ARTICLE

# A structure refinement protocol combining NMR residual dipolar couplings and small angle scattering restraints

**F. Gabel · B. Simon · M. Nilges · M. Petoukhov ·
D. Svergun · M. Sattler**

**Abstract** We present the implementation of a target function based on Small Angle Scattering data (Gabel et al. Eur Biophys J 35(4):313–327, 2006) into the Crystallography and NMR Systems (CNS) and demonstrate its utility in NMR structure calculations by simultaneous application of small angle scattering (SAS) and residual dipolar coupling (RDC) restraints. The efficiency and stability of the approach are demonstrated by reconstructing the structure of a two domain region of the 31 kDa nuclear export factor TAP (TIP-associated protein). Starting with the high resolution X-ray structures of the two individual TAP domains, the translational and orientational domain arrangement is refined simultaneously. We tested the stability of the protocol against variations of the SAS target parameters and the number of RDCs and their uncertainties. The activation of SAS restraints results in an improved translational clustering of the domain positions and lifts part of the fourfold degeneracy of their orientations (associated with a single alignment tensor). The resulting ensemble of structures reflects the conformational space that is consistent with the experimental SAS and RDC data. The SAS target function is computationally very efficient. SAS restraints can be activated at different levels of precision and only a limited SAS angular range is required. When combined with additional data from chemical shift perturbation, paramagnetic relaxation enhancement or mutational analysis the SAS refinement is an efficient approach for defining the topology of multi-domain and/or multimeric biomolecular complexes in solution based on available high resolution structures (NMR or X-ray) of the individual domains.

F. Gabel · B. Simon · M. Sattler
Structural and Computational Biology Unit, EMBL,
Meyerhofstrasse 1, 69117 Heidelberg, Germany

*Present Address:*
F. Gabel
Institut de Biologie Structurale, IBS Jean-Pierre Ebel, CEA-CNRS-UJF, 41 rue Jules Horowitz, 38027 Grenoble Cedex, France

M. Nilges
Unité de Bioinformatique Structurale, Institut Pasteur,
25-28 rue du Dr Roux, 75724 Paris, France

M. Petoukhov · D. Svergun
EMBL, Hamburg Outstation, Notkestrasse 85, 22603 Hamburg, Germany

M. Petoukhov · D. Svergun
Institute of Crystallography, Russian Academy of Sciences,
Leninsky pr. 59, 117333 Moscow, Russia

M. Sattler (✉)
Institute of Structural Biology, Helmholtz Zentrum München,
Ingolstädter Landstr. 1, 85764 Neuherberg, Germany
e-mail: sattler@helmholtz-muenchen.de

M. Sattler
Munich Center for Integrated Protein Science, Department Chemie, Technische Universität München, Lichtenbergstr. 4, 85747 Garching, Germany

**Abbreviations**

| | |
|---|---|
| ARIA | Ambiguous restraints for iterative assignment |
| CNS | Crystallography and NMR systems |

| LRR | Leucine rich repeat |
| MD | Molecular dynamics |
| NCS | Non-crystallography symmetry |
| NMR | Nuclear magnetic resonance |
| NOE | Nuclear Overhauser effect |
| NSD | Normalized spatial discrepancy |
| PDB | Protein data bank |
| PRE | Paramagnetic relaxation enhancement |
| RDC | Residual dipolar couplings |
| RMSD | Root mean square displacement |
| RRM | RNA recognition motif |
| SA | Simulated annealing |
| SAS | Small angle scattering |
| SANS | Small angle neutron scattering |
| SAXS | Small angle X-ray scattering |
| TAP | TIP-associated protein |

