PROTEINS: Structure, Function, and Genetics 53:262–272 (2003)

Efficient Minimization of Angle-Dependent Potentials for Polypeptides in Internal Coordinates

William J. Wedemeyer* **and David Baker**
University of Washington, Seattle, Washington

ABSTRACT **Angular potentials play an important role in the refinement of protein structures through angle-dependent restraints (e.g., those determined by cross-correlated relaxations, residual dipolar couplings, and hydrogen bonds). Analytic derivatives of such angular potentials with respect to the dihedral angles of proteins would be useful for optimizing such restraints and other types of angular potentials (i.e., such as we are now introducing into protein structure prediction) but have not been described. In this article, analytic derivatives are calculated for four types of angular potentials and integrated with the efficient recursive derivative calculation methods of Gō and coworkers. The formulas are implemented in publicly available software and illustrated by refining a low-resolution protein structure with idealized vector-angle, dipolar-coupling, and hydrogen-bond restraints. The method is now being used routinely to optimize hydrogen-bonding potentials in ROSETTA. Proteins 2003;53:262–272.** © **2003 Wiley-Liss, Inc.**

Key words: angular potentials; NMR structure determination; residual dipolar coupling; cross-correlated relaxations; dihedral-angle energy gradient; recursive methods; vector-angle restraint; vector-tensor restraint; vector-displacement restraint

INTRODUCTION

The most efficient optimization methods require computation of the derivatives of the energy function with respect to each degree of freedom.[1,2] The conformation of a polypeptide chain can be represented most economically by its dihedral angles, ϕ, ψ, and ω for the backbone and various χ angles for the side-chains. Angle-dependent potentials arise in a variety of situations, ranging from NMR structure restraints (e.g., those from cross-correlated relaxations[3,4] and residual dipolar couplings[5–7]) to protein structure prediction and refinement (e.g., orientation-dependent potentials for hydrogen bonds, cation-pi/stacking interactions and secondary structure-packing potentials). Hence, analytic formulas for the derivatives of such orientation-dependent potentials with respect to the dihedral angles of the protein would be helpful in various contexts. To our knowledge, such formulas have not been described in the literature (but see Appendix B). This article calculates such derivatives for three types of angular restraints: vector-angle, dipolar-coupling and hydrogen-bonding restraints. As seen below, the essence of the method is to convert the trigonometric angle restraints into linear vector restraints.

This article is organized as follows. The first section introduces our notation and some general results such as the recursive methods of Gō and coworkers.[8,9] In the next three sections, we derive explicit formulas for the dihedral-angle gradients of vector-angle, dipolar-coupling, and hydrogen-bond restraints. These three types of NMR restraints are exemplary of three general types of angular potentials, which we denote here as *vector-angle, vector-tensor,* and *vector-displacement* potentials (see below). In each section, we show how to compute these derivatives recursively. In the final section, we illustrate these gradients by refining the various types of restraints for low-resolution structures. A fourth type of angular potential, denoted as a *displacement-dihedral* potential, is considered in Appendix A.

NOTATION AND GENERAL RESULTS

Our notation is as follows. The variable α represents an arbitrary dihedral angle of the protein backbone (i.e., ϕ, ψ, or ω). The rotation axis of the dihedral angle α is denoted by the unit vector $\mathbf{n}_\alpha$, whereas $\mathbf{R}^\alpha$ denotes the position of the C-terminal atom of the bond about which the rotation takes place. For example, the $\mathbf{n}_\alpha$ of the ϕ_i dihedral angle of residue i is the unit vector pointing along the N_i-C_i^α bond of the same residue, whereas the corresponding $\mathbf{R}^\alpha$ is the position of the C_i^α atom. Nondihedral unit vectors are denoted by $\mathbf{u}$ and identified by a superscript; it should be noted that unit vectors need not point along a chemical bond, but may, for example, be linear combinations of bond unit vectors. The partial derivative of a unit vector $\mathbf{u}$ that is moved by a dihedral angle α is given by the formula[10]

$$\frac{\partial \mathbf{u}}{\partial \alpha} = \mathbf{n}_\alpha \times \mathbf{u} \tag{1}$$

The derivative of a vector displacement $\mathbf{R}^{ij} \equiv \mathbf{R}^j - \mathbf{R}^i$ between two atoms i and j separated by the dihedral angle α is given by

*Correspondence to: William J. Wedemeyer, Department of Biochemistry Michigan State University, East Lansing MI 48824. E-mail: bill_wedemeyer@usa.net

Received 12 November 2002 ; Accepted 8 March 2003

$$\frac{\partial}{\partial\alpha}\mathbf{R}^{ij} = \pm\mathbf{n}_\alpha \times \mathbf{R}^{\alpha j} \tag{2}$$

$$= \mp\,\mathbf{n}_\alpha \times \mathbf{R}^{i\alpha} \tag{3}$$

where the vectors $\mathbf{R}^{\alpha j}$ and $\mathbf{R}^{i\alpha}$ are the displacements from $\mathbf{R}^\alpha$ to $\mathbf{R}^j$ and from $\mathbf{R}^i$ to $\mathbf{R}^\alpha$, respectively. The sign is determined by whether atom i is N-terminal (and atom j C-terminal) to the dihedral angle α, or vice versa. Finally, we use the notation $k < \alpha, k = \alpha$ and $k > \alpha$ to denote that an atom k is either 1) not moved, 2) directly moved, or 3) indirectly moved by dihedral angle α. An atom is *directly* moved by dihedral angle α if it belongs to the rigid body immediately C-terminal to the dihedral-angle bond; by contrast, an atom is *indirectly* moved if it is C-terminal to the dihedral-angle bond of α but other dihedral angles intervene. For example, the angle ϕ_i of residue i directly moves the H_i^α, C_i', and C_i^β atoms of the same residue but indirectly moves the O_i' atom, which is moved directly by the ψ_i angle.

Go̅ and coworkers[8,9] have developed efficient algorithms for computing energy gradients with respect to dihedral angles. The general form of these gradients is

$$\frac{\partial}{\partial\alpha}E = \mathbf{n}_\alpha \cdot \mathbf{F}_\alpha + (\mathbf{n}_\alpha \times \mathbf{R}^\alpha)\cdot\mathbf{G}_\alpha \tag{4}$$

where the quantities $\mathbf{F}_\alpha$ and $\mathbf{G}_\alpha$ do not depend on the dihedral-angle variables $\mathbf{n}_\alpha$ and $\mathbf{R}^\alpha$. The principal advantage of such quantities is that they can be computed recursively

$$\mathbf{F}_{\alpha+1} = \mathbf{F}_\alpha + \mathbf{f}_\alpha \tag{5}$$

$$\mathbf{G}_{\alpha+1} = \mathbf{G}_\alpha + \mathbf{g}_\alpha \tag{6}$$

from even simpler quantities $\mathbf{f}_\alpha$ and $\mathbf{g}_\alpha$. The recursion can proceed either from the N-terminus to the C-terminus (using the equations above) or from the C-terminus to the N-terminus, using the equations

$$\mathbf{F}_\alpha = \mathbf{F}_{\alpha+1} - \mathbf{f}_\alpha \tag{7}$$

$$\mathbf{G}_\alpha = \mathbf{G}_{\alpha+1} - \mathbf{g}_\alpha \tag{8}$$

As an illustration, let E represent a simple distance-dependent potential

$$E = \sum_{i<j} V(R^{ij}) \tag{9}$$

where R^{ij} represents the distance between atoms i and j (i.e., the magnitude of the vector $\mathbf{R}^{ij}$). The derivative of this energy can be written in the form of Eq. 4 where the vector quantities $\mathbf{f}_\alpha$ and $\mathbf{g}_\alpha$ are defined[8,9]

$$\mathbf{f}_\alpha = \sum_{atoms\ p=\alpha}\ \sum_{atoms\ q\neq\alpha}\left(\frac{1}{R^{pq}}\right)\left(\frac{dV}{dR}\right)[\mathbf{R}^p \times \mathbf{R}^q] \tag{10}$$

$$\mathbf{g}_\alpha = \sum_{atoms\ p=\alpha}\ \sum_{atoms\ q\neq\alpha}\left(\frac{1}{R^{pq}}\right)\left(\frac{dV}{dR}\right)[\mathbf{R}^p - \mathbf{R}^q] \tag{11}$$

To compute all the $\mathbf{f}_\alpha$ and $\mathbf{g}_\alpha$, it suffices to initialize them to zero and then loop over the interactions $V(R^{ij})$, adding the summands to the $\mathbf{f}_\alpha$ and $\mathbf{g}_\alpha$ for the two dihedral angles α^i and α^j that directly move the atoms i and j. The quantities $\mathbf{F}_\alpha$ and $\mathbf{G}_\alpha$ can then be computed by using the forward recursion relations (Eqs. 5 and 6) with the initial condition $\mathbf{F}_0 = \mathbf{G}_0 = 0$. The backward recursion relations (Eqs. 7 and 8) can be used with equal efficiency, with the initial condition $\mathbf{F}_\alpha = \mathbf{G}_\alpha = 0$ for the post-C-terminal dihedral angle α. In the calculations below, we determine appropriate recursions for each type of angular potential.

