Eur Biophys J (2006) 35: 313–327
DOI 10.1007/s00249-005-0037-3

ARTICLE

Frank Gabel · Bernd Simon · Michael Sattler

A target function for quaternary structural refinement from small angle scattering and NMR orientational restraints

Received: 21 September 2005 / Revised: 25 November 2005 / Accepted: 5 December 2005 / Published online: 14 January 2006

Abstract We present a novel target function based on atomic coordinates that permits quaternary structural refinement of multi-domain protein–protein or protein–RNA complexes. It requires that the high-resolution structures of the individual domains are known and that small angle scattering (SAS) data as well as NMR orientational restraints from residual dipolar couplings (RDCs) of the complex are available. We show that, when used in combination, the translational and rotational restraints contained in SAS intensities and RDCs, respectively, define a target potential function that permits to determine the overall topology of complexes made up of domains with low internal symmetry. We apply the target function on a modestly anisotropic model system, the Barnase/Barstar complex, and discuss factors that influence the structural refinement such as data errors and the geometrical properties of the individual domains.

Keywords Small angle scattering · NMR · Residual dipolar couplings · Macromolecular complex · Annealing

Abbreviations CNS: Crystallography and NMR system · NMR: Nuclear magnetic resonance · NOE: Nuclear Overhauser effect · PDB: Protein data bank · RDCs: Residual dipolar couplings · RNA: Ribonucleic acid · SANS: Small angle neutron scattering · SAS: Small angle scattering

F. Gabel (✉) · B. Simon · M. Sattler (✉)
Structural and Computational Biology Group,
European Molecular Biology Laboratory,
Meyerhofstrasse 1, 69117 Heidelberg, Germany
E-mail: gabel@embl.de
Tel.: +49-6221-3878549
Fax: +49-6221-3878306
E-mail: sattler@embl.de
Tel.: +49-6221-3878552

Introduction

One of the greatest challenges in modern structural biology is the characterization of the topology of multi-domain macromolecular complexes that govern a major part of important cellular functions (Gavin and Superti-Furga 2003; Sali and Chiu 2005). Due to steric properties and inter-domain flexibility, crystallization of such complexes is not always easily accomplishable. Amongst the most promising approaches to this problem are multi-disciplinary methods, e.g., by combining X-ray crystallography of single domains with electron microscopy (Baumeister and Steven 2000) or small angle scattering (SAS; Koch et al. 2003) from entire complexes. Although some progress has been made in the recent years, in many cases these combinations suffer from the ambiguity of how to position high-resolution domain structures with the proper orientations into the low-resolution envelopes provided by the respective complementary technique (Rossmann et al. 2005). Nuclear magnetic resonance (NMR) residual dipolar couplings (RDCs; Tjandra et al. 1997a; Lipsitz and Tjandra 2004; Bax 2003; Prestegard et al. 2000) or heteronuclear relaxation data (Tjandra et al. 1997b; Dosset et al. 2000; Brüschweiler 2003) can provide such rotational restraints that define domain orientations but suffer from the inability to provide translational restraints, such as inter-domain distances or radii of gyration. A number of reports have described the utility of RDCs for defining domain orientations in protein complexes (Chou et al. 2001; Clore 2000; Clore and Schwieters 2003; Dobrodumov and Gronenborn 2003; Dominguez et al. 2003; Dosset et al. 2001; Jain et al. 2004; McCoy and Wyss 2002; Skrynnikov et al. 2000). In a recent approach using a grid search, RDCs and SAS data have been used in combination to refine the structure of the calmodulin/trifluoperazine complex (Mattinen et al. 2002). However, no target function has been derived that takes RDC and SAS restraints simultaneously into account for structural refinement. RDCs

and relaxation data are both well suited to provide information on domain orientations (Lipsitz and Tjandra 2004; Bax 2003; Prestegard et al. 2000; Dosset et al. 2000; Brüschweiler 2003).

In this paper, we present a target function based on atomic coordinates that permits to refine the topology of two domains in a complex using RDCs and minimal SAS data, provided that the high-resolution structures of both domains are known. Our approach exploits the fact that the distance between the centers of mass of two bodies is well defined if their individual radii of gyration and the radius of gyration of the complex they are forming are known (parallel-axes theorem; Goldstein 1977). With the orientation of both domains defined by RDCs, the problem of refining their translational positions with respect to SAS data is reduced to shift their centers of mass on a two-dimensional spherical surface determined by the radius of gyration of the complex.

Expanding the SAS intensity into a MacLaurin series, we develop a target function based on atomic coordinates and a driving force that optimize the domain positions on the sphere. Coordinate files of both domains have to be provided with orientational information from RDCs. Combined with these data, SAS intensities of the complex from a limited angular range (2–3 times the Guinier range) restrict the possible quaternary structures of the complex tremendously. The limitation to a small angular range constitutes an advantage of our algorithm compared to other SAS methods that need, in general, data covering a largest possible angular range (reviewed by Koch et al. 2003). We demonstrate the functionality of the target function on the (modestly anisotropic) Barnase/Barstar complex (Guillet et al. 1993; pdb ID "1BGS") as a model system. The efficiency and accuracy of the target function are discussed in terms of RDC and SAS errors and in terms of the geometrical properties of the domains.

Our approach is tailored to systems where domain orientations are known from RDCs. RDCs have become very popular over the recent years and are being successfully applied to an ever growing number of biological systems (Lipsitz and Tjandra 2004; Bax 2003; Prestegard et al. 2000). The target function derived can be incorporated as an energy term in structure calculation protocols using, e.g., CNS protocols (Brünger et al. 1998; Linge et al. 2003) providing an additional driving force to existing terms (NOE and H-bond restraints, etc.).

Although our approach aims at refining the relative position of two domains, it can be applied to multi-domain complexes under the condition that SAS and RDC data from two individual domains in a multi-domain complex are available. These can be experimentally obtained by deuterium labeling and contrast matching in small angle neutron scattering (SANS) experiments (Timmins and Zaccai 1988) and by appropriate isotope labeling schemes in NMR experiments (Gardner and Kay 1998; Mackereth et al. 2005) of the domains of interest.

Theory

Expansion of the scattering intensity

The angularly averaged SAS intensity $I(s)$ of an isolated particle consisting of a discrete set of scattering centers can be described by the Debye equation (Debye 1915):

$$I(s) = \sum_i \sum_j f(\mathbf{r}'_i) f(\mathbf{r}'_j) \frac{\sin(sr'_{ij})}{sr'_{ij}}, \tag{1}$$

$s = (4\pi/\lambda)\sin\,\theta$ is the absolute value of the scattering vector $\mathbf{s}$, λ is the wavelength of the scattered radiation (X-rays or neutrons) in Ångström and 2θ is the scattering angle. $\mathbf{r}'_i$ and $\mathbf{r}'_j$ denote the coordinates of the scattering centers (i.e., atoms) with respect to an arbitrary origin. $r'_{ij} = |\mathbf{r}'_i - \mathbf{r}'_j|$ is the absolute value of the distance vector between two scattering centers with scattering amplitudes $f(\mathbf{r}'_i)$ and $f(\mathbf{r}'_j)$. The sums in (1) run over all atoms of the particle. The sinc function can be expanded into a MacLaurin series (Feigin and Svergun 1987):

$$\frac{\sin(sr'_{ij})}{sr'_{ij}} = 1 - \frac{s^2 r'^2_{ij}}{6} + \frac{s^4 r'^4_{ij}}{120} - \frac{s^6 r'^6_{ij}}{5,040} + \frac{s^8 r'^8_{ij}}{362,880} - \cdots. \tag{2}$$

In the following, we make the assumption that $f(\mathbf{r}'_i) = f(\mathbf{r}'_j) = \text{const}$, i.e., that the particle can be described as a continuous body with a homogeneous scattering amplitude in the SAS approximation (in SANS, different $f(\mathbf{r}'_i)$ have to be used for proteins and RNAs). This is reasonable for biological macromolecules such as proteins where all occurring types of atoms (O, C, N, H, etc.) are, on a molecular scale, spatially uniformly distributed. The normalized scattering intensity, $I_{\text{norm}}(s) = I(s)/I(s=0)$, can then be expanded in the following form (N being the number of scattering atoms in the particle):

$$I_{\text{norm}}(s) = 1 + \frac{1}{N^2}\left(-\frac{s^2}{6}A + \frac{s^4}{120}B - \frac{s^6}{5,040}C + \frac{s^8}{362,880}D - \cdots\right). \tag{3}$$

With the coefficients

$$A = \sum_i \sum_j r'^2_{ij}, \quad B = \sum_i \sum_j r'^4_{ij},$$
$$C = \sum_i \sum_j r'^6_{ij}, \quad D = \sum_i \sum_j r'^8_{ij}, \cdots. \tag{4}$$

A is equal to twice the square of the radius of gyration, $2R_g^2$, of the particle.

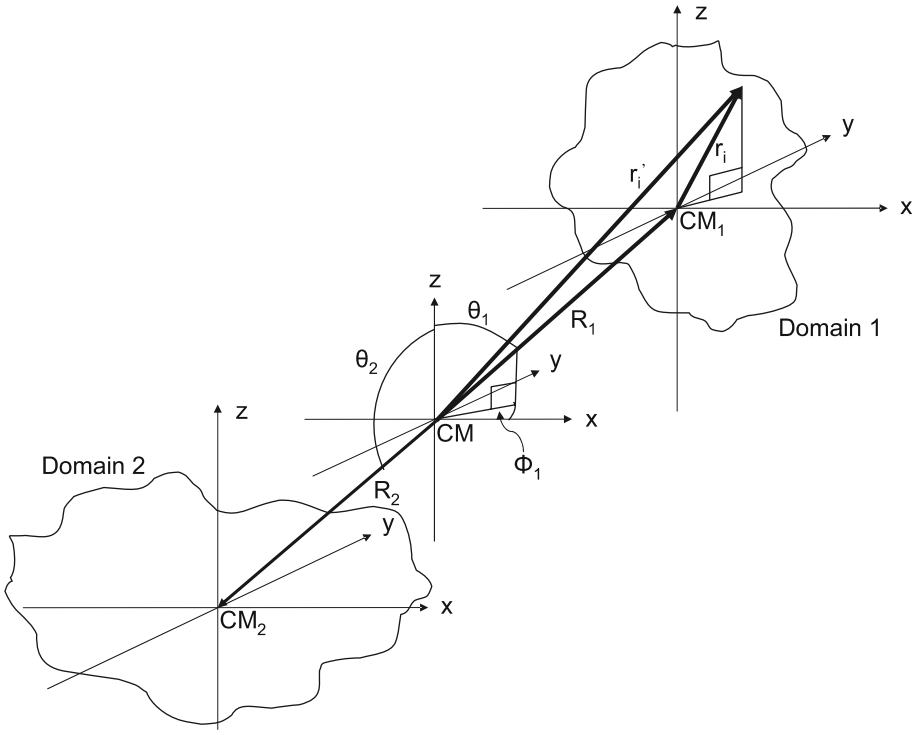

Fig. 1 Geometry and vector conventions of the two domain system

Description of the two domain geometry

We describe the two domains as illustrated in Fig. 1: The spatial coordinates $\mathbf{r'}_i$ of all atoms composing domain 1 are expressed by a sum of two vectors: the first pointing from the center of mass of the complex (CM) to the center of mass of domain 1 (CM_1), and the second pointing from CM_1 to the atomic position:

$$\mathbf{r}_i' = \mathbf{R}_1 + \mathbf{r}_i. \tag{5}$$

An analogous convention is made for atoms composing domain 2, $\mathbf{R}_1$ being replaced by $\mathbf{R}_2$. The angular pairs ($\theta_{1,} \phi_1$) and ($\theta_{2,} \phi_2$) define the vectors $\mathbf{R}_1$ and $\mathbf{R}_2$. Since ($\theta_{1,} \phi_1$) and ($\theta_{2,} \phi_2$) are not independent, a single set of angles (θ, ϕ) can be used to describe the positions of both domains:

$$\begin{aligned} \theta :=&\quad \theta_1 = \pi - \theta_2, \\ \phi :=&\quad \phi_1 = \phi_2 - \pi, \end{aligned} \tag{6}$$

θ and ϕ assume the values $0 \le \theta \le \pi$ and $0 \le \phi \le 2\pi$. Please note that the two domain-bound coordinate frames (x, y, z) and the (θ, ϕ) coordinate frame are pairwise axes-parallel.

Target functions for structural refinement

The coefficients A, B, C, ... relate the SAS intensity to the atomic coordinates of a particle and vice versa. In principle, the scattered intensity can be calculated for any particle from its atomic coordinates. However, due to the infinity of the expansion (3), this is not very practical. We therefore approximate the normalized SAS intensity at small scattering vectors s by a truncated expansion:

$$I_{\mathrm{norm}}(s) \approx 1 + \frac{1}{N^2}\left(-\frac{s^2}{6}A_{\mathrm{target}} + \frac{s^4}{120}B_{\mathrm{target}} - \frac{s^6}{5040}C_{\mathrm{target}}\right). \tag{7}$$

The coefficients A_{target}, B_{target} and C_{target} define the "target" structure, i.e., the actual structure of the particle in solution and can be determined, e.g., from fits of normalized experimental SAS intensities. Since $I_{\mathrm{norm}}(s)$ is an alternating series converging to zero, (7) describes $I_{\mathrm{norm}}(s)$ well as long as the D-term is small compared to the C-term:

$$\left|\frac{(s^8/362880)D}{-(s^6/5040)C}\right| = \frac{1}{72}\frac{\sum_i \sum_j r'^8_{ij}}{\sum_i \sum_j r'^6_{ij}}s^2 \le \frac{1}{72}\Delta^2_{\max}s^2 < 1, \tag{8}$$

$\Delta_{\max} \ge r'_{ij}$ is the maximum extension of the particle. On the other hand, the terms B_{target} and C_{target} must contribute notably to I_{norm} (s) with respect to the term A_{target}. This requires that the fit range expands beyond the so-called Guinier range (Guinier and Fournet 1955):

$$1 < R_g s. \tag{9}$$

The Guinier range is normally used to extract the radius of gyration, R_g, of a particle. Assuming $\Delta_{\max} \le 3R_g$, one obtains a combined condition (8) and (9) for the upper limit of the fit range:

$$1 < R_g s < 3. \tag{10}$$

To obtain the highest accuracy of the coefficients A_{target}, B_{target}, and C_{target}, a progressively increased s-range should be used to fit experimental data with a polynomial expansion of growing power in s. Beginning with $I(s) \approx 1 + (1/N^2)\left(-(s^2/6)A_{\text{target}}\right)$ in a very small s-range extracting A_{target}, the obtained value should be used to fit B_{target} with

$$I(s) \approx 1 + \frac{1}{N^2}\left(-\frac{s^2}{6}A_{\text{target}} + \frac{s^4}{120}B_{\text{target}}\right)$$

in a larger s-range and so on.

In the following, we express A, B, and C explicitly as functions of the domain atomic coordinates. We assume that the orientations of both domains are as determined by RDCs (Lipsitz and Tjandra 2004; Bax 2003; Prestegard et al. 2000). Domain orientational uncertainties associated with RDC restraints are treated in the "Discussion" section.

Target function $A = A_{target}$ and inter-domain distance

Knowing the radii of gyration of two bodies as well as the radius of gyration of the complex they form, the distance between their centers of mass is determined (parallel-axes theorem; Goldstein 1977). Substituting (5) into (4) for atoms from both domains and using the center of mass relationship $\sum_{i\in K_1} \mathbf{r}_i = \sum_{i\in K_2} \mathbf{r}_i = 0$, A can be expressed as follows:

$$\begin{aligned} A &= \sum_{i,j\in K_1,K_2} r'^2_{ij} = \sum_{i,j\in K_1} \left|\mathbf{r}_i - \mathbf{r}_j\right|^2 + \sum_{i,j\in K_2} \left|\mathbf{r}_i - \mathbf{r}_j\right|^2 + 2\sum_{i\in K_1, j\in K_2} \left|(\mathbf{R}_1 + \mathbf{r}_i) - (\mathbf{R}_2 + \mathbf{r}_j)\right|^2 \\ &= 2\left((N_1+N_2)\sum_{i\in K_1} r_i^2 + (N_1+N_2)\sum_{i\in K_2} r_i^2 + N_1N_2(R_1+R_2)^2\right), \end{aligned} \tag{11}$$

N_1 and N_2 are the number of atoms in domain 1 (K_1) and domain 2 (K_2), respectively. $R_1 = |\mathbf{R}_1|$ and $R_2 = |\mathbf{R}_2|$ are related by $N_1\ R_1 = N_2\ R_2$. Mixed sums, containing contributions of both domains, vanish due to the center of mass relationship. R_1 and R_2 are completely defined by $A = A_{\text{target}}$, allowing a first structural refinement of the domain positions to an inter-domain distance $R_1 + R_2$. CM_1 and CM_2 are thus confined to diametrically opposed points on a sphere of diameter $R_1 + R_2$. Further structural refinement using B_{target} and C_{target} optimizes the domain centers of mass on that sphere (i.e., refines θ and ϕ).

Target function $B = B_{target}$

B can be calculated in analogy to A using (4) and (5) and the center of mass relationship. It is composed of a constant term B_{const}, and a (θ, ϕ)-dependent term B (θ, ϕ):

$$B = \sum_{i,j\in K_1,K_2} r'^4_{ij} = B_{\text{const}} + B(\theta,\phi) \tag{12}$$

with

$$\begin{aligned} B_{\text{const}} =& 2\sum_{i,j\in K_1}\left[r_i^4 + r_i^2 r_j^2 + 2\left(\mathbf{r}_i\cdot\mathbf{r}_j\right)^2\right] \\ &+ 2\sum_{i,j\in K_2}\left[r_i^4 + r_i^2 r_j^2 + 2\left(\mathbf{r}_i\cdot\mathbf{r}_j\right)^2\right] \\ &+ 2\sum_{i\in K_1, j\in K_2}\left\{\left|\mathbf{r}_i - \mathbf{r}_j\right|^4 + \left|\mathbf{R}_2 - \mathbf{R}_1\right|^4\right. \\ &\left. + \left|\mathbf{r}_i - \mathbf{r}_j\right|^2\left|\mathbf{R}_2 - \mathbf{R}_1\right|^2\right\} \end{aligned} \tag{13}$$

and

$$\begin{aligned} B(\theta,\phi) = 4\sum_{i\in K_1, j\in K_2}\left\{-\left(\mathbf{r}_i - \mathbf{r}_j\right)\cdot\left(\mathbf{R}_2 - \mathbf{R}_1\right)\left|\mathbf{r}_i - \mathbf{r}_j\right|^2\right. \\ \left. + \left[\left(\mathbf{r}_i - \mathbf{r}_j\right)\cdot\left(\mathbf{R}_2 - \mathbf{R}_1\right)\right]^2\right\}, \end{aligned} \tag{14}$$

R_1 and R_2 are determined by $A = A_{\text{target}}$. $B(\theta, \phi)$ can be calculated explicitly ((26) in the appendix). The coefficients are moments of the domain Cartesian atomic coordinates and given in (27). The condition $B = B_{\text{target}}$ confines the positions of the domain centers of mass to one-dimensional lines on the sphere $A = A_{\text{target}}$ (compare Fig. 4). Starting from an arbitrary pair (θ, ϕ), the expression $B_{\text{target}} - B$ can be minimized. To this end, we have calculated a gradient field $\nabla B(\theta, \phi) = (\partial B/\partial\theta)\ \mathbf{e}_\theta + (\partial B/\partial\phi)\ \mathbf{e}_\phi$ (see (28)). $\mathbf{e}_\theta$ and $\mathbf{e}_\phi$ are unity vectors pointing tangentially to the sphere's surface in the directions of increasing θ and ϕ, respectively.