## Introduction

Both residual dipolar couplings (RDCs) in high-resolution nuclear magnetic resonance (NMR) spectroscopy and small angle scattering (SAS) have experienced tremendous progress in recent years as powerful tools for biomolecular structure determination in solution (Prestegard et al. 2000; Bax 2003; Blackledge 2005; Svergun and Koch 2002; Koch et al. 2003; Neylon 2008). Whereas SAS data contain translational information of biomacromolecules such as radii of gyration, inter-domain distances and arrangements, RDCs provide the respective domain orientations in multimeric complexes as well as domain-internal structural information. NMR can provide additional information about binding interfaces and conformational dynamics. Therefore, both techniques complement each other perfectly and have been applied to challenging problems in structural biology (Aliprandi et al. 2008; Bernado et al. 2005, 2007; Choy et al. 2002; Goult et al. 2007; Grishaev et al. 2008; Marino et al. 2006; Tidow et al. 2007). Especially the high potential of the combined use of SAS and RDCs for structural analysis in solution has been realized recently (Mattinen et al. 2002; Yuzawa et al. 2004; Bernado et al. 2005; Grishaev et al. 2005; Gabel et al. 2006; Marino et al. 2006; Mareuil et al. 2007).

Here, we present the implementation of a recently introduced target function, combining both NMR RDC and SAS restraints (Gabel et al. 2006), in CNS (Crystallography and NMR Systems; Brünger et al. 1998). In contrast to earlier work (Gabel et al. 2006), where the domains were kept artificially at fixed orientations and the target function was calculated directly from the atomic positions, we have now incorporated the target function into a molecular dynamics/simulated annealing protocol in CNS. The protocol is tested

by retrieving the structure of the two-domain, 31 kDa nuclear export protein TAP (Liker et al. 2000) from a set of simulated RDC and SAS data, after initial randomization of the domain positions and orientations. We find a family of structures that cluster both translationally and orientationally and that match the overall SAS curve very well. The effect of the SAS parameters and of the number and errors of RDCs on the quality of the refined structures is investigated. The main advantages of our approach are the short computing times per MD (molecular dynamics) step, the limited angular range of SAS data needed and the flexibility to incorporate the SAS target potential at different levels of accuracy in the annealing protocol. We compare our results with an alternative refinement procedure using the rigid body modeling program SASREF (Petoukhov and Svergun 2005).

## Material and methods

Simulation of RDCs

Three sets of RDCs were simulated with (a) no noise and (b) 1 Hz and (c) 2 Hz noise for all N–H bonds of TAP (residues 120–355, PDB entry 1FT8_A) using the program DC (Delaglio et al. 1995) for an alignment tensor magnitude $D_a = 7$ Hz and rhombicity $r = 0.24$. More details are provided in the Supplementary Material.

Calculation of the SAS target curve and parameters

The small angle scattering target curve of TAP was calculated with the program CRYSOL (Svergun et al. 1995) from the PDB entry 1FT8_A (residues 120–194 and 207–355) in the s-range 0.00…0.35 Å$^{-1}$ (Fig. 1). The simulated
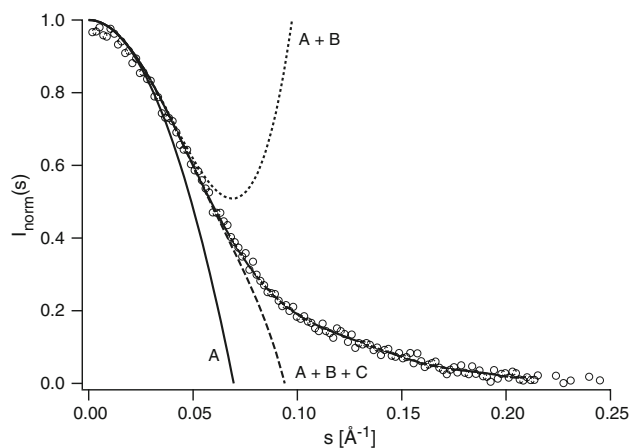


**Fig. 1** Calculated SAS data points and polynomial fit (thick continuous line) with a polynomial fit (Gabel et al. 2006) up to the order of $s^{18}$. The truncated expansions to the order of $s^2$ (A), $s^4$ (A + B) and $s^6$ (A + B + C) are also shown for a comparison

SAS curve was fitted with a polynomial function (Gabel et al. 2006) in order to extract the parameters $A_{\text{target}}$, $B_{\text{target}}$ and $C_{\text{target}}$ used in the SAS target potential (Eq. 1). In addition, the target parameters were calculated directly from the atomic positions (for details see Supplementary Material). The respective values are compared in Supplementary Table 1, along with optimum values obtained in a grid-search (see next section). For the alternative refinement with SASREF, an additional scattering curve was calculated with randomized 5% errors (Supplementary Fig. 3).