We derive analytic formulas for three types of angular potential (Fig. 1). In the first case [Fig. 1(a)], *vector-angle potentials* restrain the angle θ between two vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ that are fixed relative to the local backbone (e.g., the angle between the N—H bond vector of residue i and the C′═O′ bond vector of residue j). Such restraints may be determined from cross-correlated NMR relaxations[3] or residual dipolar coupling experiments.[11,12] Vector-angle potentials may also be useful in modeling physical interactions (e.g., hydrogen bonds[13,14]) or in knowledge-based potentials describing e.g., the preferred geometries of secondary structure elements.[15] In the second case [Fig. 1(b)], *vector-tensor potentials* restrain the double scalar product of a symmetric tensor $\mathbf{D}$ with a vector $\mathbf{u}$ that is fixed relative to the local backbone. Such potentials are already used in refining protein structures using NMR residual dipolar couplings,[6,7,16] but they may have other applications [e.g., in statistical potentials describing the orientation of secondary structure elements relative to the ellipsoidal semiaxes of a protein (i.e., the principal axes of its gyration tensor[17])]. In the third case, [Fig. 1(c)], *vector-displacement potentials* restrain the angle θ between a unit vector $\mathbf{u}$ fixed relative to the local backbone and a displacement $\mathbf{R}^{ij}$ between two atoms i and j in the protein. Such potentials are useful in modeling angle-dependent physical interactions (such as close hydrogen bonds), but again they may be useful for statistical potentials describing e.g., the preferred geometry of secondary structure elements. A fourth type, *displacement-dihedral potentials,* is treated in Appendix A of this article.

VECTOR-ANGLE POTENTIALS

We assume that the total vector-angle potential energy may be written

$$E^{VA} = \sum_{restraints\ m} V^{VA}(x_m) \tag{12}$$

where $V^{VA}(x_m)$ is the energy of restraint m and $x_m \equiv \mathbf{u}_{m1} \cdot \mathbf{u}_{m2}$ is the cosine of the angle between two unit vectors $\mathbf{u}_{m1}$ and $\mathbf{u}_{m2}$ [Fig. 1(a)]. Expressing the angular dependence in vectors is advantageous, because vectors vary in a simple way (linearly) with dihedral-angle changes, allowing recursive methods[8,9] to be used without difficulty.

The derivative of the vector-angle restraint energy E^{VA} with respect to a backbone dihedral angle α is given by the chain rule

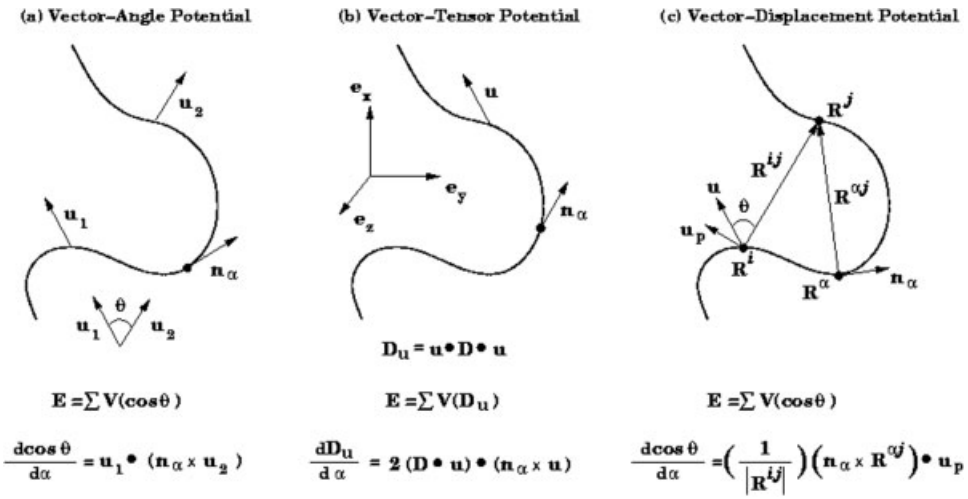

Fig. 1. Illustrations of the angular potentials treated in this article. **a:** A vector-angle potential restraining the angle between two unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$. The unit vector $\mathbf{n}_\alpha$ represents the rotation axis of the backbone dihedral angle α. **b:** A vector-tensor potential restraining the double scalar product between a unit vector $\mathbf{u}$ and a symmetric tensor $\mathbf{D}$; the vectors $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ represent the principal axes of the tensor (e.g., the principal alignment coordinate frame for a residual dipolar coupling tensor). **c:** A vector-displacement potential restraining the angle between a unit vector $\mathbf{u}$ and a displacement $\mathbf{R}^{ij} = \mathbf{R}^j - \mathbf{R}^i$ between two atoms i and j of the molecule. The vector $\mathbf{u}_p$ represents the component of $\mathbf{u}$ that is perpendicular to the displacement $\mathbf{R}^{ij}$. The vector $\mathbf{R}^{ij} = \mathbf{R}^j - \mathbf{R}^i$ represents the displacement between the C-terminal atom of the bond of dihedral angle α and atom j.

$$\frac{\partial}{\partial\alpha} E^{VA} = \sum_{\text{restraints } m} \left(\frac{dV^{VA}}{dx_m}\right)\left(\frac{\partial x_m}{\partial\alpha}\right) \tag{13}$$

The derivative of V^{VA} with respect to x_m is assumed to be known, whereas the derivative of x_m with respect to the dihedral angle α is

$$\frac{\partial x_m}{\partial\alpha} = \mathbf{u}_{m1} \cdot \left(\frac{\partial \mathbf{u}_{m2}}{\partial\alpha}\right) \tag{14}$$

$$= \mathbf{u}_{m1} \cdot (\mathbf{n}_\alpha \times \mathbf{u}_{m2}) \tag{15}$$

$$= \mathbf{n}_\alpha \cdot (\mathbf{u}_{m2} \times \mathbf{u}_{m1}) \tag{16}$$

where we adopt the convention that the dihedral angle change $d\alpha$ leaves the first vector $\mathbf{u}_{m1}$ fixed in space. Thus, the energy gradient becomes

$$\frac{\partial}{\partial\alpha} E^{VA} = \mathbf{n}_\alpha \cdot \sum_{m1<\alpha} \sum_{m2\geq\alpha} \left(\frac{dV^{VA}}{dx_m}\right)(\mathbf{u}_{m2} \times \mathbf{u}_{m1}) \tag{17}$$

In the Gō notation (Eq. 4) above, $\mathbf{G}_\alpha = 0$ and

$$\mathbf{F}_\alpha = \sum_{m1<\alpha} \sum_{m2\geq\alpha} \left(\frac{dV^{VA}}{dx_m}\right)(\mathbf{u}_{m2} \times \mathbf{u}_{m1}) \tag{18}$$

This can be computed recursively by setting the variable $\mathbf{f}_\alpha$ equal to

$$\mathbf{f}_\alpha = \sum_{m1=\alpha} \sum_{m2\neq\alpha} \left(\frac{dV^{VA}}{dx_m}\right)(\mathbf{u}_{m2} \times \mathbf{u}_{m1}) \tag{19}$$

Practically, one computes these $\mathbf{f}_\alpha$ by looping over the vector-angle restraints and adding the summand to the $\mathbf{f}_\alpha$ for each of the dihedral angles α that *directly* move the two vectors involved. The $\mathbf{F}_\alpha$ may then be calculated by using the forward recursion relations (Eqs. 5 and 6) with the initial condition $\mathbf{F}_0 = 0$. (The backward recursion may be used with equal efficiency.)