Target function $C = C_{target}$

Using (4) and (5), C can be calculated in analogy to A and B. It is composed of a constant term C_{const} and a (θ, ϕ)-dependent term $C(\theta, \phi)$:

$$C = \sum_{i,j\in K_1,K_2} r'^6_{ij} = C_{\text{const}} + C(\theta,\phi) \tag{15}$$

with

$$
\begin{aligned}
C_{\text{const}} = & 2\sum_{i,j\in K_1}\left[r_i^6+3r_i^2r_j^4-6\mathbf{r}_i\cdot\mathbf{r}_jr_i^2r_j^2+12\left(\mathbf{r}_i\cdot\mathbf{r}_j\right)^2r_i^2-4\left(\mathbf{r}_i\cdot\mathbf{r}_j\right)^3\right]\\
& +2\sum_{i,j\in K_2}\left[r_i^6+3r_i^2r_j^4-6\mathbf{r}_i\cdot\mathbf{r}_jr_i^2r_j^2+12\left(\mathbf{r}_i\cdot\mathbf{r}_j\right)^2r_i^2-4\left(\mathbf{r}_i\cdot\mathbf{r}_j\right)^3\right]\\
& +2\sum_{i\in K_1,j\in K_2}\left\{\left|\mathbf{r}_i-\mathbf{r}_j\right|^6+\left|\mathbf{R}_2-\mathbf{R}_1\right|^6+3\left[\left|\mathbf{r}_i-\mathbf{r}_j\right|^4\left|\mathbf{R}_2-\mathbf{R}_1\right|^2+\left|\mathbf{r}_i-\mathbf{r}_j\right|^2\left|\mathbf{R}_2-\mathbf{R}_1\right|^4\right]\right\}
\end{aligned}
\tag{16}
$$

and

$$
\begin{aligned}
C(\theta,\phi) = & 24\sum_{i\in K_1,j\in K_2}\left\{\left[\left(\mathbf{r}_i-\mathbf{r}_j\right)\cdot\left(\mathbf{R}_2-\mathbf{R}_1\right)\right]^2\left|\mathbf{R}_2-\mathbf{R}_1\right|^2\left[1+\frac{\left|\mathbf{r}_i-\mathbf{r}_j\right|^2}{\left|\mathbf{R}_2-\mathbf{R}_1\right|^2}\right]\right\}\\
& -24\sum_{i\in K_1,j\in K_2}\left\{\left(\mathbf{r}_i-\mathbf{r}_j\right)\cdot\left(\mathbf{R}_2-\mathbf{R}_1\right)\left|\mathbf{r}_i-\mathbf{r}_j\right|^2\left|\mathbf{R}_2-\mathbf{R}_1\right|^2\left[1+\frac{1}{2}\frac{\left|\mathbf{r}_i-\mathbf{r}_j\right|^2}{\left|\mathbf{R}_2-\mathbf{R}_1\right|^2}\right]\right\}\\
& -16\sum_{i\in K_1,j\in K_2}\left\{\left[\left(\mathbf{r}_i-\mathbf{r}_j\right)\cdot\left(\mathbf{R}_2-\mathbf{R}_1\right)\right]^3\right\}.
\end{aligned}
\tag{17}
$$

$C(\theta, \phi)$ can be explicitly calculated in analogy to $B(\theta, \phi)$ ((29) in the appendix). The coefficients are moments of the atomic coordinates of both domains in analogy to the B-potential but with higher powers involved (see (30)).

In analogy to the B-potential, $C_{\text{target}}-C$ defines one-dimensional lines on the sphere $A = A_{\text{target}}$. The gradient field $\nabla\, C(\theta, \phi) = (\partial C/\partial\theta)\, \mathbf{e}_\theta + (\partial C/\partial\phi)\, \mathbf{e}_\phi$ is given in the appendix (31).

The combined conditions $A = A_{\text{target}}$, $B = B_{\text{target}}$, and $C = C_{\text{target}}$ define sets of angular coordinates (θ, ϕ) for the domain positions (compare Fig. 4) that match the scattering behavior of the target structure to the sixth power of the scattering vector s.

Combined target function

We validate the potentials A, B, and C by refining the structure of the Barnase/Barstar complex (Guillet et al. 1993) with a sequential optimization against A_{target}, B_{target} and C_{target} (cf. "Material and methods" and "Results"). More generally, the structural refinement against A, B, and C can be performed simultaneously, e.g., in the form of a linearly combined potential Ψ, where $\lambda_{a,b,c}$ are weighing factors for the three potentials:

$$
\begin{aligned}
\Psi(\theta,\phi,R_1+R_2) = & \frac{\lambda_a}{6}\left|A-A_{\text{target}}\right|^2+\frac{\lambda_b}{120}\left|B-B_{\text{target}}\right|^2\\
& +\frac{\lambda_c}{5040}\left|C-C_{\text{target}}\right|^2.
\end{aligned}
\tag{18}
$$

Such a target function is better suited to be implemented into NMR structure calculation protocols using simulated annealing (Brünger et al. 1998; Tjandra et al. 1997a; Linge et al. 2003) than a sequential refinement. The target function generates a driving force

$$
\nabla\Psi(\theta,\phi,R_1+R_2)=\left(\frac{\partial}{\partial\theta}\Psi,\frac{\partial}{\partial\phi}\Psi,\frac{\partial}{\partial(R_1+R_2)}\Psi\right)
$$

in addition to existing terms, specifically for RDCs but also NOE-derived distance restraints and other terms. The advantage of the target function and the derived driving force is that they are directly based on atomic coordinates and can be easily used in parallel to domain-internal refinement. Knowing the coefficients A_{target}, B_{target} and C_{target}, the number and kind of possible domain positions can be easily calculated and depicted graphically (compare Fig. 4).

Materials and methods

Refinement protocol

We have implemented the potentials A, B, and C in a sequential structural refinement protocol using Python and C-shell scripts (available upon request). An implementation into XPLOR/CNS (Brünger et al. 1998; Linge et al. 2003) modules is under way for use in structural NMR calculation. We assume that the domain orientations are determined by RDC restraints and fixed.

Step 1 Minimization of $A_{target}-A$. R_1 and R_2 are calculated from the domain coordinate files and from A_{target} using (11) and $N_1\ R_1 = N_2\ R_2$. The domains are then translated to positions $\mathbf{R}_1$ and $\mathbf{R}_2$ which may be chosen arbitrarily but must point into anti-parallel directions.

Step 2 Minimization of $B_{target}-B$. Starting with the geometry obtained in step 1, $B_{\text{diff}} = B_{\text{target}}-B$ is determined. $B(\theta, \phi)$ is calculated from the current geometry using (13) and (26). The domains 1 and 2 are shifted parallely or anti-parallely to $\nabla\, B(\theta, \phi)$ by an incremental vector $k\nabla\, B(\theta, \phi)$, according to the sign of B_{diff}. k is different for domains 1 and 2 and is proportional to R_1 and R_2, respectively. After this tangential motion, a

motion radial to the new vectors $\mathbf{R}_1^* = \mathbf{R}_1 \pm k_1 \nabla B(\theta, \phi)$ and $\mathbf{R}_2^* = \mathbf{R}_2 \pm k_2 \nabla B(\theta, \phi)$ is performed for both domains in order to place their centers of mass at the correct distance $R_1 + R_2$. These two steps are iterated until $|B_{\text{diff}}|$ is smaller than a predefined value $B_{\min}$.

Step 3 Minimization of $C_{target} - C$. In this third refinement step, the parameter $C_{\text{target}} - C$ is refined by moving the domains by an incremental vector $\pm m\,\mathbf{T}$ along the equipotential line $B = B_{\text{target}}$. The direction of motion, $\pm\mathbf{T}$, is determined by the difference $C_{\text{diff}} = C_{\text{target}} - C$, and is orthogonal to the gradient of B:

$$\mathbf{T} = \left(\frac{\partial B}{\partial \theta} \mathbf{e}_{\varphi,} - \frac{\partial B}{\partial \varphi} \mathbf{e}_{\theta} \right). \tag{19}$$

This step is iterated until $|C_{\text{diff}}|$ is smaller than a predefined value $C_{\min}$. Correctional radial motions are interposed as in step 2. If incremental shifts by $\pm m\mathbf{T}$ lead to a geometry with $|B_{\text{diff}}| > B_{\min}$, step 2 is interposed and repeated until $|B_{\text{diff}}| < B_{\min}$. The refinement algorithm is stopped once both conditions, $|B_{\text{diff}}| < B_{\min}$ and $|C_{\text{diff}}| < C_{\min}$ are fulfilled simultaneously.

Step 4 Sorting of structural solutions. The combined conditions $A = A_{\text{target}}$, $B = B_{\text{target}}$, and $C = C_{\text{target}}$ define one or more structural solutions for the domain positions, i.e., pairs (θ, ϕ) for the domain centers of mass (compare Fig. 4). These solutions can be sorted according to the following criteria.

1. Steric and linker constraints: both domains are not allowed to interpenetrate (a fact not incorporated in steps 1–3). The maximum inter-domain distance may be limited by a linker. Further structural information, e.g., spin-labeling (Battiste and Wagner 2000) or chemical shift perturbations (Zuiderweg 2002) may as well be used at this stage.
2. The coefficient D may be calculated for the different solutions fitting an extended (7), including the D-term, to the experimental scattering intensity. As an alternative, the complete scattering curves from the different structural solutions can be calculated (Svergun et al. 1995) and compared to the experimental intensities. Both options will eliminate some of the equivalent solutions and help to find the structure that best matches the experimental information and the scattering intensity.

The Barnase/Barstar model system

We have applied the refinement protocol on a model system, the Barnase/Barstar complex (Guillet et al. 1993; pdb ID "1BGS", units B and C, molecular weight 21 kDa). The coordinate files contained 1,335 non-water atoms: 462 in domain 1 (Barstar) and 873 in domain 2 (Barnase). For each atom we assumed $f(\mathbf{r}_i) = \text{const} = 1$. The coordinate pair (θ, ϕ) described the position of the center of mass of the Barstar domain.

Shifting both domains translationally to arbitrary initial positions, we tried to retrieve the Barnase/Barstar structure as deposited in the protein data bank (i.e., the target structure) using the refinement protocol presented earlier. The target parameters A_{target}, B_{target}, and C_{target} were calculated from the deposited structure according to (4). We simulated RDC restraints by keeping the domain orientations of the deposited structure fixed during the refinement protocol and only allowing translational degrees of freedom.

