum mechanical in origin is hydrogen bonding cooperativity in protein secondary structure elements and in clusters and infinite chains of small hydrogen bonded molecules (see review by Dannenberg). The origin of hydrogen bonding cooperativity lies in the interaction between hydrogen bonds forming extended chains and networks, which results in the dependence of hydrogen bonding energies on the number and orientation of neighboring hydrogen bonds. For example, energies of α -helical hydrogen bonds depend on the length of the helix (Ireta *et al.*, 2003; Park and Goddard, 2000). Hydrogen

bonding cooperativity in secondary structure elements and small molecule clusters can be due to classical electrostatic interactions, polarization, and quantum resonance effects (Dykstra, 1993). Kobko and Dannenberg (2003) argued that the electrostatic dipole–dipole interaction model is inadequate for the description of hydrogen bonding cooperativity in formamide chains and that hydrogen bond strength is enhanced further by a combination of polarization and covalent interactions. Wiczorek and Dannenberg (2003) noted that pairwise electrostatic potentials lacking polarization cannot properly describe hydrogen bonding energies in α helices. In the case of multiply stranded β sheets, Zhao and Wu (2002) argued that the cooperativity is due to long-range electrostatics and polarization rather than to short-range resonance effects. In contrast to α helices and multiple stranded β sheets, no significant cooperativity was exhibited by single β strands and 2_7 ribbons (Wu and Zhao, 2001). On the basis of the studies described earlier, it appears that hydrogen bonds can involve electrostatics, polarization, and covalent interactions in different proportions depending on the molecular system. Therefore, empirical electrostatics models neglecting both polarization and resonance effects will, in some cases, be inadequate for the quantitative analysis of hydrogen bonding cooperativity.

How strongly are energies and geometries of protein hydrogen bonds affected by their macromolecular environment? *Ab initio* quantum mechanical calculations show that energies gained when a hydrogen bond adopts its most favorable orientation are typically on the order of 1–2 kcal/mol compared to less favorable orientations. Energies of this magnitude can be offset relatively easily if other favorable interactions are made or if the rest of the molecule imposes structural constraints onto the range of possible hydrogen bonding orientations. In proteins, this phenomenon leads to the secondary structure dependence of hydrogen bonding geometries: for example, main chain–main chain hydrogen bonds in α helices are more constrained by the helical backbone than side chain–side chain hydrogen bonds. Surprisingly, in the latter case, experimentally observed distributions of hydrogen bond geometries are reproduced accurately with *ab initio* calculations on small molecule model systems. This observation supports a fairly local picture of hydrogen bond formation and the limited impact of macromolecular environment on geometries of side chain hydrogen bonds in proteins.

The orientation-dependent hydrogen bonding potential has proven very useful in such diverse applications as protein structure prediction, protein–protein docking, and protein sequence design. Due to the short range of hydrogen bonding interactions, energy funnels on the hydrogen bonding landscape leading to native conformations of monomeric

AU3

proteins and protein–protein complexes are relatively narrow, with RMSD of about 2–3 Å from the native structure. Native structures and decoys in this range can be reasonably well discriminated from nonnative decoys using the hydrogen bonding potential alone, provided that *ab initio* folding and docking algorithms are capable of producing sufficiently many native-like decoys. However, further away from the native well discrimination becomes poor, especially for side chain–side chain hydrogen bonds, which can be equally well optimized in all decoys due to less compact decoy conformations. In protein design applications, hydrogen bonds in designed proteins have been shown by crystallographic analysis to be formed as predicted computationally and may be crucial in defining protein interaction specificity.

Subtle physical mechanisms of hydrogen bond formation and the role of hydrogen bonds in experimentally observed protein and small molecule structures have been investigated for decades and continue to be a focus of many recent studies. Some of these studies are empirical surveys of hydrogen bond geometries in structural databases of proteins and small molecules, whereas others are theoretical calculations of energies and geometries of hydrogen bonds in a variety of molecules and molecular complexes. Theoretical approaches to modeling hydrogen bonds range from empirical molecular mechanics descriptions of biological macromolecules to high-level *ab initio* electronic structure calculations on model hydrogen bonded systems. Taken together, theoretical and empirical studies provide a unified and consistent picture of the hydrogen bond and underscore its role as an important determinant of macromolecular interactions. Including hydrogen bonding potentials into empirical energy functions leads to marked improvement in the performance of algorithms developed for computational structure prediction of monomeric proteins and protein–protein complexes and for protein sequence design.