VECTOR-TENSOR POTENTIALS

Residual dipolar couplings have been the focus of much study in recent years.[6,7] Although such couplings may suffice in some cases to determine the protein structure,[18] they are generally used to refine structures determined by other methods. Such refinements may be done by using vector-angle restraints[11,12] or restraints on the dipolar couplings themselves[16]

$$E^{DC} = \sum_{\text{unit vectors } \mathbf{u}} V^{DC}(D_u) \tag{20}$$

where $D_u \equiv \mathbf{u} \cdot \mathbf{D} \cdot \mathbf{u}$ represents the residual dipolar coupling associated with a unit vector $\mathbf{u}$ for a dipolar coupling tensor $\mathbf{D}$ [Fig. 1(b)]. The derivative of this energy is given by the chain rule

$$\frac{\partial}{\partial\alpha} E^{DC} = \sum_{\text{unit vectors } \mathbf{u}} \left(\frac{dV^{DC}}{dD_u}\right)\left(\frac{\partial D_u}{\partial\alpha}\right) \tag{21}$$

The derivative of V^{DC} with respect to D_u is assumed known, whereas the derivative of the dipolar coupling D_u is given by

$$\frac{\partial}{\partial\alpha} D_u = 2(\mathbf{D} \cdot \mathbf{u}) \cdot \left(\frac{\partial \mathbf{u}}{\partial\alpha}\right) \tag{22}$$

$$= 2(\mathbf{D} \cdot \mathbf{u}) \cdot [\mathbf{n}_\alpha \times \mathbf{u}] \tag{23}$$

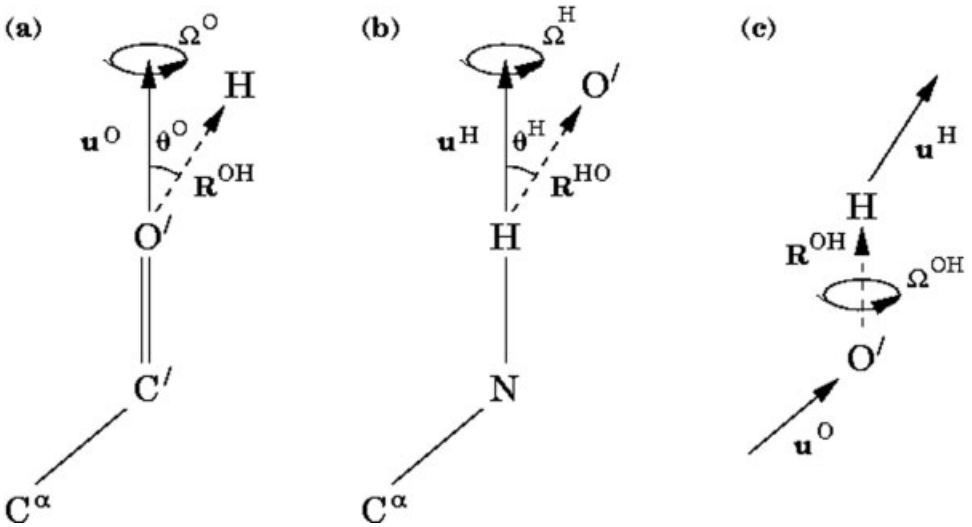

Fig. 2. **a:** The unit vector $\mathbf{u}^O$ is fixed in the acceptor coordinate frame (e.g., that defined by the C^α—C′—O′ atoms of the acceptor peptide group). The variable x^O is the cosine of the angle θ^O between such a unit vector and the vector $\mathbf{R}^{OH}$ from the acceptor atom to the donor hydrogen. The unit vector $\mathbf{u}^O$ need not be aligned with the C′═O′ bond. The azimuthal angle Ω^O of $\mathbf{R}^{OH}$ about $\mathbf{u}^O$ is also depicted; the reference vector defining $\Omega^O = 0$ is chosen to make the distribution of Ω^O as even as possible. **b:** The unit vector $\mathbf{u}^H$ is fixed in the donor coordinate frame (e.g., that defined by the C^α—N—H atoms of the donor peptide group). The variable x^H is the cosine of the angle θ^H between such a unit vector and the vector $\mathbf{R}^{HO}$ from donor hydrogen to the acceptor atom. The unit vector $\mathbf{u}^H$ need not be aligned with the N—H bond. The azimuthal angle Ω^H of $\mathbf{R}^{HO}$ in the coordinate frame is also depicted; the reference vector defining $\Omega^H = 0$ is chosen to make the distribution of Ω^H as even as possible. **c:** The torsional angle Ω^{OH} is formed by three vectors $\mathbf{u}^O$, $\mathbf{R}^{OH}$, and $\mathbf{u}^H$.

$$= \mathbf{n}_\alpha \cdot 2[\mathbf{u} \times (\mathbf{D} \cdot \mathbf{u})] \tag{24}$$

where we treat the dipolar tensor $\mathbf{D}$ here as constant (e.g., as when the structure is being solved in the principal alignment frame]. Under these conditions, the derivative of the total restraint energy equals

$$\frac{\partial}{\partial \alpha} E^{DC} = \mathbf{n}_\alpha \cdot 2 \sum_{\mathbf{u} \geq \alpha} \left(\frac{dV^{DC}}{dD_u} \right) [\mathbf{u} \times (\mathbf{D} \cdot \mathbf{u})] \tag{25}$$

In the Gō notation (Eq. 4) above, $\mathbf{G}_\alpha = 0$ and

$$\mathbf{F}_\alpha = 2 \sum_{\mathbf{u} \geq \alpha} \left(\frac{dV^{DC}}{dD_u} \right) [\mathbf{u} \times (\mathbf{D} \cdot \mathbf{u})] \tag{26}$$

This can be computed recursively by setting the variable $\mathbf{f}_\alpha$ equal to

$$\mathbf{f}_\alpha = -2 \sum_{\mathbf{u} = \alpha} \left(\frac{dV^{DC}}{dD_u} \right) [\mathbf{u} \times (\mathbf{D} \cdot \mathbf{u})] \tag{27}$$

where the negative sign should be noted. The corresponding $\mathbf{F}_\alpha$ can be computed by using the backward recursion relations (Eqs. 7 and 8) with the post-C-terminal $\mathbf{F}_\alpha = 0$. The forward recursion is not as efficient, because it requires computing the full gradient (Eq. 26) at the first dihedral angle. As an aside, the vector $[\mathbf{u} \times (\mathbf{D} \cdot \mathbf{u})]$ is never zero except in the limit of zero anisotropy; because the dipolar coupling tensor is traceless with nonzero eigenvalues, the vector $(\mathbf{D} \cdot \mathbf{u})$ is never parallel to $\mathbf{u}$ in the principal alignment frame.

VECTOR-DISPLACEMENT POTENTIALS

Several methods have been developed recently to detect hydrogen bonds.[19–21] It has long been conjectured that H-bonds (together with side-chain packing) provide the structural specificity that distinguishes the native fold from other plausibly folded structures.[22,23] Hence, it is to be expected that the refinement of H-bonds will improve the protein-like character of a low-resolution structure. This expectation is indeed borne out, as shown below [see Fig. 3(d)].