### Optimization of SAS target potential parameters by a grid-search

While the SAS target values $A_{\text{target}}$, $B_{\text{target}}$ and $C_{\text{target}}$ used in Eq. 1 are extracted from a polynomial fit of a CRYSOL-generated scattering curve from the TAP crystal structure (including a hydration shell), the parameters $A$, $B$ and $C$ are directly calculated from the atomic positions (Supplementary Eq. 2) of the structures during the refinement process and thus ignore a hydration shell. Therefore, a grid-search of the target parameters $A_{\text{target}}$, $B_{\text{target}}$ and $C_{\text{target}}$ around the theoretically predicted ones is necessary to consider the hydration shell in modified target parameters (Supplementary Table 2). The optimum values were obtained as those that gave the best $\chi^2$ fit of all calculated structures in a run with the simulated SAS curve.

### CNS protocol

We incorporated the SAS target potential (Gabel et al. 2006)

$$\Psi = \left(\frac{\lambda_a}{6}\right)(A - A_{\text{target}})^2 + \left(\frac{\lambda_b}{120}\right)(B - B_{\text{target}})^2 + \left(\frac{\lambda_c}{5040}\right)(C - C_{\text{target}})^2 \quad (1)$$

into the existing CNS program suite (Brünger et al. 1998) by adding several FORTRAN modules and modifying existing CNS protocols. A schematic overview of the individual steps of the CNS protocol is given in Fig. 2.

It is noteworthy that the target potential $A_{\text{target}}$ (i.e. excluding the higher order terms) may be incorporated into the CNS refinement protocol in a simplified way as distance restraints ("pseudo-NOE restraints") between rigid domains (see Supplementary Material).

### Structure calculations

To test the refinement protocol, the following combinations of SAS and RDC potentials were used: (a) No SAS target potential active, all 217N–H RDC restraints active, (b) activation of the target parameter $A_{\text{target}}$ and all RDC restraints, (c) activation of the target parameters $A_{\text{target}}$ and $B_{\text{target}}$ and all RDC restraints, (d) activation of the target parameters $A_{\text{target}}$, $B_{\text{target}}$ and $C_{\text{target}}$ and all RDC restraints, (e) same setup as in (b–d) but all RDCs with 1 or 2 Hz experimental noise, and control runs with no SAS and no RDC restraints active (Table 1 and Supplementary Table 2). In each case, a total of 500 structures were calculated. The 50 structures with the highest CNS energies were discarded from the subsequent structural analysis since most of them had obvious steric problems.

### Selection of refined structures by scoring against the full SAS curve

The SAS scattering curves of the refined structures were calculated with CRYSOL and scored against the target SAS curve (for details see Supplementary Material). The refined structures were evaluated in terms of the following parameters (Table 1 and Supplementary Table 2): Radius of gyration, root mean square displacement (RMSDs) of the RRM domain around its mean position and with respect to the crystal (target) RRM domain relative to the LRR domain, $\chi^2$ of the fit with the target SAS curve, mean RDC $Q$-value (Cornilescu et al. 1998) of the ten best refined structures and the percentage of clustered structures (a cluster being defined by structures with a maximum orientational deviation of $\pm15°$ in a molecule-fixed coordination frame).

### Alternative refinement with SASREF

We also compared the results of our approach to an alternative rigid body refinement using the program SASREF (Petoukhov and Svergun 2005). Details are provided in the Supplementary Material.

## Results

### The CNS refinement protocol

The flowchart in Fig. 2 shows the implementation of the SAS potential into the CNS program suite. The modules are available upon request and will be provided in an extended version of CNS/ARIA.