ACKNOWLEDGMENTS

We thank David Baker and many members of the Baker laboratory for stimulating discussions and insights. A.V.M. is supported by a postdoctoral fellowship from the Leukemia and Lymphoma Society. T.K. acknowledges support from the Human Frontier Science Program.

REFERENCES

- Allinger, N. L., Yuh, Y. H., and Lii, J.-H. (1989). Molecular mechanics: The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **111**, 8551–8566.

- Baker, E. N., and Hubbard, R. E. (1984). Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **44**, 97–179.
- Bolon, D. N., Marcus, J. S., Ross, S. A., and Mayo, S. L. (2003). Prudent modeling of core polar residues in computational protein design. *J. Mol. Biol.* **329**, 611–622.
- Boys, S. F., and Bernardi, F. (1970). Calculation of small molecular interactions by differences of separate total energies: Some procedures with reduced errors. *Mol. Phys.* **19**, 553–566.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187–217.
- Buck, M., and Karplus, M. (2001). Hydrogen bond energetics: A simulation and statistical analysis of N-methyl acetamide (NMA), water, and human lysozyme. *J. Phys. Chem. B* **105**, 11000–11015.
- Camacho, C. J., and Vajda, S. (2002). Protein–protein association kinetics and protein docking. *Curr. Opin. Struct. Biol.* **12**, 36–40.
- Chen, Y., Kortemme, T., Robertson, T., Baker, D., and Varani, G. (2004). A new hydrogen-bonding potential for the design of protein-RNA interactions predicts specific contacts and discriminates decoys. *Nucleic Acids Res.* **32**, 5147–5162.
- Chevalier, B. S., Kortemme, T., Chadsey, M. S., Baker, D., Monnat, R. J. J., and Stoddard, B. L. (2002). Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell* **10**, 895–905.
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
- Dahiyat, B. I., and Mayo, S. L. (1997). De novo protein design: Fully automated sequence selection. *Science* **278**, 82–87.
- Dwyer, M. A., Looger, L. L., and Hellinga, H. W. (2004). Computational design of a biologically active enzyme. *Science* **304**, 1967–1971.
- Dykstra, C. E. (1993). Electrostatic interaction potentials in molecular force fields. *Chem. Rev.* **93**, 2339–2353.
- Fabiola, F., Bertram, R., Korostelev, A., and Chapman, M. S. (2002). An improved hydrogen bond potential: Impact on medium resolution protein structures. *Protein Sci.* **11**, 1415–1423.
- Feller, D. (1992). Application of systematic sequences of wave-functions to the water dimer. *J. Chem. Phys.* **96**, 6104–6114.
- Gao, J., and Freindorf, M. (1997). Hybrid ab initio QM/MM simulation of N-methylacetamide in aqueous solution. *J. Phys. Chem. A* **101**, 3182–3188.
- Gavezzotti, A., and Filippini, G. (1994). Geometry of the intermolecular X-H...Y (X, Y = N, O) hydrogen bond and the calibration of empirical hydrogen-bond potentials. *J. Phys. Chem.* **98**, 4831–4837.
- Gordon, D. B., Marshall, S. A., and Mayo, S. L. (1999). Energy functions for protein design. *Curr. Opin. Struct. Biol.* **9**, 509–513.
- Gray, J. J., Moughon, S., Kortemme, T., Schueler-Furman, O., Misura, K. M. S., Morozov, A. V., and Baker, D. (2003a). Protein-protein docking predictions for the CAPRI experiment. *Proteins Struct. Funct. Gen.* **52**, 118–122.
- Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A., and Baker, D. (2003b). Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **331**, 281–299.