Hydrogen bonds are well known to have an angular component to their energy, deriving both from dipolar electrostatic interactions between the hydrogen-bonding groups and also (for sufficiently close H-bonds) from covalent interactions.[24,25] This angular dependence has been incorporated into many potentials over the years (e.g., those of Lippincott and Schroeder,[26,27] Coulson,[13] Liquori and coworkers,[14] Scheraga and coworkers,[28,29] Hagler and coworkers,[39] and more recent potentials (e.g., Refs. 31–36). Typically, the angular potentials involve an angle θ^O between the acceptor-hydrogen displacement vector $\mathbf{R}^{OH}$ and a unit vector $\mathbf{u}^O$ fixed in the acceptor coordinate frame [Fig. 2(a)] or an angle θ^H between a unit vector $\mathbf{u}^H$ fixed in the donor hydrogen coordinate frame [Fig. 2(b) and (c)]. As in the earlier sections, we parameterize the potentials in terms of the cosines x^O and x^H of these angles, which can be written as vector dot products

$$x^O \equiv \mathbf{u}^O \cdot (\mathbf{R}^{OH}/R^{OH}) \tag{28}$$

$$x^H \equiv \mathbf{u}^H \cdot (\mathbf{R}^{OH}/R^{OH}) \tag{29}$$

where R^{OH} represents the distance between the acceptor atom O′ and the donor hydrogen H. More generally, we use the term *vector-displacement restraint* to denote potentials that depend on the angle between a vector $\mathbf{u}$ fixed in a local coordinate frame and the displacement $\mathbf{R}^{ij}$ between two atoms i and j [Fig. 1(c)].

For hydrogen bonds, the unit vectors $\mathbf{u}^O$ and $\mathbf{u}^H$ are often chosen as the unit vectors along the C′—O′ and H—N bonds, respectively, but this need not be; $\mathbf{u}^O$ and $\mathbf{u}^H$ can represent any unit vector fixed in the acceptor or donor coordinate frames, respectively, such as vectors aligned with the lone-pair orbitals,[28,29] the amide dipole moment, or statistically optimized vectors (Appendix C). The dot product of two vectors $\mathbf{u}^O$ and $\mathbf{u}^H$ *with each other* has also been used in hydrogen-bond potentials (e.g., the dot product between the C═O and N—H vectors[13,14]) but such vector-angle potentials were treated above. H-bond potentials may also use the azimuthal angles Ω^O and Ω^H of $\mathbf{R}^{OH}$ in the acceptor and donor coordinate frames, or the dihedral angle Ω^{OH} made by the two vectors $\mathbf{u}^O$ and $\mathbf{u}^H$ about the axis defined by $\mathbf{R}^{OH}$ [Figs. 2(a)–(c); Appendix C]. Such dihedral restraints may be treated directly (using methods described in Appendix A) or by the methods derived below, because the orientation of a vector $\mathbf{R}^{OH}$ in a coordinate frame can be specified not only by its polar angles θ and Ω but also by its direction cosines to three orthogonal vectors.[37] Thus, restraints on Ω^O and Ω^H can be mimicked by introducing vector-displacement restraints between $\mathbf{R}^{OH}$ and a set of vectors $\mathbf{u}$ in the chosen coordinate frame.

For brevity, we restrict our derivations to x^O, although the analogous results for x^H are given below. We assume that the angular part of the hydrogen-bond potential can be written as

$$E^O = \sum_{donors\ i} \sum_{acceptors\ j} V^O(x_{ij}^O) \tag{30}$$

The gradient of this energy with respect to a dihedral angle α is given by the chain rule

$$\frac{\partial}{\partial \alpha} E^O = \sum_{donors\ i} \sum_{acceptors\ j} \left(\frac{\partial V^O}{\partial x_{ij}^O}\right)\left(\frac{\partial x_{ij}^O}{\partial \alpha}\right) \tag{31}$$

The derivative of $V(x_{ij}^O)$ with respect to x_{ij}^O is assumed to be known, whereas the derivative of x_{ij}^O is obtained by differentiating Eq. 28

$$\frac{\partial x_{ij}^O}{\partial \alpha} = \left(\frac{1}{R_{ij}^{OH}}\right)\left\{\mathbf{u}_j^O \cdot \left(\frac{\partial}{\partial \alpha}\mathbf{R}_{ij}^{OH}\right) - x_{ij}^O\left(\frac{\partial}{\partial \alpha} R_{ij}^{OH}\right)\right\} \tag{32}$$

where we adopt the convention that the dihedral angle change $d\alpha$ leaves the acceptor atoms (and, hence, the $\mathbf{u}_j^O$ vector) fixed in space. The derivative of the distance R_{ij}^{OH} between the O′ atom of the j^{th} acceptor and the H atom of the i^{th} donor is

$$\frac{\partial}{\partial \alpha} R_{ij}^{OH} = \mathbf{u}_{ij}^{OH} \cdot \frac{\partial}{\partial \alpha}\mathbf{R}_{ij}^{OH} \tag{33}$$

where $\mathbf{u}_{ij}^{OH}$ is the unit vector in the direction of $\mathbf{R}_{ij}^{OH}$. Substituting this result into Eq. 32 yields

$$\frac{\partial x_{ij}^O}{\partial \alpha} = \left(\frac{1}{R_{ij}^{OH}}\right)\left(\frac{\partial}{\partial \alpha}\mathbf{R}_{ij}^{OH}\right) \cdot (\mathbf{u}_j^O - x_{ij}^O \mathbf{u}_{ij}^{OH}) \tag{34}$$

Because $x_{ij}^O \equiv \mathbf{u}_j^O \cdot \mathbf{u}_{ij}^{OH}$, the quantity $(\mathbf{u}_j^O - x_{ij}^O \mathbf{u}_{ij}^{OH})$ represents the component of $\mathbf{u}_j^O$ that is perpendicular to $\mathbf{u}_{ij}^{OH}$. This is geometrically intuitive [Fig. 1(c)], because the angle between $\mathbf{u}^O$ and $\mathbf{R}_{ij}^{OH}$ is not affected by changes in $\mathbf{R}_{ij}^{OH}$ that are parallel to its length.

Finally, the derivative of $\mathbf{R}_{ij}^{OH}$ is obtained from the standard formula[10]

$$\frac{\partial}{\partial \alpha}\mathbf{R}_{ij}^{OH} = \pm \mathbf{n}_\alpha \times \mathbf{R}_i^{\alpha H} \tag{35}$$

where the plus and minus signs correspond to whether the acceptor atoms are N-terminal or C-terminal, respectively, to the dihedral-angle bond. The vector $\mathbf{R}_i^{\alpha H}$ represents the vector from the dihedral atom position $\mathbf{R}^\alpha$ to the position $\mathbf{R}_i^H$ of the i^{th} donor hydrogen. Substitution of Eq. 35 into Eq. 34 yields the final equation

$$\frac{\partial x_{ij}^O}{\partial \alpha} = \pm\left(\frac{1}{R_{ij}^{OH}}\right)(\mathbf{n}_\alpha \times \mathbf{R}_i^{\alpha H}) \cdot (\mathbf{u}_j^O - x_{ij}^O \mathbf{u}_{ij}^{OH}) \tag{36}$$

The equivalent formula for x_{ij}^H is obtained analogously

$$\frac{\partial x_{ij}^H}{\partial \alpha} = \mp\left(\frac{1}{R_{ij}^{OH}}\right)(\mathbf{n}_\alpha \times \mathbf{R}_j^{O\alpha}) \cdot (\mathbf{u}_i^H - x_{ij}^H \mathbf{u}_{ij}^{OH}) \tag{37}$$

where the vector $\mathbf{R}_j^{O\alpha}$ represents the vector from the position $\mathbf{R}_j^O$ of acceptor atom j to the dihedral atom position $\mathbf{R}^\alpha$. The plus and minus signs correspond to whether the acceptor atoms are N-terminal or C-terminal, respectively, to the dihedral-angle bond.