The SAS potential is only activated during the second cooling step *sa_l_cool2.cns*. This late activation ensures that the domains are already oriented properly by the RDC data and that the SAS potential terms $B$ and $C$ (that depend explicitly on the domain orientation; Gabel et al. 2006) act correctly. The exact time point of activation can be specified by the logical parameter $i_cool/$ncycle that

**Fig. 2** Schematic, sequential flowchart of the CNS protocol used
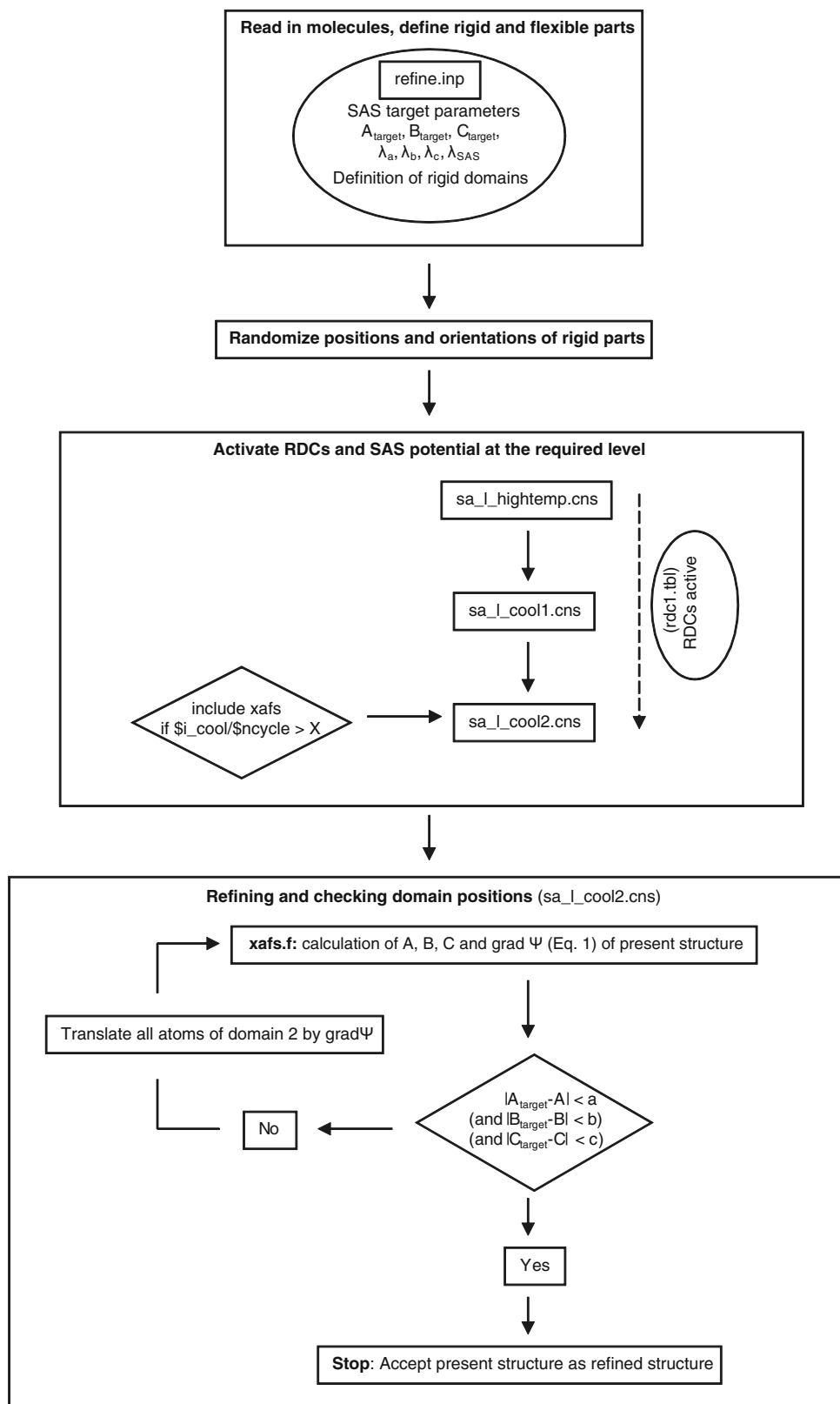


represents the remaining fraction of molecular dynamics steps within *sa_l_cool2.cns*. In order to save computing time, we chose X = $i_cool/$ncycle = 0.9, *i.e.* the SAS