- Grishaev, A., and Bax, A. (2004). An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation. *J. Am. Chem. Soc.* **126**, 7281–7292.
- Grzybowski, B. A., Ishchenko, A. V., DeWitte, R. S., Whitesides, G. M., and Shakhnovich, E. I. (2000). Development of a knowledge-based potential for crystals of small organic molecules: Calculation of energy surfaces for C=O...H-N hydrogen bonds. *J. Phys. Chem. B* **104**, 7293–7298.
- Guerois, R., Nielsen, J. E., and Serrano, L. (2002). Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387.
- Guo, H., and Karplus, M. (1992). Ab initio studies of hydrogen bonding of N-methylacetamide: Structure, cooperativity, and internal rotational barriers. *J. Phys. Chem.* **96**, 7273–7287.
- Guo, H., and Karplus, M. (1994). Solvent influence on the stability of the peptide hydrogen bond: A supramolecular cooperative effect. *J. Phys. Chem.* **98**, 7104–7105.
- Hagler, A. T., Huler, E., and Lifson, S. (1974). Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *J. Am. Chem. Soc.* **96**, 5319–5327.
- Hagler, A. T., and Lifson, S. (1974). Energy functions for peptides and proteins. II. Amide hydrogen bond and calculation of amide crystal properties. *J. Am. Chem. Soc.* **96**, 5327–5335.
- Halgren, T. A., and Damm, W. (2001). Polarizable force fields. *Curr. Opin. Struct. Biol.* **11**, 236–242.
- Halperin, I., Ma, B., Wolfson, H., and Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins Struct. Funct. Gen.* **47**, 409–443.
- Hansson, T., Oostenbrink, C., and van Gunsteren, W. F. (2002). Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **12**, 190–196.
- Hardin, C., Pogorelov, T. V., and Luthey-Schulten, Z. (2002). Ab initio protein structure prediction. *Curr. Opin. Struct. Biol.* **12**, 176–181.
- Harrison, R. J., Nichols, J. A., Straatsma, T. P., Dupuis, M., Bylaska, E. J., Fann, G. I., Windus, T. L., Apra, E., de Jong, W., Hirata, S., *et al.* (2002). NWChem, a computational chemistry package for parallel computers, version 4.1. Pacific Northwest National Laboratory, Richland, Washington.
- Hassan, S. A., Guarnieri, F., and Mehler, E. L. (2000). Characterization of hydrogen bonding in a continuum solvent model. *J. Phys. Chem. B* **104**, 6490–6498.
- Havranek, J. J., and Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nature Struct. Biol.* **10**, 45–52.
- Ireta, J., Neugebauer, J., Scheffler, M., Rojo, A., and Galvan, M. (2003). Density functional theory study of the cooperativity of hydrogen bonds in finite and infinite α -helices. *J. Phys. Chem. B* **107**, 1432–1437.
- Isaacs, E. D., Shukla, A., Platzman, P. M., Hamann, D. R., Barbiellini, B., and Tulk, C. A. (1999). Covalency of the hydrogen bond in ice: A direct X-ray measurement. *Phys. Rev. Lett.* **82**, 600–603.
- Jernigan, R. L., and Bahar, I. (1996). Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* **6**, 195–209.
- Jorgensen, W. L., and Tirado-Rives, J. (1988). The OPLS potential functions for proteins: Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666.

- Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236.
- Kaschner, R., and Hohl, D. (1998). Density functional theory and biomolecules: A study of glycine, alanine, and their oligopeptides. *J. Phys. Chem. A* **102**, 5111–5116.
- Kobko, N., and Dannenberg, J. J. (2003). Cooperativity in amide hydrogen bonding chains: Relation between energy, position, and H-bond chain length in peptide and protein folding models. *J. Phys. Chem. A* **107**, 10389–10395.
- Kollman, P. A. (1977). Noncovalent interactions. *Accounts Chem. Res.* **10**, 365–371.
- Kortemme, T., and Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. USA* **99**, 14116–14121.
- Kortemme, T., and Baker, D. (2004). Computational design of protein-protein interactions. *Curr. Opin. Chem. Biol.* **8**, 91–97.
- Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L., and Baker, D. (2004). Computational redesign of protein-protein interaction specificity. *Nature Struct. Mol. Biol.* **11**, 371–379.
- Kortemme, T., Morozov, A. V., and Baker, D. (2003). An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.* **326**, 1239–1259.
- Kuhlman, B., and Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA* **97**, 10383–10388.
- Lawrence, M. C., and Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950.
- Lazaridis, T., and Karplus, M. (1998). Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *J. Mol. Biol.* **288**, 477–487.
- Lazaridis, T., and Karplus, M. (2000). Effective energy functions for protein structure prediction. *Curr. Opin. Struct. Biol.* **10**, 139–145.
- Lee, M. R., Tsai, J., Baker, D., and Kollman, P. A. (2001). Molecular dynamics in the endgame of protein structure prediction. *J. Mol. Biol.* **313**, 417–430.
- Lii, J.-H., and Allinger, N. L. (1994). Directional hydrogen bonding in the MM3 force field. I. *J. Phys. Org. Chem.* **7**, 591–609.
- Lii, J.-H., and Allinger, N. L. (1998). Directional hydrogen bonding in the MM3 force field: II. *J. Comp. Chem.* **19**, 1001–1016.
- Lipsitz, R. S., Sharma, Y., Brooks, B. R., and Tjandra, N. (2002). Hydrogen bonding in high-resolution protein structures: A new method to assess NMR protein geometry. *J. Am. Chem. Soc.* **124**, 10621–10626.
- Looger, L. L., Dwyer, M. A., Smith, J. J., and Hellinga, H. W. (2003). Computational design of receptor and sensor proteins with novel functions. *Nature* **423**, 185–190.
- Lumb, K. J., and Kim, P. S. (1995). A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* **34**, 8642–8648.
- Ma, B., Lii, J.-H., and Allinger, N. L. (2000). Molecular polarizabilities and induced dipole moments in molecular mechanics. *J. Comp. Chem.* **21**, 813–825.
- MacKerrell, A. D., Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Jr., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616.
- Mayo, S. L., Olafson, B. D., and Goddard, W. A., III (1990). DREIDING: A generic force field for molecular simulations. *J. Phys. Chem.* **94**, 8897–8909.

- McDonald, I. K., and Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.
- McGuire, R. F., Momany, F. A., and Scheraga, H. A. (1972). Energy parameters in polypeptides. V. Empirical hydrogen bond potential function based on molecular orbital calculations. *J. Phys. Chem.* **76**, 375–393.
- Morozov, A. V., Kortemme, T., and Baker, D. (2003). Evaluation of models of electrostatic interactions in proteins. *J. Phys. Chem. B* **107**, 2075–2090.
- Morozov, A. V., Kortemme, T., Tsemekhman, K., and Baker, D. (2004). Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc. Natl. Acad. Sci. USA* **101**, 6946–6951.
- Morokuma, K. J. (1971). Molecular orbital studies of hydrogen bonds. III. C=O...H-O hydrogen bond in H₂CO...H₂O and H₂CO...2H₂O. *J. Chem. Phys.* **55**, 1236–1244.
- Neria, E., Fischer, S., and Karplus, M. (1996). Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **105**, 1902–1921.
- No, K. T., Kwon, O. Y., Kim, S. Y., Jhon, M. S., and Scheraga, H. A. (1995). A simple functional representation of angular-dependent hydrogen-bonded systems. I. Amide, carboxylic acid, and amide-carboxylic acid pairs. *J. Phys. Chem.* **99**, 3478–3486.
- Park, C., and Goddard, W. A., III (2000). Stabilization of α -helices by dipole-dipole interactions within α -helices. *J. Phys. Chem. B* **104**, 7784–7789.
- Parr, R. G., and Yang, W. (1989). “Density-Functional Theory of Atoms And Molecules.” Oxford University Press, Oxford.
- Pauling, L., and Corey, R. B. (1951). Configurations of polypeptide chains with favored orientations around single bonds: Two new pleated sheets. *Proc. Natl. Acad. Sci. USA* **37**, 729–740.
- Perdew, J., Burke, K., and Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868.
- Perez, C., and Ortiz, A. R. (2001). Evaluation of docking functions for protein-ligand docking. *J. Med. Chem.* **44**, 3768–3785.
- Petrey, D., and Honig, B. (2000). Free energy determinants of tertiary structure and the evaluation of protein models. *Protein Sci.* **9**, 2181–2191.
- Pokala, N., and Handel, T. M. (2001). Review: Protein design-where we were, where we are, where we’re going. *J. Struct. Biol.* **134**, 269–281.
- Ponder, J. W., and Case, D. A. (2003). Force fields for protein simulations. *Adv. Protein Chem.* **66**, 27–85.
- Qian, W., Mirkin, N. G., and Krimm, S. (1999). A spectroscopically effective molecular mechanics model for the intermolecular interactions of the hydrogen-bonded N-methylacetamide dimer. *Chem. Phys. Lett.* **315**, 125–129.
- Ramakrishnan, B., and Qasba, P. K. (2001). Crystal structure of lactose synthase reveals a large conformational change in its catalytic component, the beta1, 4-galactosyltransferase-I. *J. Mol. Biol.* **310**, 205–218.
- Reiher, W. E., III (1985). “Theoretical Studies of Hydrogen Bonding.” Ph.D. Thesis, Harvard University.
- Reina, J., Lacroix, E., Hobson, S. D., Fernandez-Ballester, G., Rybin, V., Schwab, M. S., Serrano, L., and Gonzalez, C. (2002). Computer-aided design of a PDZ domain to recognize new target sequences. *Nature Struct. Biol.* **9**, 621–627.
- Rohl, C. A., Strauss, C. E. M., Misura, K., and Baker, D. (2004). Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93.