Substitution of Eq. 36 into the derivative Eq. 31 yields

$$\frac{\partial}{\partial \alpha} E^O = -\sum_{donors\ i<\alpha} \sum_{acceptors\ j\geq\alpha} \mathbf{B}_{ij}^O \cdot (\mathbf{n}_\alpha \times \mathbf{R}_i^{\alpha H}) + \sum_{donors\ i\geq\alpha} \sum_{acceptors\ j<\alpha} \mathbf{B}_{ij}^O \cdot (\mathbf{n}_\alpha \times \mathbf{R}_i^{\alpha H}) \tag{38}$$

where the vector $\mathbf{B}_{ij}^O$ is defined as

$$\mathbf{B}_{ij}^O = \left(\frac{\partial V^O}{\partial x_{ij}^O}\right)\left(\frac{1}{R_{ji}^{OH}}\right)(\mathbf{u}_j^O - x_{ij}^O \mathbf{u}_{ji}^{OH}) \tag{39}$$

Equation (38) can be simplified by separating the displacement $\mathbf{R}_i^{\alpha H}$ into its component parts

$$\mathbf{R}_i^{\alpha H} = \mathbf{R}_i^H - \mathbf{R}^\alpha \tag{40}$$

Substitution of Eq. 40 into the energy gradient (Eq. 38) yields the final result in the Gō notation

$$\frac{\partial}{\partial \alpha} E^O = \mathbf{n}_\alpha \cdot \mathbf{F}_\alpha + (\mathbf{n}_\alpha \times \mathbf{R}^\alpha) \cdot \mathbf{G}_\alpha \tag{41}$$

where the vector quantities are defined

$$\mathbf{F}_\alpha = \left\{ \sum_{donors\ i<\alpha} \sum_{acceptors\ j\geq\alpha} \mathbf{B}_{ij}^O \times \mathbf{R}_i^H \right\} - \left\{ \sum_{acceptors\ j<\alpha} \sum_{donors\ i\geq\alpha} \mathbf{B}_{ij}^O \times \mathbf{R}_i^H \right\} \quad (42)$$

$$\mathbf{G}_\alpha = \left\{ \sum_{donors\ i<\alpha} \sum_{acceptors\ j\geq\alpha} \mathbf{B}_{ij}^O \right\} - \left\{ \sum_{acceptors\ j<\alpha} \sum_{donors\ i\geq\alpha} \mathbf{B}_{ij}^O \right\} \quad (43)$$

The quantities $\mathbf{F}_\alpha$ and $\mathbf{G}_\alpha$ can be computed efficiently by using the forward recursion relations (Eqs. 5 and 6), where the four quantities are defined

$$\mathbf{f}_\alpha = \left\{ \sum_{donors\ i=\alpha} \sum_{acceptors\ j\neq\alpha} \mathbf{B}_{ij}^O \times \mathbf{R}_i^H \right\} - \left\{ \sum_{acceptors\ j=\alpha} \sum_{donors\ i\neq\alpha} \mathbf{B}_{ij}^O \times \mathbf{R}_i^H \right\} \quad (44)$$

$$\mathbf{g}_\alpha = \left\{ \sum_{donors\ i=\alpha} \sum_{acceptors\ j\neq\alpha} \mathbf{B}_{ij}^O \right\} - \left\{ \sum_{acceptors\ j=\alpha} \sum_{donors\ i\neq\alpha} \mathbf{B}_{ij}^O \right\} \quad (45)$$

with the initial condition $\mathbf{F}_0 = \mathbf{G}_0 = 0$. (The backward recursion may be used with equal efficiency.) By analogous reasoning, the x^H derivative can be written as

$$\frac{\partial}{\partial\alpha} E^H = \mathbf{n}_\alpha \cdot \mathbf{F}_\alpha + (\mathbf{n}_\alpha \times \mathbf{R}^\alpha) \cdot \mathbf{G}_\alpha \quad (46)$$

where the vector quantities are defined

$$\mathbf{F}_\alpha = \left\{ \sum_{donors\ i<\alpha} \sum_{acceptors\ j\geq\alpha} \mathbf{B}_{ij}^H \times \mathbf{R}_j^O \right\} - \left\{ \sum_{acceptors\ j<\alpha} \sum_{donors\ i\geq\alpha} \mathbf{B}_{ij}^H \times \mathbf{R}_j^O \right\} \quad (47)$$

$$\mathbf{G}_\alpha = \left\{ \sum_{donors\ i<\alpha} \sum_{acceptors\ j\geq\alpha} \mathbf{B}_{ij}^H \right\} - \left\{ \sum_{acceptors\ j<\alpha} \sum_{donors\ i\geq\alpha} \mathbf{B}_{ij}^H \right\} \quad (48)$$

The quantities $\mathbf{F}_\alpha$ and $\mathbf{G}_\alpha$ can be computed efficiently by using the forward recursion relations (Eqs. 5 and 6), where the four quantities are defined

$$\mathbf{f}_\alpha = \left\{ \sum_{donors\ i=\alpha} \sum_{acceptors\ j\neq\alpha} \mathbf{B}_{ij}^H \times \mathbf{R}_j^O \right\} - \left\{ \sum_{acceptors\ j=\alpha} \sum_{donors\ i\neq\alpha} \mathbf{B}_{ij}^H \times \mathbf{R}_j^O \right\} \quad (49)$$

$$\mathbf{g}_\alpha = \left\{ \sum_{donors\ i=\alpha} \sum_{acceptors\ j\neq\alpha} \mathbf{B}_{ij}^H \right\} - \left\{ \sum_{acceptors\ j=\alpha} \sum_{donors\ i\neq\alpha} \mathbf{B}_{ij}^H \right\} \quad (50)$$

and where the vector $\mathbf{B}_{ij}^H$ is defined as

$$\mathbf{B}_{ij}^H = \left(\frac{dV^H}{dx^H}\right)\left(\frac{1}{R_{ji}^{OH}}\right)(\mathbf{u}_j^H - x_{ij}^H \mathbf{u}_{ji}^{OH}) \quad (51)$$

Taken together, these formulas provide efficiently computed, analytic formulas for the gradient of an angular H-bond potential with respect to the backbone dihedral angles of a protein. These results can be extended to side-chain dihedral angles by the methods of Gō and coworkers.[8,9]

APPLICATIONS TO PROTEIN STRUCTURE REFINEMENT

We applied the refinement of these angular restraints to a low-resolution structure of the B1 domain of protein L from *Peptostreptococcus magnus* (PDB accession code 1hz6), which has been studied both theoretically and experimentally in our laboratory. The initial decoy had a C^α RMSD of 6.7 Å to the native structure; although the α-helix is well-formed, the β-sheet is poorly developed, having only four hydrogen bonds between β-strands compared with 22 such hydrogen bonds in the native structure [Fig. 3(a)]. The folding of β-sheets is known to be more challenging than that of α-helices because of the greater entropic cost in making the correct long-range contacts. Therefore, the refinement of such decoys is representative of problems encountered in NMR structure determination and protein structure prediction. We use ideal restraint data in the following tests of our methods, because we want merely to illustrate that the formulas and software are correct; further testing will be required to determine how best to apply these methods to practical structure refinements.

We generated ideal angle-restraint data by searching for C^α atoms separated by <7 Å in the native structure 1hz6. For each such pair m, we computed the cosine x_m^{nat} of the angle between their C^α—C^β unit vectors in the native structure; 166 vector-angle restraints were generated in this way. Starting from the initial 6.7 Å structure, we minimized the potential

$$E^{VA} = \sum_m (x_m - x_m^{nat})^2 \quad (52)$$

by using the Broyden–Fletcher–Goldfarb–Shanno method.[1] This minimization produced a structure that agreed with the native structure to 1.2 Å C^α RMSD. Moreover, the C^α—C^β vector-angle restraints sufficed to make 11 hydro-

Fig. 3. **a:** Initial low-resolution structure, which has a C^α RMSD of 6.7 Å to the true native structure. The four β-sheet hydrogen bonds are indicated in magenta, and the ribbon is colored from blue (N-terminus) to red (C-terminus). **b:** Refinement of 166 C^α—C^β vector-angle restraints produces a structure that agrees with the native structure to 1.2 Å C^α RMSD and has 11 of the 22 native β-sheet H-bonds. **c:** Minimization of 200 residual dipolar coupling restraints produces a structure that agrees with the native structure to 0.2 Å C^α RMSD and has all the correct H-bonds. **d:** Minimization of a statistical H-bond potential (Appendix C) for the 22 β-sheet hydrogen bonds succeeds in forming all of these H-bonds and reduces the C^α RMSD to 4.3 Å. Although topologically similar to the native structure (c), the β-strands of this structure bow and twist differently, and the α helix is displaced relative to the β-sheet.