restraints were only active during the last 10% of the *sa_l_cool2.cns* protocol (=2000 MD steps). The target parameters $A_{target}$, $B_{target}$ and $C_{target}$ (=$a_{xaf}$, $b_{xaf}$, $c_{xaf}$) and

**Table 1** Structural parameters of the refined structures as a function of the activation degree of the SAS-potential

| Data set # | SAXS potential | Interdomain distance RRM-LRR (Å, all structures) | RMSD RRM to mean (Å, structures with correct orientation) | RMSD RRM to target (Å, structures with correct orientation) | $\chi^2$ (structures with correct orientation) |
|---|---|---|---|---|---|
| 1 | No SAS | $38.63 \pm 6.68$ | $19.27 \pm 7.93$ | $21.63 \pm 9.29$ | $0.859 \pm 0.346$ |
| 4 | $A = 1180$ | $41.63 \pm 0.46$ | $18.75 \pm 8.99$ | $22.32 \pm 10.85$ | $0.546 \pm 0.045$ |
| 9 | $A = 1180$, $B = 2.8 \times 10^6$ | $42.02 \pm 0.36$ | $11.48 \pm 6.26$ | $20.93 \pm 7.41$ | $0.539 \pm 0.013$ |
| 13 | $A = 1180$, $B = 2.8 \times 10^6$, $C = 8.5 \times 10^9$ | $41.96 \pm 0.38$ | $12.63 \pm 5.20$ | $19.12 \pm 8.52$ | $0.530 \pm 0.011$ |
| – | Target crystal structure | $41.22$ | – | – | $0.516$ |

The optimum target parameters shown here were determined by the grid-search method (Supplementary Table 2)

their individual weighing factors (Eq. 1) $\lambda_a$, $\lambda_b$ and $\lambda_c$ (=$w_{axf}$, $w_{bxf}$, $w_{cxf}$, see Supplement) are provided by the user, along with the overall weighing factor $\lambda_{SAS}$ (=$w_{xaf}$) for the energy of the SAS potential and the definition of the rigid domains. A typical example is given in the Supplementary Material.

The FORTRAN77 file *xafs.f* is invoked during the *sa_l_cool2.cns* protocol in order to calculate the SAS parameters *A*, *B* and *C* of the structures during the refinement protocol. This calculation is done from the atomic coordinates as described (Gabel et al. 2006). *xafs.f* calculates also the potential Ψ (Eq. 1) and its gradient according to Gabel et al. (2006). If the parameters *A*, *B* and *C* of the present structure differ from their respective target value by more than specific percentage a, b and c (defined in the *xafs.f* file), all atoms of domain 2 (RRM in our case) will be translationally shifted by grad Ψ with the partial derivatives $\frac{\partial \Psi}{\partial \theta}$ and $\frac{\partial \Psi}{\partial \varphi}$ multiplied with a weighing factor $k = 10^{-7}$. In the present protocol we chose a = 0.03, b = 0.05 and c = 0.05. The optimized value for *k* is defined in *xafs.f*. If the parameters *A*, *B* and *C* match their target values (within the marges provided) the present structure will be considered as the final, SAS-refined structure and the protocol will stop.

Optimization of SAS target potential values by a grid-search

The SAS parameters *A*, *B* and *C* (Eq. 1) are approximate and do not consider the hydration shell of the protein. Furthermore, the finite polynomial fit and experimental errors of the scattering curve lead to uncertainties in the determination of $A_{target}$, $B_{target}$ and $C_{target}$ (discussed in Gabel et al. 2006). Therefore, a grid-search around their initial values (determined by a $\chi^2$-fit against the SAS curve) is necessary. The optimum values for $A_{target}$, $B_{target}$ and $C_{target}$ were obtained as those that gave the best $\chi^2$ fit of all calculated structures in a run with the simulated SAS curve (Materials and Methods and Supplementary Table 2). The

optimal $A_{target} = 1180$ found by a grid-search is within 5% of its predicted value. $B_{target} = 2.8 \times 10^6$ and $C_{target} = 8.5 \times 10^9$ are within 10% and 25% of their predicted values, respectively (Supplementary Table 1).