- Scheiner, S. (1997). "Hydrogen Bonding: A Theoretical Perspective." Oxford University Press, Oxford.
- Singh, U. C., and Kollman, P. A. (1985). A water dimer potential based on ab initio calculations using Morokuma component analyses. *J. Chem. Phys.* **83**, 4033–4040.
- Smith, G. R., and Sternberg, M. J. E. (2002). Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct. Biol.* **12**, 28–35.
- Szabo, A., and Ostlund, N. S. (1982). "Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory." Macmillan, New York.
- Thomas, P. D., and Dill, K. A. (1996). An iterative method for extracting energy-like quantities from protein structures. *Proc. Natl. Acad. Sci.* **93**, 11628–11633.
- Topol, I. A., Burt, S. K., and Rashin, A. A. (1995). Can contemporary density functional theory yield accurate thermodynamics for hydrogen bonding? *Chem. Phys. Lett.* **247**, 112–119.
- Torii, H., Tatsumi, T., Kanazawa, T., and Tasumi, M. (1998). Effects of intermolecular hydrogen-bonding interactions on the amide I mode of N-methylacetamide: Matrix-isolation infrared studies and ab initio molecular orbital calculations. *J. Phys. Chem. B* **102**, 309–314.
- Tsai, C. J., Kumar, S., Ma, B., and Nussinov, R. (1999). Folding funnels, binding funnels, and protein function. *Protein Sci.* **8**, 1181–1190.
- Tuma, C., Boese, A. D., and Handy, N. C. (1999). Predicting the binding energies of H-bonded complexes: A comparative DFT study. *Phys. Chem. Chem. Phys.* **1**, 3939–3947.
- Umeyana, H., and Morokuma, K. J. (1977). The origin of hydrogen bonding: An energy decomposition study. *J. Am. Chem. Soc.* **99**, 1316–1332.
- Vajda, S., and Camacho, C. J. (2004). Protein-protein docking: Is the glass half-full or half-empty? *Trends Biotech.* **22**, 110–116.
- Vargas, R., Garza, J., Friesner, R. A., Stern, H., Hay, B. P., and Dixon, D. A. (2001). Strength of the N–H···O=C and C–H···O=C bonds in formamide and N-methylacetamide dimers. *J. Phys. Chem. A* **105**, 4963–4968.
- Vorobjev, Y. N., Almagro, J. C., and Hermans, J. (1998). Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. *Proteins Struct. Funct. Gen.* **32**, 399–413.
- Vorobjev, Y. N., and Hermans, J. (1999). ES/IS: Estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. *Biophys. Chem.* **78**, 195–205.
- Watson, T. M., and Hirst, J. D. (2002). Density functional theory vibrational frequencies of amides and amide dimers. *J. Phys. Chem. A* **106**, 7858–7867.
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr., and Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784.
- Wieczorek, R., and Dannenberg, J. J. (2003). H-bonding cooperativity and energetics of α -helix formation of five 17-amino acid peptides. *J. Am. Chem. Soc.* **125**, 8124–8129.
- Wu, Y., and Zhao, Y. (2001). A theoretical study on the origin of cooperativity in the formation of 3_{10} - and α -helices. *J. Am. Chem. Soc.* **123**, 5313–5319.
- Zagrovic, B., Snow, C. D., Shirts, M. R., and Pande, V. S. (2002). Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.* **323**, 927–937.