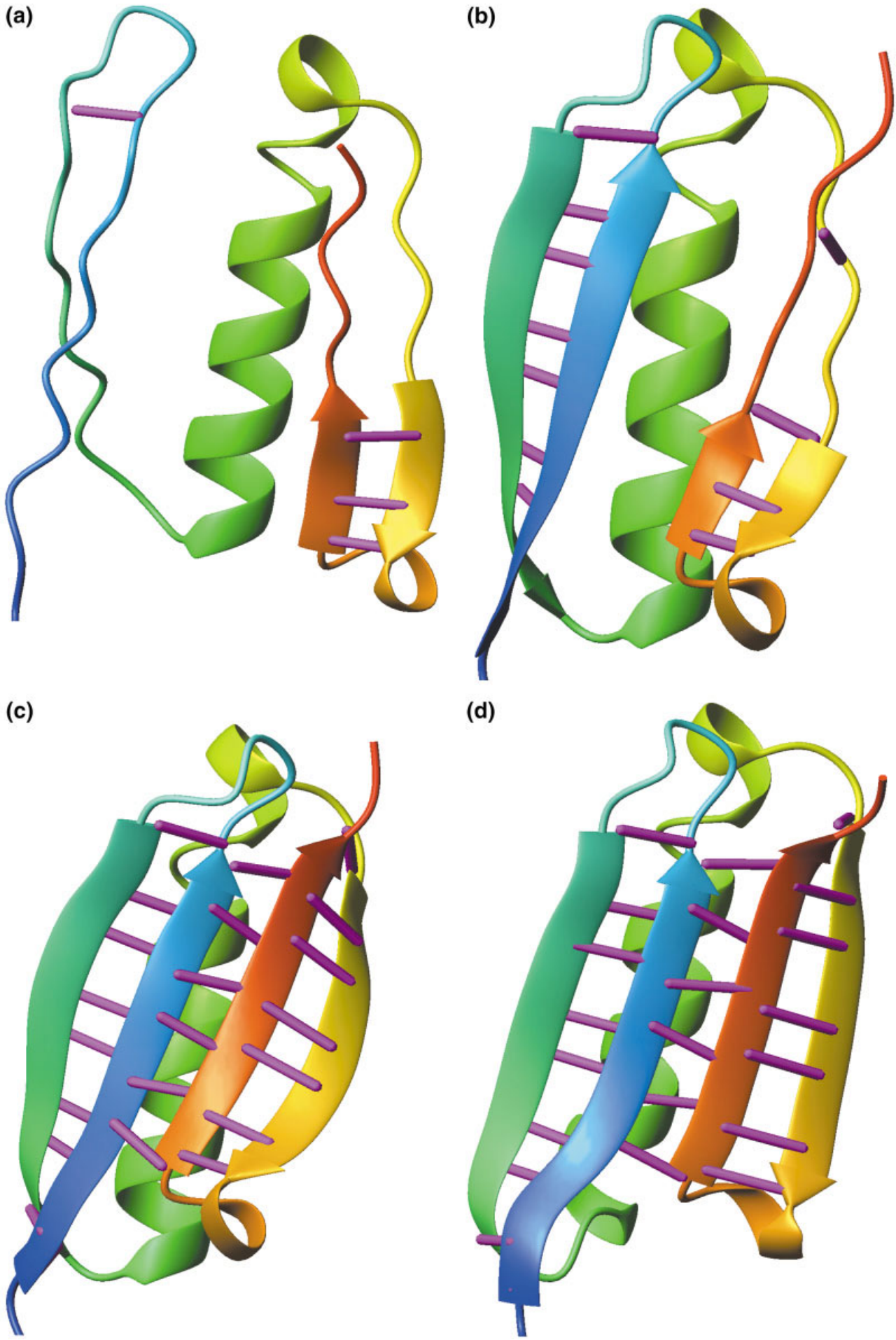

Figure 3.

gen bonds between the antiparallel β-strands (i.e., in the two hairpins), roughly a threefold improvement over the initial structure. However, no hydrogen bonds were created between the two central parallel β-strands. The entire minimization required <2 s of CPU time on a 1 GHz Intel Pentium III (Coppermine) processor.

We generated simulated residual dipolar coupling data for 1hz6 by assuming that the dipolar coupling tensor **D** was diagonal in the PDB coordinate frame of the native structure, using a $D_a = 1$ and a rhombicity of one third. Dipolar couplings D_u^{nat} were computed for the C^α—H^α, C^α—C', and C'—N vectors of the native structure, resulting in 200 restraints. The quadratic function

$$E^{DC} = \sum_{\mathbf{u}} (D_u - D_u^{nat})^2 \tag{53}$$

was minimized as above, where the dipolar tensor **D** was estimated by using the method of Losonczi et al.[38] and held fixed for each round of minimization. Twenty rounds of minimization were conducted; after each round, the dipolar tensor was reestimated and another round of minimization begun. The entire 20 rounds of minimization required roughly 40 s of CPU time on the 1 GHz processor. The final minimized structure differed from the native structure by 0.2 Å C^α RMSD and had all of its hydrogen bonds. Clearly, such good agreement should not be expected from experimental data, which are affected by measurement errors and dynamic internal motions of the protein. Nevertheless, it illustrates the potential effectiveness of the minimization methods described here.

Finally, we minimized the initial structure by using the 22 β-sheet hydrogen bonds found in the native structure. We used a differentiable, orientation-dependent H-bond potential (described briefly in Appendix C). The statistical potential was supplemented by a weak quadratic distance restraint between the hydrogen-bonded O′ and H atoms; the goal of this distance restraint is to bring these atoms into the range where the (relatively short-ranged) statistical potential can act effectively (i.e., under 2.5 Å). Starting from the initial 6.7 Å structure, minimization of the combined potential produces a structure that has all 22 of the restrained β-sheet hydrogen bonds and is qualitatively much more protein-like. The entire minimization required 12 s on the 1 GHz processor. The minimized structure has a C^α RMSD of 4.3 Å, a significant improvement for only 22 restraints. The residual errors lie in the twist of the β-sheet and in its disposition relative to the α helix. Such errors are not likely to be fixed by minimizing a generic hydrogen-bond potential but can of course be amended by adding other distance or angle restraints. These results support the idea that refinement of hydrogen bonds can improve high-resolution NMR structures.[39]

We have since applied our method to 14 other proteins with similarly good results. As seen for protein L, refinement of H-bonds typically results in modest, but significant, improvements in the C^α RMSD to the native, and significant alterations in the backbone structure from the starting decoy. However, refinement can also result in atomic clashes and unphysical backbone dihedral angles, especially for models with low topological similarity to the native protein; additional potentials are required to prevent such unphysical structures.

CONCLUSION

The minimization of three types of angle-dependent potentials has been studied, specifically those corresponding to vector-angle, dipolar-coupling, and hydrogen-bond restraints. Analytical formulas for the dihedral-angle gradient of all three types of angular potentials were derived and integrated with the efficient recursive derivative calculation methods of Gō and coworkers.[8,9] Refinement of vector-angle, dipolar-coupling, and H-bond potentials led to significant improvements in protein structure [Figs. 3(a)–(d)]. These methods can also be applied to other angular potentials whether physical, knowledge-based, or experimental constraints. The analytical formulas were implemented in software written in ANSI-compliant C and freely available (upon request from the corresponding author) to all members of the scientific community, being covered by the GNU General Public License.

ACKNOWLEDGMENTS

Figures 3(a)–(d) were prepared with MOLMOL.[40] WJW was supported by a postdoctoral fellowship from the Howard Hughes Medical Institute. We thank Drs. A. T. Brünger, P. Güntert, and especially G. M. Clore for answering questions about Xplor/CNS, DYANA, and Xplor-NIH and for providing the source code relevant to angular restraints in Xplor-NIH.