Results

A, B- and C-potentials of the Barnase/Barstar complex

$R_1 = 17.48$ Å and $R_2 = 9.25$ Å were determined with $A = A_{\text{target}} = 627.4$, corresponding to $R_g = 17.7$ Å. The potentials B and C are represented graphically in Figs. 2 and 3. All numerical values given for the potentials are scaled by $1/N^2$ with $N = 1{,}335$. The units of the potentials are: $[A] = \text{Å}^2$, $[B] = \text{Å}^4$, and $[C] = \text{Å}^6$. B and C are topologically similar due to the analogies in (26) and (29). This point is discussed further in the "Discussion" section. B varied between a minimum value of 6.42×10^5 and a maximum value of 7.34×10^5 . The equipotential lines $B = B_{\text{target}} = 7.032\times10^5$ are depicted in Fig. 4. They correspond to a small, closed circle on the "northern" hemisphere ($\theta \leq \pi/2$) and an extended line on the "southern" hemisphere ($\theta \geq \pi/2$), assuming all possible values of ϕ. C varied between 0.835×10^9 and 1.226×10^9 . The equipotential lines $C = C_{\text{target}} = 1.062\times10^9$ are depicted in Fig. 4. They are topologically and spatially close to the lines $B = B_{\text{target}}$. Only two distinguished points in the (θ, ϕ)-plane fulfill the combined conditions $B = B_{\text{target}}$ and $C = C_{\text{target}}$: $(\theta, \phi) = (0.25/1.19)$ (target structure) and $(\theta, \phi) = (0.97/2.11)$ (an alternative structure). They are both situated in the northern hemisphere. The extended lines in the southern hemisphere do not intersect and there is no structural solution in this region. Both structural solutions have very similar D-values: $D = 1.90283\times10^{12}$ (target structure) and $D = 1.90276\times10^{12}$ (the alternative structure).

In addition to the equipotential lines $B = B_{\text{target}}$ and $C = C_{\text{target}}$, Fig. 4 depicts the angular coordinates (θ, ϕ) of $\mathbf{R}_1$ (Barstar) at different stages of the refinement process. The Barstar domain was initialized to $\mathbf{R}_1 = (0, 0, -R_1)$ and $\mathbf{R}_1 = (0, R_1, 0)$, corresponding to angular coordinate sets $(\theta, \phi) = (\pi/2, \pi)$ and $(\theta, \phi) = (\pi/2, \pi/2)$, respectively. After minimization of $B_{\text{target}} - B$ (refinement step 2), $\mathbf{R}_1$ is positioned on the line $B = B_{\text{target}}$ in both cases. In refinement step 3, $\mathbf{R}_1$ moves along the equipotential line $B = B_{\text{target}}$ until it matches the condition $C = C_{\text{target}}$. We checked the stability and the sampling of the refinement algorithm by randomizing the initial coordinate pairs (θ, ϕ) in the ranges $0 \leq \theta < \pi$ and $0 \leq \phi < 2\pi$ by the Python built-in routine *whrandom.random()*. The initial starting positions of the Barstar domain are depicted in black in Fig. 9 in the

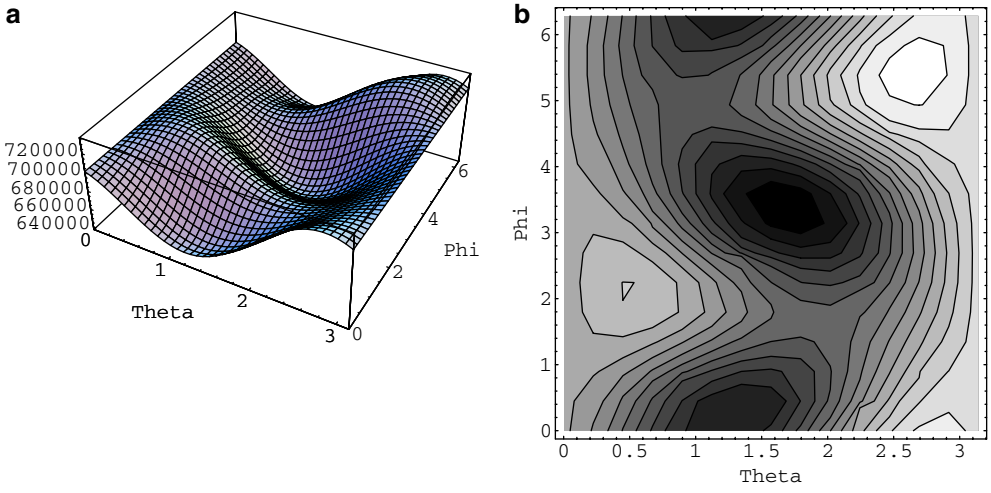

Fig. 2 Three-dimensional representation (*left*) and contour plot (*right*) of the *B*-potential (12) of the Barnase–Barstar complex as a function of (θ, ϕ), scaled with $1/N^2$, $N = 1{,}335$. In the contour plot, equipotential lines of *B* are traced

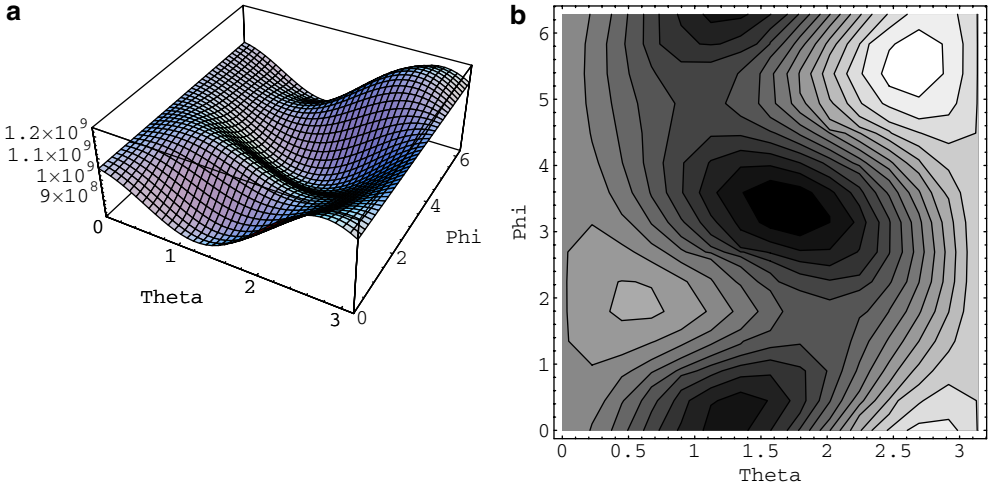

Fig. 3 Three-dimensional representation (*left*) and contour plot (*right*) of the *C*-potential (15) of the Barnase–Barstar complex as a function of (θ, ϕ), scaled with $1/N^2$, $N = 1{,}335$. In the contour plot, equipotential lines of *C* are traced

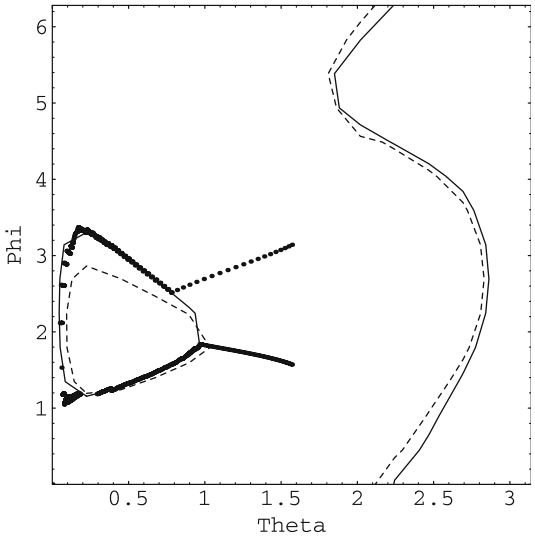

Fig. 4 Graphical representation of the equipotential lines $B = B_{\text{target}}$ (*continuous lines*) and $C = C_{\text{target}}$ (*dashed lines*) for the Barnase–Barstar complex as a function of (θ, ϕ) for $A = A_{\text{target}} = 627.4$. The ($\theta$, ϕ)-values denoting the center of mass coordinates of domain 1 (Barstar) during the structural refinement beginning at the two different starting points, $\mathbf{R}_1 = (0, 0, -R_1)$ and $\mathbf{R}_1 = (0, R_1, 0)$, are shown with *filled circles*

appendix. The distribution is homogeneous and explores the whole angular space uniformly. The minimized end positions (θ, ϕ) of the Barstar domain are also depicted in Fig. 9: out of 1,925 calculated structures, 1,021 (53%) were situated in the side-minimum (extended line in the southern hemisphere) and are depicted in red. None had converged after 1,000 refinement steps and their $|B_{\text{diff}}|$ and $|C_{\text{diff}}|$ were larger than those of the converged structures. The structures that converged at less than 1,000 refinement steps are depicted in green: 523 structures (27%) clustered around the target structure (θ, ϕ) = (0.25/1.19) and 380 structures (20%) clustered around the alternative solution (θ, ϕ) = (0.97/2.11). The clustering around both solutions was excellent. Out of the 523 structures clustering around the target position (θ, ϕ) = (0.25/1.19), only 44 lay outside the interval ($\theta \pm 0.1, \phi \pm 0.1$). For the alternative solution, the convergence was even better. A single outlier converging at (θ, ϕ) = (0.0/3.85) was observed, most probably trapped at the $\theta = 0$ singularity.

A spatial representation of the coordinate structures at different stages of the refinement pathway starting from $\mathbf{R}_1 = (0, R_1, 0)$ is given in Fig. 5. For all structures,

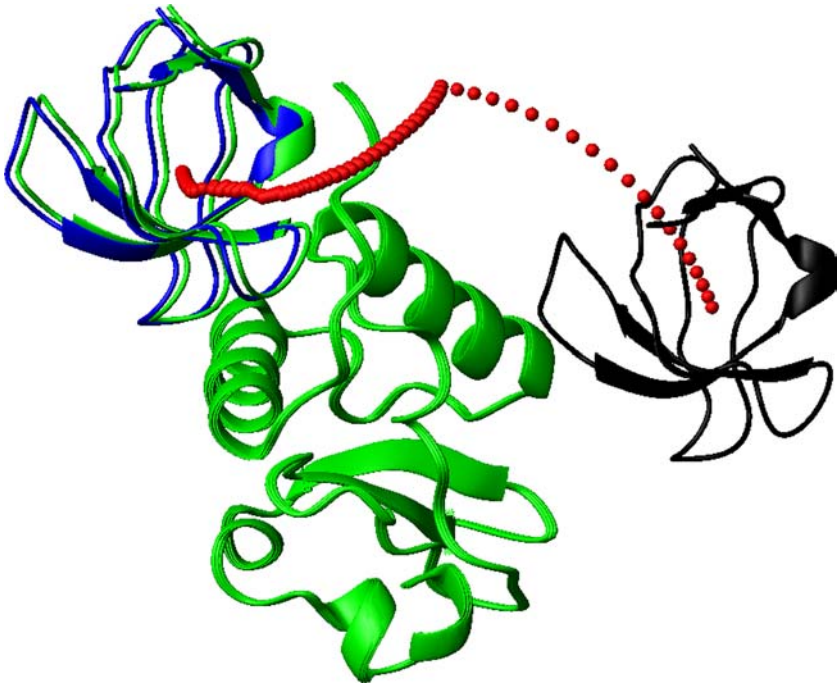

Fig. 5 Superposition of Barnase–Barstar structures at different stages of the refinement process starting at $\mathbf{R}_1 = (0, R_1, 0)$: the target structure is in *green*, the structure after the initialization step 1 is in *black* and the refined structure is in *blue*. The $\mathbf{R}_1$ coordinates are in *red*. The Barnase domain is superimposed for all structures and depicted in *green*

the larger Barnase domains have been superimposed. The two phases of the structural refinement are clearly distinguished by a kink in the $\mathbf{R}_1$ coordinate trace: the Barstar domain is first moved along the B-gradient until it encounters the $B = B_{\text{target}}$ line. It then moves tangentially to that line until it is situated close to its target position defined by the combined conditions $B = B_{\text{target}}$ and $C = C_{\text{target}}$. In both refinement pathways depicted in Fig. 4, the target structure as well as the alternative structure were retrieved with a precision of $(\theta, \phi) \pm (5,5)$ degrees.