The A-potential lifts part of the orientational degeneracy

The activation of the *A*-term of the SAS-potential in addition to RDC restraints improved the structural parameters of the 20 best structures ($\chi^2$-scored against the SAS-curve) in several regards (Fig. 3). The translational spread of the RRM centres of mass is decreased and its rotational degeneracy is partially lifted by reducing the number of orientations present from three to two (more details are provided in the Supplementary Material).

Structural parameters as a function of the activation level of the SAS-potential

The effect of the progressive activation of the SAS-potential terms on all structures (with correct domain orientation; no $\chi^2$-scoring applied) is best summarized in Fig. 4, showing the spatial distribution of the TAP alignment tensor frames (placed on the RRM domain centre of mass after superposition of the LRR domain).

In the absence of a SAS-potential, all RRM domains are properly oriented and aligned by the RDCs but distributed randomly within a sphere that is limited by the steric restraints due to the finite length of the linker connecting both domains. The ample translational distribution of the RRM domains is reflected in large RMSDs of the vector connecting both domains and a poor mean $\chi^2$-fit of all structures with the SAS curve (Table 1, data set 1).

The activation of the *A*-term induces a significant decrease of the translational degrees of freedom of the structures in the sense that all RRM domain centres (while being correctly oriented by the RDCs) are now placed at
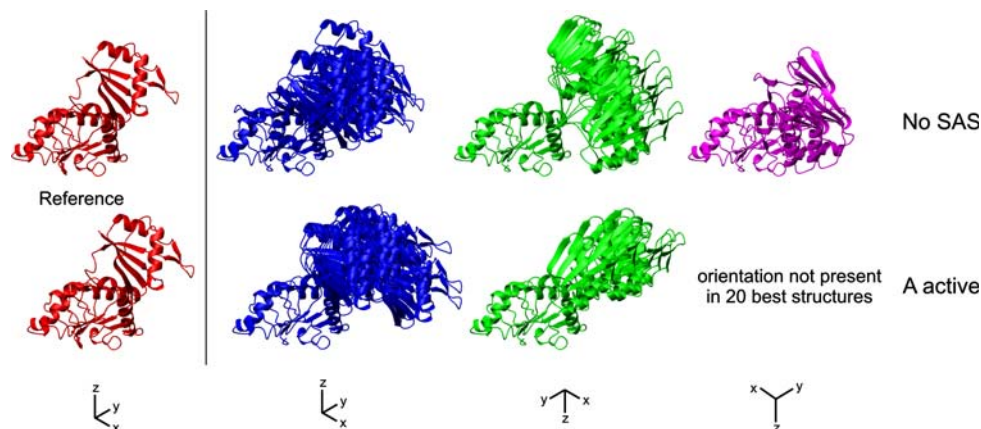
**Fig. 3** 20 best refined structures (in terms of $\chi^2$). Top: no SAS term active, RDCs active (dataset 1 in Table 1), bottom: $A_{target}$ active, RDCs active (dataset 4). In the bottom part, three structures out of the 20 best have been omitted since they did not belong to any of the four possible clusters. The structures in the domain orientation of the reference structure are depicted in blue, the crystal structure itself in red. For all structures the LRR domain is superimposed and shown in the same orientation. Structures in green and magenta belong to clusters other than the crystal structure orientation, where the RRM domain is rotated by 180° compared to the reference structure orientation. The fourth possible orientation is not found, presumably due to steric constraints introduced by the linker of finite length
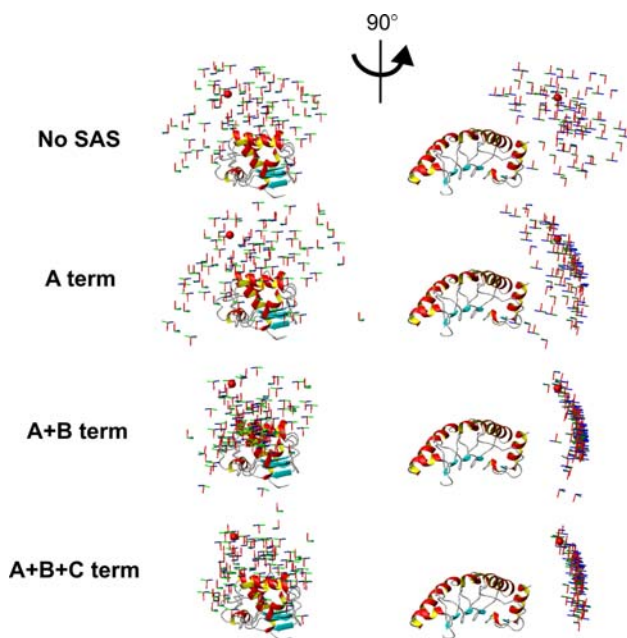