- Zhang, C., Liu, S., Zhu, Q., and Zhou, Y. (2005). A knowledge-based energy function for protein-ligand, protein-protein, and protein-DNA complexes. *J. Med. Chem.* **48**, 2325–2335.
- Zhao, Y., and Wu, Y. (2002). A theoretical study of β -sheet models: Is the formation of hydrogen-bond networks cooperative? *J. Am. Chem. Soc.* **124**, 1570–1571.

FURTHER READING

- Dunbrack, R. L., Jr., and Karplus, M. (1993). Backbone-dependent rotamer library for proteins: Application to side-chain prediction. *J. Mol. Biol.* **230**, 543–574.
- Dunbrack, R. L., Jr., and Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661–1681.
- Elcock, A. H., Sept, D., and McCammon, J. A. (2001). Computer simulation of protein-protein interactions. *J. Phys. Chem. B* **105**, 1504–1518.
- Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325.
- Najmanovich, R., Kuttner, J., Sobolev, V., and Edelman, M. (2000). Side-chain flexibility in proteins upon ligand binding. *Proteins Struct. Funct. Gen.* **39**, 261–268.

Author Query Form



Journal: Advances in Protein Chemistry, 72
Article No.: Chapter 2

Dear Author,

During the preparation of your manuscript for typesetting some questions have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list. This form should then be returned with your marked proof/list of corrections to Elsevier Science.

Disk use

In some instances we may be unable to process the electronic file of your article and/or artwork. In that case we have, for efficiency reasons, proceeded by using the hard copy of your manuscript. If this is the case the reasons are indicated below:

- Disk damaged
- Incompatible file format
- LaTeX file for non-LaTeX journal
- Virus infected
- Discrepancies between electronic file and (peer-reviewed, therefore definitive) hard copy.
- Other:

We have proceeded as follows:

- Manuscript scanned
- Manuscript keyed in
- Artwork scanned
- Files only partly used (parts processed differently:.....)

Bibliography

If discrepancies were noted between the literature list and the text references, the following may apply:

- The references listed below were noted in the text but appear to be missing from your literature list. Please complete the list or remove the references from the text.
- Uncited references:** This section comprises references which occur in the reference list but not in the body of the text. Please position each reference in the text or, alternatively, delete it. Any reference not dealt with will be retained in this section.

Query Refs.	Details Required
Author	<p>Further Reading section:</p> <p>Some of your references may appear in a "Further Reading" section at the end of the chapter. These are references that were found to be uncited in the text, and have been called out to your attention. This is only temporary. To ensure that the called out references appear in the reference list, please cite them in the appropriate spot within the text.</p> <p>We appreciate your cooperation.</p>

Query Refs.	Details Required	Author's response
AU1	What is this?	
AU2	Sentence ok?	
AU3	Sentence ok	

Typesetter Query

Attention authors: Please address every typesetter query below. Failure to do so may result in references being deleted from your proofs. Thanks for your cooperation.

Query Refs.	Details Required	Author's response
TS1	Please provide further author names for the below references: Harrison, R. J., <i>et al.</i> (2002) MacKerrell, A. D., <i>et al.</i> (1998).	