Discussion

Solutions provided by the B- and C-potentials

The solutions of the structural refinement depend very much on the geometry of both domains, contained in (weighted) moments of their atomic coordinates (27) and (30). Generally speaking, the more anisotropic the domains are the more restricted and non-ambiguous the allowed domain positions are. For two point-symmetric domains the coefficients g_{B}, h_{B} and i_{B} as well as the coefficients g_{C} to s_{C} vanish since to each vector $\mathbf{r}_i$ there exists a vector $-\mathbf{r}_i$, i.e., the sums over all odd powers of the Cartesian coordinates are zero. In the special, very symmetric, case of two spherical but not necessarily identical particles, a structural refinement is not possible at all since the potentials B and C do not depend on (θ, ϕ) and a driving force is not defined: a structural refinement beyond the inter-domain distance is not possible in this case.

The solutions $B = B_{\text{target}}$ and $C = C_{\text{target}}$ are topologically similar in the (θ, ϕ)-diagram (cf. Fig. 4). Since $r_{ij}^2/(R_1 + R_2)^2 < 1$, the coefficients a to i are almost identical in (27) and (30), i.e., their corresponding terms in (26) and (29) are almost proportional. This and the fact that the coefficients j_{C} to s_{C} are differences of atomic coordinates and therefore small, explains the similarity of the graphical solutions $B = B_{\text{target}}$ and $C = C_{\text{target}}$. The larger the quotients $r_{ij}^2/(R_1 + R_2)^2$ and the larger the coefficients j_{C} to s_{C} are, the more both potentials differ. In terms of particle geometry, this means that the more the domains are anisotropic and extended and the more proximate their centers of mass are, the better their potentials B and C are suited for a structural refinement with our approach.

Effect of target coefficient errors

We have validated the refinement capacity of the target functions using coefficients A_{target}, B_{target}, and C_{target} calculated directly from the deposited pdb structure. In practice, these target coefficients are obtained from experimental SAS data by a polynomial regression (7), using a least χ^2 criterion. Their errors are two-fold: (1) errors from fitting an experimental SAS intensity with a truncated polynomial expansion and (2) experimental data errors.

Experimental data errors (with standard deviations σ_i) are automatically contained in a weighted least χ^2 regression over all experimental data points s_i:

$$\chi^2 = \frac{1}{n-1} \sum_i \left(\frac{I_{\text{norm}}(s_i) - I_{\text{fit}}(s_i)}{\sigma_i} \right)^2. \qquad (20)$$

$I_{\text{fit}}(s)$ is the regression from (7) and n is the number of experimental scattering vectors s_i.

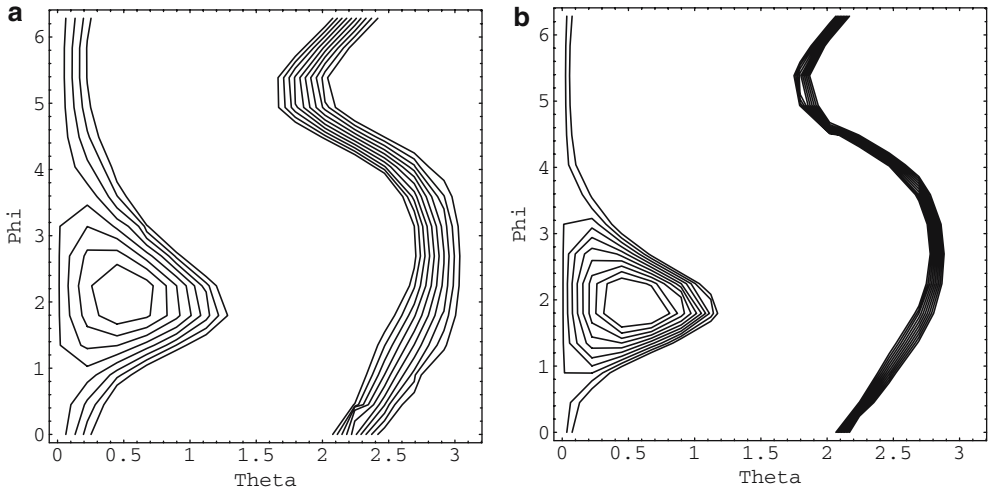

Fig. 6 Variations of Barnase/Barstar equipotential lines $B = B_{\text{target}}$ (*left*) and $C = C_{\text{target}}$ (*right*) when the target functions vary between $\pm 2\%$

The error $\Delta I(s)$ from fitting experimental data by not including terms greater than s^6 in the McLaurin series can be estimated by the absolute value of the member in s^8:

$$\Delta I(s) = \frac{1}{N^2} \sum_{i,j} \left| \frac{s^8 r'^8_{ij}}{362880} \right| \leq \frac{s^8 \Delta^8_{\max}}{362880} \tag{21}$$

$\Delta_{\max}$ is the maximal spatial extension of the complex. Assuming 3 $R_g < \Delta_{\max}$, $\Delta I(s)$ is smaller than 2% of $I_{\text{norm}}(s)$. In this case (7) can be safely applied to experimental data up to three times the Guinier range.

In order to investigate errors of A_{target}, B_{target}, and C_{target}, we have plotted the equipotential lines for $B = B_{\text{target}} \pm 2\%$ and $C = C_{\text{target}} \pm 2\%$ in Fig. 6. Both equipotential lines display a notable dependence on the target values. Fortunately, the variations of the target values are not independent when obtained from fits of experimental scattering curves: since B_{target} and C_{target} have opposite signs (7), they will increase or decrease simultaneously. This property increases the stability of the structural solutions obtained with our method against small variations of the target values. In addition, the topology and spatial extension of the equipotential lines on the northern hemisphere appear distorted due to their proximity to the pole $\theta = 0$ and are actually confined to a small spatial region close to the pole.

A combined potential function Ψ (18) will, even for errors of A_{target}, B_{target}, and C_{target} of up to ~5% and for not too anisotropic particles like the Barnase/Barstar complex, allow a structural refinement: the combined conditions $A = A_{\text{target}}$, $B = B_{\text{target}}$, and $C = C_{\text{target}}$ restrict the possible domain positions to the vicinity of the lines $B = B_{\text{target}}$ and $C = C_{\text{target}}$ and an inter-domain distance close to $R_1 + R_2$. The minimum of the target function Ψ and the refined structure are thus confined to this restricted conformational space, making a sampling of the whole angular range unnecessary. Linker or steric constraints as well as complementary structural information, e.g., from mutational analysis, NMR chemical shift interface mapping (Zuiderweg 2002) or spin labeling (Battiste and Wagner 2000), restrict the conformational space for structural solutions further and allow to distinguish between equivalent solutions.

Influence of domain orientational uncertainties

The solutions provided by the refinement potentials A, B, and C depend on the domain orientations. How do domain orientational uncertainties influence the refined structures? A change in the domain orientations can be described by a rotation of one domain by an Euler matrix **A**. The individual atomic coordinates undergo the following transformation:

$$\begin{pmatrix} x'_i \\ y'_i \\ z'_i \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}. \tag{22}$$

With the Euler matrix **A**:

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} = \begin{pmatrix} \cos\gamma\cos\beta\cos\alpha - \sin\gamma\sin\alpha & \cos\gamma\cos\beta\sin\alpha + \sin\gamma\cos\alpha & -\cos\gamma\sin\beta \\ -\sin\gamma\cos\beta\cos\alpha - \cos\gamma\sin\alpha & -\sin\gamma\cos\beta\sin\alpha + \cos\gamma\cos\alpha & \sin\gamma\sin\beta \\ \sin\beta\cos\alpha & \sin\beta\sin\alpha & \cos\beta \end{pmatrix}. \tag{23}$$

This matrix describes a clockwise rotation around axis z by an angle α, followed by a clockwise rotation around the new axis y by an angle β and finally a clockwise

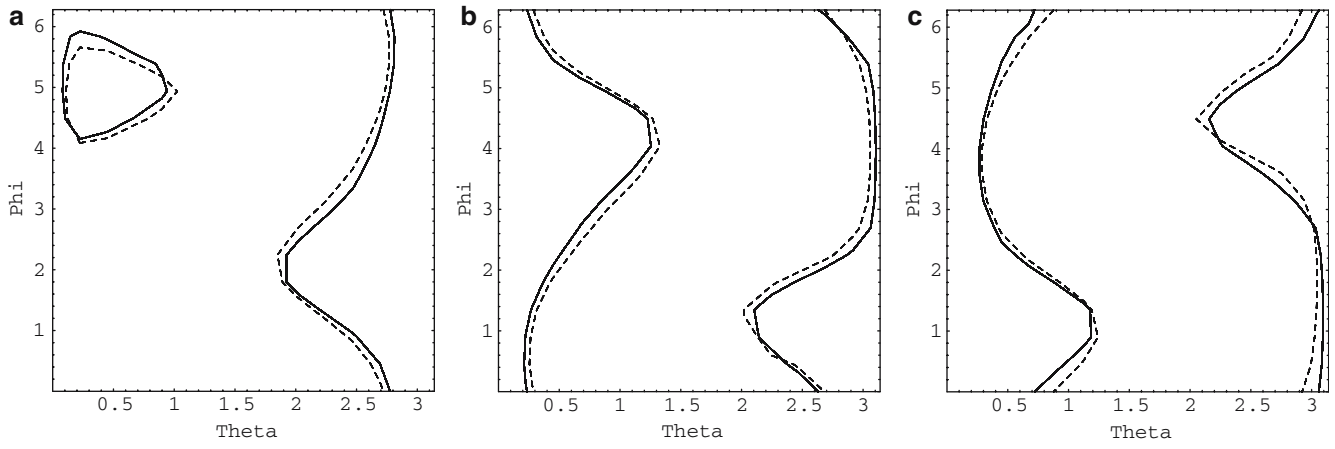

Fig. 7 Equipotential lines $B = B_{\text{target}}$ and $C = C_{\text{target}}$ (*continuous* and *dashed* lines, respectively) of the Barnase/Barstar complex of Barnase domain flips as defined in the text. *Left*: $\alpha = \pi$, $\beta = \gamma = 0$; *middle*: $\alpha = \gamma = 0$, $\beta = \pi$; and *right*: $\alpha = 0$, $\beta = \gamma = \pi$

rotation around the new axis z by an angle γ. (Conventions adopted from Arfken and Weber 1995; angles are in radians.) We discuss two special cases: domain flips and small rotations.