**Fig. 4** Alignment tensor reference frames of all refined TAP structures (with correct domain orientation) as a function of the activation level of the SAS-potential from two different views. LRR domains have been superposed and the alignment tensor frames are placed at the centres of mass of the RRM domains. The centre of mass of the target (crystal) RRM domain is depicted as a small red sphere. The structures on the right-hand side have been generated by a rotation of 90° around the vertical axis

equal distance from the LRR centre of mass on a two-dimensional (spherical) surface. However, since no driving force orthogonal to the domain-connecting vector is active, the refined RRM positions continue to be widely spread over this spherical surface, comparable to the case without SAS restraints. This is reflected by large RRM RMSDs to

the mean refined structure and to the target structure (Table 1, data set 4) comparable to the ones without SAS restraints. However, the $\chi^2$-values of all structures are significantly improved, due to the adjustment of the radius of gyration induced by the activation of the $A$-potential. As in the case without SAS-potentials, the target position of the RRM domain is situated within the set of structures refined.

The additional activation of the $B$-term, which induces a driving force on the RRM domain tangentially to the spherical surface defined by the $A$-term, reduces mainly the translational spread of the RRM domain positions and their RMSDs to the mean refined structure but improves the RMSDs to the target structure only slightly (Table 1, data set 9). The $\chi^2$-values of all structures decrease further and their standard deviation is less pronounced with respect to the case where only $A$ is active. All structures remain correctly oriented.

Finally, the full activation of the SAS-potential ($A + B + C$) decreases further the degrees of freedom of the RRM domain positions on the spherical surface into a crescent moon shaped region and reduces the mean $\chi^2$-values (Table 1, data set 13). All structures match the scattering curve very well and the positions of the RRM alignment tensor frames can be considered as a geometric representation of a set of structures that are all compatible with the scattering curve (in terms of a low $\chi^2$-fit). The RMSDs of the RRM to the mean structure increases slightly with respect to the value in the case of $A + B$, the RMSDs to the target structure are comparable to the preceeding cases. Note, however, that the coordinate RMSD is not a good measure for characterizing the improvement of a spherical distribution of structures into a disk- or line-like distribution as is achieved by the activation of the SAS potential (Fig. 4).
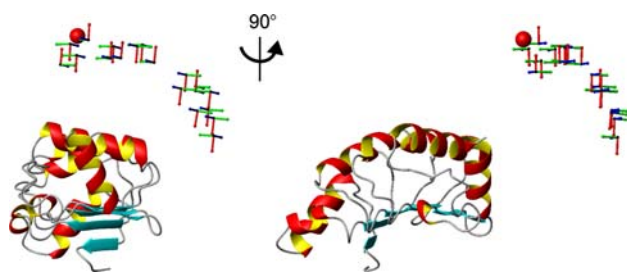
Figure 5 illustrates how an additional $\chi^2$-scoring (against the complete scattering curve) selects a subset of structures out of those already refined against the $A + B + C$ potential: the translational spread of these 20 best structures is limited to a core region of the crescent moon shaped surface but a unique, unambiguous structure cannot be pointed out even in this case. Note, that all structures shown in Fig. 5 have virtually indistinguishable $\chi^2$-values (=0.515…0.518). Figure 5 can therefore be considered as a geometrical representation of an ensemble of TAP structures equivalent from a SAS point of view (within the error bars of the SAS curve).