REFERENCES

1. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in C, 2nd ed. New York: Cambridge University Press; 1992. p 420–430.
2. Li Z, Scheraga HA. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. Proc Natl Acad Sci USA 1987;84:6611–6615.
3. Reif B, Hennig M, Griesinger C. Direct measurement of angles between bond vectors in high-resolution NMR. Science 1997;276: 1230–1233.
4. Schwalbe H, Carlomagno T, Hennig M, Junker J, Reif B, Richter C, Griesinger C. Cross-correlated relaxation for measurement of angles between tensorial interactions. Methods Enzymol 2001;338: 35–81.
5. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. Proc Natl Acad Sci USA 1995;92:9279–9283.
6. Prestegard JH, Al-Hashimi HM, Tolman JR. NMR structures of biomolecules using field oriented media and residual dipolar couplings. Quart Rev Biophys 2000;33:371–424.
7. Bax A, Kontaxis G, Tjandra N. Dipolar couplings in macromolecular structure determination. Methods Enzymol 2001;339:127–174.
8. Abe H, Braun W, Noguti T, Gō N. Rapid calculation of 1st and 2nd derivatives of conformational energy with respect to dihedral angles for proteins—general recurrent equations. Comput Chem 1984;8:239–247.
9. Braun W, Gō N. Calculation of protein conformations by proton proton distance constraints—a new efficient algorithm. J Mol Biol 1985;186:611–626.
10. Landau LD, Lifshitz EM. Mechanics, 3rd ed. New York: Pergamon Press; 1976. p 18–19.
11. Meiler J, Blomberg N, Nilges M, Griesinger C. A new approach for applying residual dipolar couplings as restraints in structure elucidation. J Biomol NMR 2000;16:245–252.
12. Skrynnikov NR, Kay LE. Assessment of molecular structure using

frame-independent orientational restraints derived from residual dipolar couplings. J Biomol NMR 2000;18:239–252.
13. Coulson CA. The hydrogen bond. In: Hadži D, editor. Hydrogen bonding. New York: Pergamon Press; 1959. p 339–360.
14. de Santis P, Giglio E, Liquori AM, Ripamonti A. van der Waals interaction and the stability of helical polypeptide chains. Nature 1965;206:456–458.
15. Simons KT, Ruczinski I, Kooperberg C, Fox BA, Bystroff C, Baker D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins 1999;34:82–95.
16. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. Use of ^{1}H-^{15}N and ^{1}H-^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat Struct Biol 1997;4:732–738.
17. Flory PJ, Yoon DY. Moments and distribution functions for polymer chains of finite length. I. Theory. J Chem Phys 1974;61: 5358–5365.
18. Wedemeyer WJ, Rohl CA, Scheraga HA. Exact solutions for chemical bond orientations from residual dipolar couplings. J Biomol NMR 2002;22:137–151.
19. Cordier F, Grzesiek S. Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{NC'}$ scalar couplings. J Am Chem Soc 1999;121:1601–1602.
20. Cornilescu G, Hu JS, Bax A. Identification of the hydrogen bonding network in a protein by scalar couplings. J Am Chem Soc 1999;121:2949–2950.
21. Grzesiek S, Cordier F, Dingley AT. Scalar couplings across hydrogen bonds. Methods Enzymol 2001;338:111–133.
22. Ptitsyn OB. Stagewise mechanism of self-organization of protein molecules. Dokl Akad Nauk SSSR 1973;210:1213–1215.
23. Levitt M, Warshel A. Computer simulation of protein folding. Nature 1975;253:694–698.
24. Hadži D, editor. Theoretical treatments of hydrogen bonding. New York: John Wiley and Sons; 1997.
25. Scheiner S. Hydrogen bonding: a theoretical perspective. New York: Oxford University Press; 1997.
26. Moulton WG, Kromhout RA. Nuclear magnetic resonance: structure of the amino group. II. J Chem Phys 1956;25:34–37.
27. Schroeder R, Lippincott ER. Potential function model of hydrogen bonds. II. J Phys Chem 1957;61:921–928.
28. Scott RA, Scheraga HA. Conformational analysis of macromolecules. III. Helical structures of polyglycine and poly-L-alanine. J Chem Phys 1966;45:2091–2101.
29. Ooi T, Scott RA, Vanderkooi G, Scheraga HA. Conformational analysis of macromolecules. IV. Helical structures of poly-L-alanine, poly-L-valine, poly-β-methyl-L-aspartate, poly-γ-methyl-L-glutamate, and poly-L-tyrosine. J Chem Phys 1967;46:4410–4426.
30. Hagler AT, Huler E, Lifson S. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. J Am Chem Soc 1974;96:5319–5327.
31. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4: 187–217.
32. Mayo SL, Olafson BD, Goddard WA III. DREIDING: a generic force field for molecular simulations. J Phys Chem 1990;94:8897–8909.
33. Gordon DB, Marshall SA, Mayo SL. Energy functions for protein design. Curr Opin Struct Biol 1999;9:509–513.
34. Fabiola F, Bertram R, Korostelev A, Chapman MS. An improved hydrogen bond potential: impact on medium resolution protein structures. Protein Sci 2002;11:1415–1423.
35. Lipsitz RS, Sharma Y, Brooks BR, Tjandra N. Hydrogen bonding in high-resolution protein structures: a new method to assess NMR protein geometry. J Am Chem Soc 2002;124:10621–10626.
36. Kortemme T, Morozov AV, Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. J Mol Biol 2003;326:1239–1259.
37. Bronshtein IN, Semendyayev KA. Handbook of mathematics. Frankfurt: Verlag Harri Deutsch; 1985. p 186–190.
38. Losonczi JA, Andrec M, Fischer MWF, Prestegard JH. Order matrix analysis of residual dipolar couplings using singular value decomposition. J Magn Reson 1999;138:334–342.
39. Spronk CAEM, Linge JP, Hilbers CW, Vuister GW. Improving the quality of protein structures derived by NMR spectroscopy. J Biomol NMR 2002;22:281–289.
40. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graphics 1996;14:51–55.
41. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The XplorNIH NMR molecular structure determination package. J Magn Reson 2003;160:65–73.
42. McGuire RF, Momany FA, Scheraga HA. Energy parameters in polypeptides. V. An empirical hydrogen bond potential function based on molecular orbital calculations. J Phys Chem 1972;76:375–393.

APPENDIX A
Displacement-Dihedral Potentials

A *displacement-dihedral* potential is defined as a restraint on the dihedral angle made by two fixed vectors and the displacement between them, e.g., the dihedral angle Ω^{OH} made by the vectors $\mathbf{u}^O$, $\mathbf{R}^{OH}$, and $\mathbf{u}^H$ of a hydrogen bond [Fig. 2(b)]. The spherical law of cosines[37] may be used to express the *cosine* of such dihedral angles in terms of vector dot products. Specifically, the spherical law of cosines states that the spherical angle Ω^{uvw} between three vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ obeys the equation

$$\cos\,\theta_{uw} = \cos\,\theta_{uv}\cos\,\theta_{vw} + \sin\,\theta_{uv}\sin\,\theta_{vw}\cos\,\Omega^{uvw} \quad (54)$$

where θ_{uv}, θ_{uw}, and θ_{vw} represent the angles between $\mathbf{u}$ and $\mathbf{v}$, between $\mathbf{u}$ and $\mathbf{w}$, and between $\mathbf{v}$ and $\mathbf{w}$, respectively. Using x_{uv}, x_{uw}, and x_{vw} to denote the cosines of these angles, we obtain the formula

$$\cos\,\Omega^{uvw} = \frac{x_{uw} - x_{uv}x_{vw}}{\sqrt{(1 - x_{uv}^2)(1 - x_{vw}^2)}} \quad (55)$$

Hence, a restraint on the cosine of the dihedral angle Ω^{uvw} can be expressed as a restraint $V(x_{uv}, x_{uw}, x_{vw})$ on the cosines of the three intervector angles, which may be treated by vector-angle or vector-displacement restraint methods above. The cosine of the dihedral angle determines the relative geometry of the three vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ up to a chirality, which can presumably be determined from other vector-angle restraints. Hence, most displacement-dihedral restraints should be replaceable by other restraints, such as vector-angle or vector-displacement restraints.