Domain flips

In general, four different domain orientations, corresponding to 180° flips of a single domain, are possible for RDCs in an alignment frame (Prestegard et al. 2000). What conditions are required on the domain geometries to discriminate them by SAS data? The four possible orientations can be described by sets of angles α, β, and γ in the coordinate frame: (1) $\alpha = \beta = \gamma = 0$, (2) $\alpha = \pi$, $\beta = \gamma = 0$, (3) $\alpha = \gamma = 0$, $\beta = \pi$, and (4) $\alpha = 0$, $\beta = \gamma = \pi$. In each case, the atomic coordinates of the rotated domain are transformed as follows:

$$\text{(1) Identity,} \quad \text{(2)} \begin{pmatrix} x'_i \\ y'_i \\ z'_i \end{pmatrix} = \begin{pmatrix} -x_i \\ -y_i \\ z_i \end{pmatrix},$$

$$\text{(3)} \begin{pmatrix} x'_i \\ y'_i \\ z'_i \end{pmatrix} = \begin{pmatrix} -x_i \\ y_i \\ -z_i \end{pmatrix}, \quad \text{(4)} \begin{pmatrix} x'_i \\ y'_i \\ z'_i \end{pmatrix} = \begin{pmatrix} x_i \\ -y_i \\ -z_i \end{pmatrix}.$$

Two domain orientations, generated by a flip, can therefore be distinguished by their B- and C-potentials in the absence of a rotational symmetry of π around the rotation axis that transfers one structure into the other. The potentials corresponding to the transformations (2), (3) and (4) are depicted in Fig. 7 for the Barnase/Barstar complex. While transformation (2) affects only the ϕ-dependence, transformations (3) and (4) affect, via the coordinate z, also the θ-dependence of the potential. In the light of these results, it is preferable that the fourfold orientational degeneracy of the RDCs is resolved prior to SAS refinement. This can be achieved by measuring RDCs in two or more alignment media (Bax 2003).

Small errors of domain orientations

Small uncertainties of the domain orientations can be described by small Euler angles α, β, and γ. Approximating $\sin x = x$ and $\cos x = 1$, the Euler matrix can be written as

$$\mathbf{A} \approx \begin{pmatrix} 1 - \alpha\gamma & \alpha + \gamma & -\beta \\ -(\alpha + \gamma) & 1 - \alpha\gamma & \beta\gamma \\ \beta & \alpha\beta & 1 \end{pmatrix} \approx \begin{pmatrix} 1 & \alpha + \gamma & -\beta \\ -(\alpha + \gamma) & 1 & 0 \\ \beta & 0 & 1 \end{pmatrix}. \tag{24}$$

In this approximation, x'_i, y'_i and z'_i are expressed as linear combinations of the old coordinates:

$$\begin{pmatrix} x'_i \\ y'_i \\ z'_i \end{pmatrix} \approx \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \begin{pmatrix} (\alpha + \gamma)y_i - \beta z_i \\ -(\alpha + \gamma)x_i \\ \beta x_i \end{pmatrix}. \tag{25}$$

We calculated the modifications of the B and C-potentials explicitly for rotations of $\alpha = \pm 0.1$ and $\beta = \pm 0.1$ radians of the Barnase domain. Two of the resulting potentials are depicted in Fig. 8 ($\alpha = +0.1$ and $\beta = -0.1$). The solutions $B = B_{\text{target}}$ and $C = C_{\text{target}}$ were within an angular range of $\pm(0.1/0.1)$ radians (i.e., $\pm(5/5)$ degrees) of the original solutions (Fig. 4; a similar result was obtained for $\alpha = -0.1$ and $\beta = +0.1$). Therefore, small rotations of the Barnase domain do not affect the topology of the refined solutions drastically. Our results demonstrate the stability of the target potentials in the presence of minor ($< 5°$) domain orientational uncertainties that may result from experimental RDCs.

Comparison to existing algorithms and programs

The main advantages of our algorithm with respect to other refinement programs (reviewed by Koch et al.

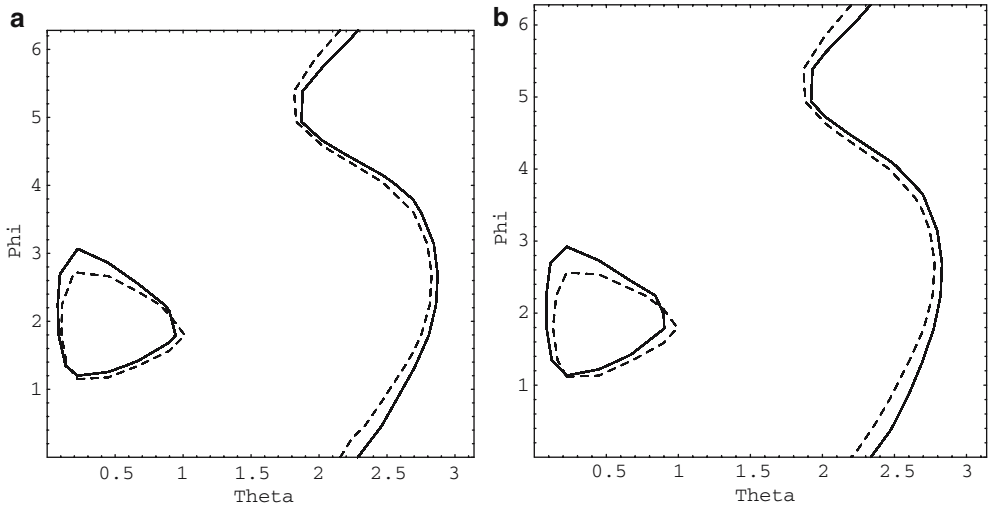

Fig. 8 Equipotential lines for $B = B_{target}$ and $C = C_{target}$ (*continuous* and *dashed lines*, respectively) after a small Barnase domain rotation according to (25): *Left*: $\alpha = +0.1$. *Right*: $\beta = -0.1$

2003) are: (1) the experimental scattering data needed are restricted to small angles, requiring only data from about 2 to 3 times the Guinier range; (2) only a limited conformational sampling of domain positions against scattering data is required; (3) the scattering curves of the individual domains do not have to be known or calculated; (4) the use of driving forces based on atomic coordinates facilitates the parallel use of our approach with domain-internal refinement protocols (Chou et al. 2000; Sibille et al. 2001).

It is important to state that the approach presented here is not limited to two domain systems: by using specific deuteration in SANS (Timmins and Zaccai 1988) and appropriate isotope labeling schemes for NMR, multi-domain complexes comprising more than two subunits are accessible to the method by pair-wise domain structural refinement.

Our approach is tailored to protein–protein or protein–RNA complexes with orientational restraints provided by RDCs. For example, Mattinen et al. (2002), used a combination of RDC restraints and a grid search against SAS data to determine the quaternary structure of a calmodulin/trifluoperazine complex. However, efforts to combine translational restraints from SAS and rotational restraints from RDCs in a general way have, to our knowledge, not been successful up to date. Using a target function built on RDC and SAS restraints, we have presented for the first time an integrated solution to this refinement problem that is applicable to a large number of multi-domain macromolecular complexes. The combined target function Ψ (18), based on atomic coordinates, is very suited to be implemented as an additional energy term in existing annealing protocols (Brünger et al. 1998; Tjandra et al. 1997a).

Conclusions

We have presented a novel target function based on atomic coordinates for multi-domain quaternary refinement. It combines translational restraints from SAS with rotational restraints from NMR RDCs. Our refinement approach is tailored to problems where high-resolution structures and orientations of individual domains in multi-domain complexes are known. Provided that NMR RDCs are available, the SAS data needed to solve the complex topology are limited to small angles (~2–3 times the Guinier range). We have demonstrated the functionality of the target function on a model system, the Barnase/Barstar complex. The influences of experimental data errors as well as the domain geometries on the stability and efficiency have been discussed. The refinement protocol has been shown to be robust against small experimental data errors.

Acknowledgements The authors thank Drs. Dmitri Svergun and Giuseppe Zaccai for critical reading of the manuscript. This work was supported by the European Union (3D repertoire contract LSHG-CT-2005-512028).

Appendix

Coefficients and partial derivatives of the *B*- and *C*-potentials are

Distribution of initial and refined Barstar positions

See Fig. 9.