For completeness, a restraint $V(\sin\,\Omega^{uvw})$ on the *sine* of a dihedral angle Ω^{uvw} can be treated as follows. Such a sine restraint can be reformulated as a vector restraint by expressing it as a restraint $V^{DD}(p)$ on the triple scalar product p of the three vectors

$$p \equiv \mathbf{u} \cdot [\mathbf{v} \times \mathbf{w}] = |\mathbf{u}||\mathbf{v}||\mathbf{w}|\sin\,\theta_{uv}\sin\,\theta_{vw}\sin\,\Omega^{uvw} \quad (56)$$

Proceeding as before, the derivative of the dihedral-displacement restraint energy is

$$\frac{\partial}{\partial\alpha} E^{DD} = \sum_p \left(\frac{dV^{DD}}{dp}\right)\left(\frac{\partial p}{\partial\alpha}\right) \quad (57)$$

The derivative of V^{DD} with respect to p is assumed known, whereas the derivative of p with respect to α is given by the

$$\frac{\partial p}{\partial \alpha} = \left(\frac{\partial \mathbf{u}}{\partial \alpha}\right) \cdot [\mathbf{v} \times \mathbf{w}] \quad (58)$$

$$+ \mathbf{u} \cdot \left[\left(\frac{\partial \mathbf{v}}{\partial \alpha}\right) \times \mathbf{w}\right] \quad (59)$$

$$+ \mathbf{u} \cdot \left[\mathbf{v} \times \left(\frac{\partial \mathbf{w}}{\partial \alpha}\right)\right] \quad (60)$$

This formula can be brought into the form (Eq. 4) needed for the recursive methods of Gō and coworkers[8,9] by using the rotation formulas (Eqs. 1 and 35) and the vector identity $(\mathbf{A} \times \mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})$.

APPENDIX B
Angle-Dependent Potentials in Other Structure Determination Programs

The formulas for the gradients of vector-angle, vector-tensor, and vector-displacement restraints have not been described previously, and the most common structure determination programs (CNS, XPLOR and DYANA) do not provide routines for the continuous minimization of such restraints (A. T. Brunger, G. M. Clore, and P. Güntert, personal communications). However, after submission of this manuscript, the source code of the NIH version of Xplor (Xplor-NIH) became available to academic researchers under a licensing arrangement with Accelrys.[41] Examination of the code relating to angular restraints (kindly provided by Dr. Clore) reveals that Xplor-NIH does have continuous minimization routines for cross-correlated relaxations and residual dipolar coupling restraints; moreover, the gradients for these restraints are analytical and, hence, mathematically equivalent to those derived here, albeit different in form and implementation. The chief difference is that Xplor-NIH calculates its derivatives with respect to the Cartesian components of the atomic positions (consistent with Xplor's origins as a flexible-geometry molecular simulation package), which are then converted into dihedral-angle derivatives by a separate routine. This Cartesian approach is an excellent alternative to the direct vector approach used here and may provide simpler gradient formulas for some angular potentials. The direct vector approach is simpler for potentials involving vectors that do not correspond to atomic positions (e.g., those of Appendix C).

APPENDIX C
Statistical Hydrogen-Bond Potentials

This appendix describes the statistical hydrogen-bond potential used here, which is a variant of one published earlier from our laboratory.[36] In this work, we group hydrogen bonds into 12 classes with distinct geometric distributions, and an independent statistical potential was developed for each class. The classes of H-bonds are distinguished by the sequence separation and secondary-structure class of the H-bonded residues. (Sequence separation is defined as the number of residues from the acceptor to the donor; for illustration, a typical α-helical H-bond has a residue separation of +4, whereas the H-bonds of a β-turn have sequence separations ±3.) The 12 classes of hydrogen bonds were as follows: helical H-bonds with sequence separations +2, +3, and +4 (3 classes); turn H-bonds with residue separations +2, ±3, ±4, and ±5 (7 classes); and long-range parallel and antiparallel β-strand H-bonds between acceptors/donors separated by more than five residues (2 classes). The disadvantage of grouping all H-bonds into a single class (as done in some earlier statistical potentials) is that the potential is apt to be dominated by the alpha helical H-bonds, which have a very different angular distribution compared to, for example, long-range antiparallel β-strand H-bonds.

For each class of backbone H-bond, statistical potentials were computed as follows. A set of single-chain X-ray structures were culled from the PDB, having resolution 1.8 Å or better and an R-factor of 20% or better. Proteins under 40 residues were eliminated, as were structures that were overly diffuse (such as 3ezm). Hydrogens were added at 1.1 Å along the bisector of the C—N—C^{α} bonds. Roughly 180 thousand backbone H-bonds were recognized by the criteria of an O—H distance <2.15 Å and a C—O—H and O—H—N angles >90°; the relatively short distance was chosen to ensure that the H-bonds had a strong covalent (vs electrostatic) component and, hence, a well-defined angular dependence. After eliminating bifurcated H-bonds, the remaining H-bonds were grouped by their sequence separation, and the distribution of the $\mathbf{R}^{OH}$ unit vector in the coordinate frames fixed relative to the C^{α}—N—H and C^{α}—C—O peptide groups. Two distinct populations were observed for the sequence separations +2, +3, and +4, corresponding to the difference between the helical (2_7, 3_{10}, and α) and turn (γ, β, and α) conformations, and were separated into two classes accordingly. (H-bonds can be easily sorted into the helix/turn classes by the azimuthal angle of $\mathbf{R}^{OH}$ in C^{α}—C—O coordinate frame.) For residue separation +5, three distinct clusters were observed, corresponding to various forms of π helix/turns; however, they lie reasonably close together on the unit spheres, so we chose not to separate them into classes.

For each class of backbone H-bond, we computed statistical potentials for six degrees of freedom that are geometrically independent and specify the relative disposition of the two H-bonded peptide groups. (As usual, statistical potentials were derived by taking the negative logarithm of the probability distribution and fitting to a functional form.) First, we computed the mean unit vectors $\mathbf{u}^H$ and $\mathbf{u}^O$ of $\mathbf{R}^{OH}$ in the two coordinate frames C^{α}—N—H and C^{α}—C—O and reference vectors $\mathbf{v}^H$ and $\mathbf{v}^O$ that are perpendicular to the mean vectors and chosen to make the distribution of the dihedral (i.e., azimuthal) angles **v**-**u**-$\mathbf{R}^{OH}$ as even as possible. Thus, the coordinate frames defined by the **u** and **v** vectors are statistically optimized and are not aligned with any interatomic vector; this freedom in choosing the coordinate frame makes it relatively simple to model arbitrarily oriented elliptical distributions on the unit sphere. We then determined the distributions of the polar angles θ^H, Ω^H and θ^O, Ω^O in the coordinate frames defined by $\mathbf{u}^H$, $\mathbf{v}^H$ and $\mathbf{u}^O$, $\mathbf{v}^O$, respec-

tively, by binning (with light smoothing) in cos θ and Ω. The cos θ potentials were well-fit to a term linear in cos θ plus an exponential that decays rapidly as cos θ decreases from one, whereas the Ω potentials could be well-fit to a weighted sum of cos Ω and $\cos^2 \Omega$. For the fifth degree of freedom, we computed the statistical potential for the O—H distance, which was fit to a repulsive exponential function plus an attractive tanh function, and designed to be negligible above 2.5 Å. It is surprising that the traditional 10–12 potential[42] did not fit the distributions well, being overly "hard"; a softer 1–2 potential fit much better in the range considered (up to 2.15 Å) but seemed overly long-ranged. Finally, the sixth degree of freedom was taken to be the torsional angle Ω_{OH} of the two statistical vectors $\mathbf{u}^O$ and $\mathbf{u}^O$ about the displacement vector $\mathbf{R}^{OH}$; the distributions of Ω^{OH} were remarkably even and could be well-fit to a term linear in cos Ω^{OH}. Expanding the statistical potentials for Ω^H, Ω^O, and Ω^{OH} in powers of cos Ω has the advantage that the simpler methods of Appendix A can be used.

The distributions of the $\mathbf{R}^{OH}$ unit vector on the unit spheres is roughly elliptical, with a low eccentricity for most classes of H-bonds. Hence, the potentials for the dihedral angles Ω^H, Ω^O, and Ω^{OH} may be neglected to first approximation, because they are dominated by the potentials for O—H distance and the angles θ^H and θ^O. As a contrast, H-bonds with sequence separations +2 and +3 (both the helical and turn types) have highly eccentric distributions; however, such interactions may be augmented by short-range backbone dihedral-angle potentials (or specialized conformational sampling of the loop) to achieve similar distributions.