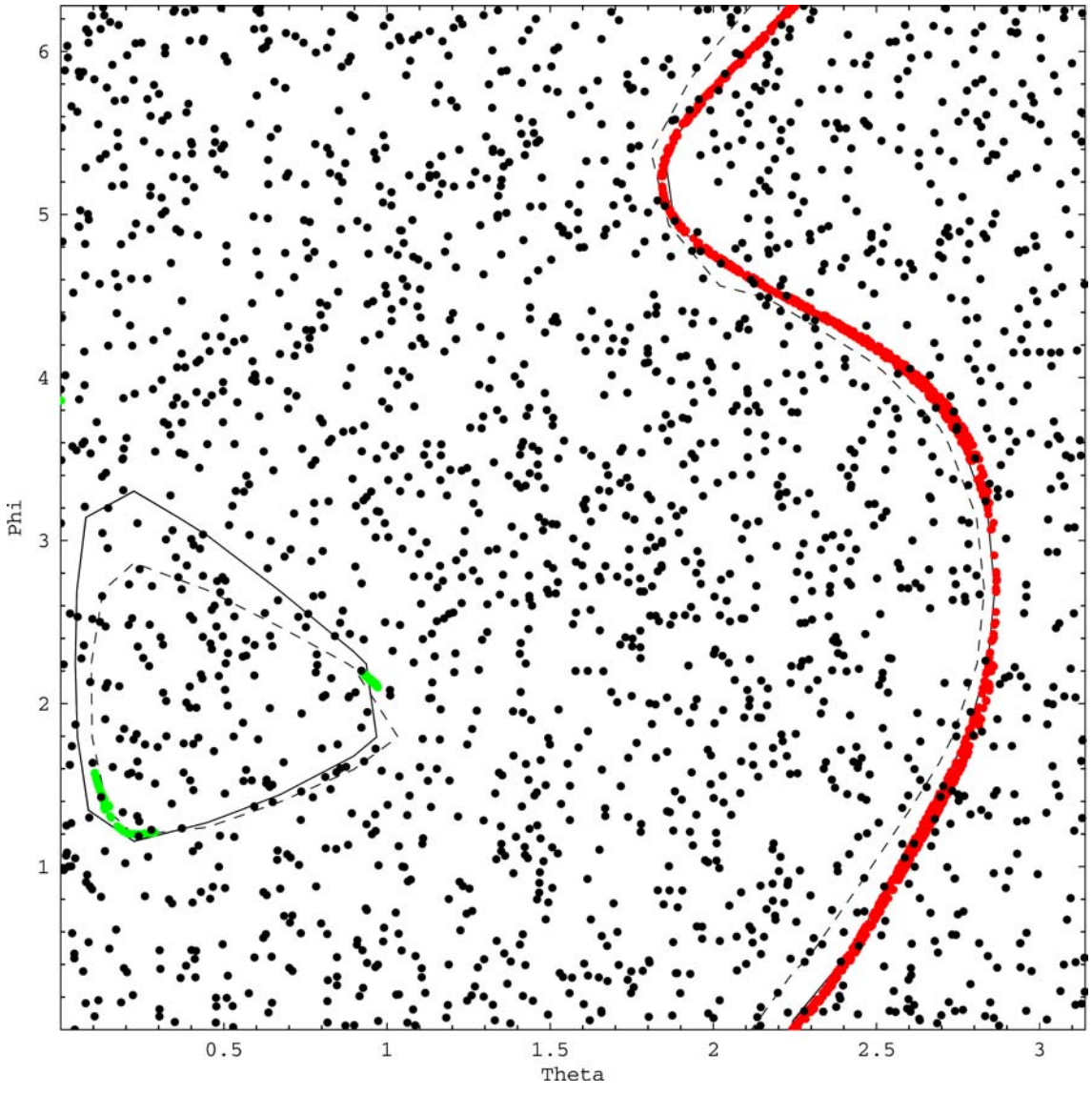

Fig. 9 Distribution of the initial and of the refined end position coordinates (θ, ϕ) of the Barstar domain. The initial positions are depicted in *black*, structures that converged at less than 1,000 refinement steps in *green*, and structures that had not converged after 1,000 steps are depicted in *red*

$$
\begin{aligned}
B(\theta,\phi) &= 4(R_1+R_2)^2\left[\begin{array}{l}\sin^2\theta\left(a_{\mathrm{B}}\cos^2\phi + b_{\mathrm{B}}\sin^2\phi + c_{\mathrm{B}}\sin\phi\cos\phi\right) + d_{\mathrm{B}}\cos^2\theta \\ \quad + \sin\theta\cos\theta\left(e_{\mathrm{B}}\cos\phi + f_{\mathrm{B}}\sin\phi\right)\end{array}\right] \\
&\quad + 4(R_1+R_2)[\sin\theta(g_{\mathrm{B}}\cos\phi + h_{\mathrm{B}}\sin\phi) + i_{\mathrm{B}}\cos\theta], \qquad (26)\\
a_{\mathrm{B}} &= \sum_{i\in K_1, j\in K_2}\left(x_i^2 + x_j^2\right), \quad b_{\mathrm{B}} = \sum_{i\in K_1, j\in K_2}\left(y_i^2 + y_j^2\right), \quad c_{\mathrm{B}} = 2\sum_{i\in K_1, j\in K_2}\left(x_i y_i + x_j y_j\right),\\
d_{\mathrm{B}} &= \sum_{i\in K_1, j\in K_2}\left(z_i^2 + z_j^2\right), \quad e_{\mathrm{B}} = 2\sum_{i\in K_1, j\in K_2}\left(x_i z_i + x_j z_j\right), \quad f_{\mathrm{B}} = 2\sum_{i\in K_1, j\in K_2}\left(y_i z_i + y_j z_j\right),\\
g_{\mathrm{B}} &= \sum_{i\in K_1, j\in K_2}\left(x_i^3 + y_i^2 x_i + z_i^2 x_i - x_j^3 - y_j^2 x_j - z_j^2 x_j\right), \qquad (27)\\
h_{\mathrm{B}} &= \sum_{i\in K_1, j\in K_2}\left(x_i^2 y_i + y_i^3 + z_i^2 y_i - x_j^2 y_j - y_j^3 - z_j^2 y_j\right),\\
i_{\mathrm{B}} &= \sum_{i\in K_1, j\in K_2}\left(x_i^2 z_i + y_i^2 z_i + z_i^3 - x_j^2 z_j - y_j^2 z_j - z_j^3\right),
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial}{\partial\phi}B(\theta,\phi) &= 4(R_1+R_2)^2\begin{bmatrix}\sin^2\theta\big(-2a_{\mathrm{B}}\cos\phi\sin\phi+2b_{\mathrm{B}}\sin\phi\cos\phi+c_{\mathrm{B}}\big(\cos^2\phi-\sin^2\phi\big)\big)\\ +\sin\theta\cos\theta(-e_{\mathrm{B}}\sin\phi+f_{\mathrm{B}}\cos\phi)\end{bmatrix}\\
&\quad+4(R_1+R_2)[\sin\theta(-g_{\mathrm{B}}\sin\phi+h_{\mathrm{B}}\cos\phi)],\\
\frac{\partial}{\partial\theta}B(\theta,\phi) &= 4(R_1+R_2)^2\begin{bmatrix}2\sin\theta\cos\theta\big(a_{\mathrm{B}}\cos^2\phi+b_{\mathrm{B}}\sin^2\phi+c_{\mathrm{B}}\sin\phi\cos\phi\big)\\ -2d_{\mathrm{B}}\cos\theta\sin\theta+\big(\cos^2\theta-\sin^2\theta\big)(e_{\mathrm{B}}\cos\phi+f_{\mathrm{B}}\sin\phi)\end{bmatrix}\\
&\quad+4(R_1+R_2)[\cos\theta(g_{\mathrm{B}}\cos\phi+h_{\mathrm{B}}\sin\phi)-i_{\mathrm{B}}\sin\theta],
\end{aligned}\tag{28}$$

$$\begin{aligned}
C(\theta,\phi) &= 24(R_1+R_2)^4\begin{bmatrix}\sin^2\theta\big(a_{\mathrm{C}}\cos^2\phi+b_{\mathrm{C}}\sin^2\phi+c_{\mathrm{C}}\sin\phi\cos\phi\big)+d_{\mathrm{C}}\cos^2\theta\\ +\sin\theta\cos\theta(e_{\mathrm{C}}\cos\phi+f_{\mathrm{C}}\sin\phi)\end{bmatrix}\\
&\quad+24(R_1+R_2)^3[\sin\theta(g_{\mathrm{C}}\cos\phi+h_{\mathrm{C}}\sin\phi)+i_{\mathrm{C}}\cos\theta]\\
&\quad+16(R_1+R_2)^3\begin{bmatrix}\sin^3\theta\big(j_{\mathrm{C}}\cos^3\phi+k_{\mathrm{C}}\sin^3\phi+l_{\mathrm{C}}\sin\phi\cos^2\phi+m_{\mathrm{C}}\sin^2\phi\cos\phi\big)\\ +n_{\mathrm{C}}\cos^3\theta+\sin^2\theta\cos\theta\big(o_{\mathrm{C}}\cos^2\phi+p_{\mathrm{C}}\sin^2\phi+q_{\mathrm{C}}\cos\phi\sin\phi\big)\\ +\sin\theta\cos^2\theta(r_{\mathrm{C}}\cos\phi+s_{\mathrm{C}}\sin\phi)\end{bmatrix},
\end{aligned}\tag{29}$$

$$a_{\mathrm{C}}=a_{\mathrm{B}}+\sum_{i\in K_1,j\in K_2}\left(x_i^2+x_j^2\right)\frac{r_{ij}^2}{(R_1+R_2)^2},\quad b_{\mathrm{C}}=b_{\mathrm{B}}+\sum_{i\in K_1,j\in K_2}\left(y_i^2+y_j^2\right)\frac{r_{ij}^2}{(R_1+R_2)^2},$$

$$c_{\mathrm{C}}=c_{\mathrm{B}}+2\sum_{i\in K_1,j\in K_2}\left(x_iy_i+x_jy_j\right)\frac{r_{ij}^2}{(R_1+R_2)^2},\quad d_{\mathrm{C}}=d_{\mathrm{B}}+\sum_{i\in K_1,j\in K_2}\left(z_i^2+z_j^2\right)\frac{r_{ij}^2}{(R_1+R_2)^2},$$

$$e_{\mathrm{C}}=e_{\mathrm{B}}+\sum_{i\in K_1,j\in K_2}\left(x_iz_i+x_jz_j\right)\frac{r_{ij}^2}{(R_1+R_2)^2},\quad f_{\mathrm{C}}=f_{\mathrm{B}}+\sum_{i\in K_1,j\in K_2}\left(y_iz_i+y_jz_j\right)\frac{r_{ij}^2}{(R_1+R_2)^2},$$

$$g_{\mathrm{C}}g_{\mathrm{B}}+\sum_{i\in K_1,j\in K_2}\left(x_i^3+y_i^2x_i+z_i^2x_i-x_j^3-y_j^2x_j-z_j^2x_j\right)\frac{1}{2}\frac{r_{ij}^2}{(R_1+R_2)^2},$$

$$h_{\mathrm{C}}=h_{\mathrm{B}}+\sum_{i\in K_1,j\in K_2}\left(x_i^2y_i+y_i^3+z_i^2y_i-x_j^2y_j-y_j^3z_j^2y_j\right)\frac{1}{2}\frac{r_{ij}^2}{(R_1+R_2)^2},\tag{30}$$

$$i_{\mathrm{C}}=i_{\mathrm{B}}+\sum_{i\in K_1,j\in K_2}\left(x_i^2z_i+y_i^2z_i+z_i^3-x_j^2z_j-y_j^2z_j-z_j^3\right)\frac{1}{2}\frac{r_{ij}^2}{(R_1+R_2)^2},$$

$$j_{\mathrm{C}}=\sum_{i\in K_1,j\in K_2}\left(x_i^3-x_j^3\right),\quad k_{\mathrm{C}}=\sum_{i\in K_1,j\in K_2}\left(y_i^3-y_j^3\right),\quad l_{\mathrm{C}}=3\sum_{i\in K_1,j\in K_2}\left(x_i^2y_i-x_j^2y_j\right),$$

$$m_{\mathrm{C}}=3\sum_{i\in K_1,j\in K_2}\left(y_i^2x_i-y_j^2x_j\right),\quad n_{\mathrm{C}}=\sum_{i\in K_1,j\in K_2}\left(z_i^3-z_j^3\right),\quad o_{\mathrm{C}}=3\sum_{i\in K_1,j\in K_2}\left(x_i^2z_i-x_j^2z_j\right),$$

$$p_{\mathrm{C}}=3\sum_{i\in K_1,j\in K_2}\left(y_i^2z_i-y_j^2z_j\right),\quad q_{\mathrm{C}}=6\sum_{i\in K_1,j\in K_2}\left(x_iy_iz_i-x_jy_jz_j\right),\quad r_C=3\sum_{i\in K_1,j\in K_2}\left(z_i^2x_i-z_j^2x_j\right),$$

$$s_{\mathrm{C}}=3\sum_{i\in K_1,j\in K_2}\left(z_i^2y_i-z_j^2y_j\right),$$

$$\begin{aligned}
\frac{\partial}{\partial\phi}C(\theta,\phi) &= 24(R_1+R_2)^4\begin{bmatrix}\sin^2\theta\big(-2a_{\mathrm{C}}\cos\phi\sin\phi+2b_{\mathrm{C}}\sin\phi\cos\phi+c_{\mathrm{C}}\big(-\sin^2\phi+\cos^2\phi\big)\big)\\ +\sin\theta\cos\theta(-e_{\mathrm{C}}\sin\phi+f_{\mathrm{C}}\cos\phi)\end{bmatrix}\\
&\quad+24(R_1+R_2)^3[\sin\theta(-g_{\mathrm{C}}\sin\phi+h_{\mathrm{C}}\cos\phi)]\\
&\quad+16(R_1+R_2)^3\begin{bmatrix}\sin^3\theta\begin{pmatrix}-3j_{\mathrm{C}}\cos^2\phi\sin\phi+3k_{\mathrm{C}}\sin^2\phi\cos\phi\\ +l_{\mathrm{C}}\big(\cos^3\phi-2\sin^2\phi\cos\phi\big)+m_{\mathrm{C}}\big(-\sin^3\phi+2\sin\phi\cos^2\phi\big)\end{pmatrix}\\ +\sin^2\theta\cos\theta\begin{pmatrix}-2o_{\mathrm{C}}\cos\phi\sin\phi+2p_{\mathrm{C}}\sin\phi\cos\phi\\ +q_{\mathrm{C}}\big(-\sin^2\phi+\cos^2\phi\big)\end{pmatrix}\\ +\sin\theta\cos^2\theta(-r_{\mathrm{C}}\sin\phi+s_{\mathrm{C}}\cos\phi)\end{bmatrix},
\end{aligned}$$

$$\frac{\partial}{\partial\theta}C(\theta,\phi) = 24(R_1+R_2)^4\begin{bmatrix} 2\sin\theta\cos\theta\left(a_C\cos^2\phi + b_C\sin^2\phi + c_C\sin\phi\cos\phi\right) - 2d_C\cos\theta\sin\theta \\ +\left(-\sin^2\theta+\cos^2\theta\right)\left(e_C\cos\phi + f_C\sin\phi\right)\end{bmatrix}$$
$$+ 24(R_1+R_2)^3\left[\cos\theta\left(g_C\cos\phi + h_C\sin\phi\right) - i_C\sin\theta\right]$$
$$+ 16(R_1+R_2)^3\begin{bmatrix} 3\sin^2\theta\cos\theta\left(j_C\cos^3\phi + k_C\sin^3\phi + l_C\sin\phi\cos^2\phi + m_C\sin^2\phi\cos\phi\right) \\ -3n_C\cos^2\theta\sin\theta \\ +\left(\sin^3\theta + 2\sin\theta\cos^2\theta\right)\left(o_C\cos^2\phi + p_C\sin^2\phi + q_C\cos\phi\sin\phi\right) \\ +\left(\cos^3\theta - 2\sin^2\theta\cos\theta\right)\left(r_C\cos\phi + s_C\sin\phi\right)\end{bmatrix}. \quad (31)$$

References

Arfken GB, Weber HJ (1995) Mathematical methods for physicists, 4th edn. Academic, San Diego

Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. Biochemistry 39:5355–5365

Baumeister W, Steven AC (2000) Macromolecular electron microscopy in the era of structural genomics. Trends Biochem Sci 25(12):624–631

Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci 12(1):1–16

Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54(Pt 5):905–921

Brüschweiler R (2003) New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. Curr Opin Struct Biol 13(2):175–183

Chou JJ, Li S, Bax A (2000) Study of conformational rearrangement and refinement of structural homology models by the use of heteronuclear dipolar couplings. J Biomol NMR 18:217–227

Chou JJ, Li S, Klee CB, Bax A (2001) Solution structure of Ca(2+)-calmodulin reveals flexible hand-like properties of its domains. Nat Struct Biol 8:990–997

Clore GM (2000) Accurate and rapid docking of protein–protein complexes on the basis of intermolecular nuclear overhauser enhancement data and dipolar couplings by rigid body minimization. Proc Natl Acad Sci USA 97:9021–9025

Clore GM, Schwieters CD (2003) Docking of protein–protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from 1H/15N chemical shift mapping and backbone 15N–1H residual dipolar couplings using conjoined rigid body/torsion angle dynamics. J Am Chem Soc 125:2902–2912

Debye P (1915) Zerstreuung von Röntgenstrahlen. Ann Phys (Leipzig) 28:809–823

Dobrodumov A, Gronenborn AM (2003) Filtering and selection of structural models: combining docking and NMR. Proteins 53:18–32

Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125:1731–1737

Dosset P, Hus JC, Blackledge M, Marion D (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. J Biomol NMR 16(1):23–28

Dosset P, Hus JC, Marion D, Blackledge M (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. J Biomol NMR 20:223–231

Feigin LA, Svergun DI (1987) Structure analysis by small-angle X-ray and neutron scattering. Plenum Press, New York, GW Taylor (ed)

Gardner KH, Kay LE (1998) The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. Annu Rev Biophys Biomol Struct 27:357–406

Gavin AC, Superti-Furga G (2003) Protein complexes and proteome organization from yeast to man. Curr Opin Chem Biol 7:21–27

Goldstein H (1977) Classical mechanics. Addison-Wesley. Reading, MA

Guillet V, Lapthorn A, Hartley RW, Mauguen Y (1993) Recognition between a bacterial ribonuclease, barnase, and its natural inhibitor, barstar. Structure 1(3):165–176

Guinier A, Fournet G (1955) Small angle scattering of X-rays. Wiley, New York

Jain NU, Wyckoff TJ, Raetz CR, Prestegard JH (2004) Rapid analysis of large protein–protein complexes using NMR-derived orientational constraints: the 95 kDa complex of LpxA with acyl carrier protein. J Mol Biol 343:1379–1389

Koch MH, Vachette P, Svergun DI (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q Rev Biophys 36(2):147–227

Linge JP, Habeck M, Rieping W, Nilges M (2003) ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics 19:315–316

Lipsitz RS, Tjandra N (2004) Residual dipolar couplings in NMR structure analysis. Annu Rev Biophys Biomol Struct 33:387–413

Mackereth CD, Simon B, Sattler M (2005) Extending the size of protein–RNA complexes studied by nuclear magnetic resonance spectroscopy. Chembiochem 6:1578–1584

Mattinen ML, Paakkonen K, Ikonen T, Craven J, Drakenberg T, Serimaa R, Waltho J, Annila A (2002) Quaternary structure built from subunits combining NMR and small-angle X-ray scattering data. Biophys J 83(2):1177–1183

McCoy MA, Wyss DF (2002) Structures of protein–protein complexes are docked using only NMR restraints from residual dipolar coupling and chemical shift perturbations. J Am Chem Soc 124:2104–2105

Prestegard JH, al-Hashimi HM, Tolman JR (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. Q Rev Biophys 33(4):371–424

Rossmann MG, Morais MC, Leiman PG, Zhang W (2005) Combining X-ray crystallography and electron microscopy. Structure 13:355–362

Sali A, Chiu W (2005) Macromolecular assemblies highlighted. Structure 13:339–341

Sibille N, Pardi A, Simorre JP, Blackledge M (2001). Refinement of local and long-range structural order in theophylline-binding RNA using (13)C-(1)H residual dipolar couplings and restrained molecular dynamics. J Am Chem Soc 123:12135–12146

Skrynnikov NR, Goto NK, Yang D, Choy WY, Tolman JR, Mueller GA, Kay LE (2000). Orienting domains in proteins using dipolar couplings measured by liquid-state NMR:

differences in solution and crystal forms of maltodextrin binding protein loaded with beta-cyclodextrin. J Mol Biol 295:1265–1273

Svergun DI, Barberato C, Koch MHJ (1995). CRYSOL—a program to evaluate X-ray solution scattering from biological macromolecules from atomic coordinates. J Appl Cryst 28:768–773

Timmins PA, Zaccai G (1988). Low resolution structures of biological complexes studied by neutron scattering. Eur Biophys J 15(5):257–268

Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A (1997a). Use of dipolar 1H–15N and 1H–13C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat Struct Biol 4:732–739

Tjandra N, Garrett DS, Gronenborn AM, Bax A, Clore GM (1997b) Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. Nat Struct Biol 4:443–449

Zuiderweg ER (2002). Mapping protein–protein interactions in solution by NMR spectroscopy. Biochemistry 41(1):1–7