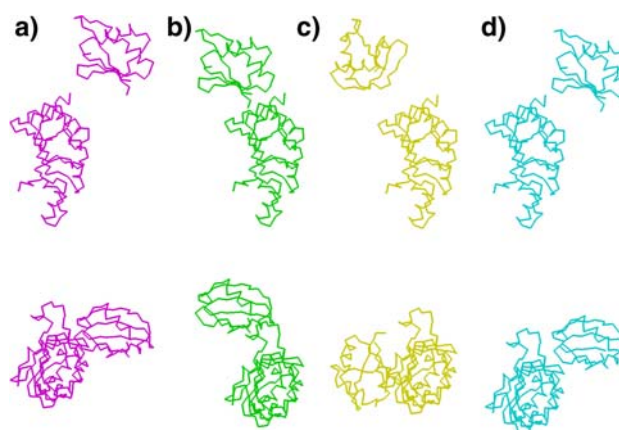### Effect of errors on the refined structures

The introduction of experimental noise (1 and 2 Hz) on the RDC data sets did not change notably the main structural parameters, i.e. radius of gyration, translational spread etc. (for more details see Supplementary Material). This suggests that the protocol is stable against errors in the experimental RDCs.

### Alternative refinement with SASREF

The CNS-based structure calculation protocol with simultaneous refinement against SAS and RDC data was compared with rigid body modelling using the program SASREF (Petoukhov and Svergun 2005). SASREF allowed the reconstruction of the initial model in some runs but also yielded several other solutions equally compatible with the simulated data (see Supplementary Fig. 3, top). In particular, two "false positives", one with a different position of the RRM relative to the LRR domain (Fig. 6b) and the other with different relative orientation of domains (Fig. 6c) were obtained yielding the same or even better fit quality than the correct rigid body model (Fig. 6d,



**Fig. 6** TAP rigid body models reconstructed by SASREF with RDC restraints. (**a**) reference structure, (**b**) alternative position, (**c**) alternative orientation, (**d**) correct reconstruction. The models are shown with the same orientation for the LRR domain. The bottom view is rotated by 90° about horizontal axis

Supplementary Table 3). Not unexpectedly, although the RMSD between the atomic coordinates of the false positives and the initial model are rather high, the NSD criteria (Kozin and Svergun 2001) were all below 1.0 (Supplementary Table 3). This indicated that the models with "alternative" configuration of domains still kept essentially the same low resolution shape (and thus the same fit to the scattering data). A possible further reason for the observed ambiguity is the influence of the solvation shell. In CRYSOL, used to compute the simulated profile, this shell is calculated as an envelope surrounding the entire TAP particle. SASREF employs the scattering amplitudes calculated by CRYSOL from each individual domain, surrounded by its solvation shell. The influence of the hydration shell was tested by generating a second scattering curve with smaller errors (Supplementary Material).

### Discussion

The activation of an inter-domain distance restraint ($A_{target}$) was sufficient to reduce the possible domain orientations that match the scattering curve well (Fig. 3, bottom). This demonstrates that the joint refinement against RDC and SAS data can lift the four-fold orientational degeneracy associated with a single alignment tensor at least partially. However, as becomes clear from Fig. 3, some degeneracy can remain, which is due to the rather isotropic shape of the RRM domain (effects of domain anistropy are discussed in more detail in Gabel et al. 2006). In order to completely lift the orientational degeneracy it is necessary to include additional data. These could include RDC data from an additional distinct alignment tensor (Prestegard et al. 2000; Bax 2003; Blackledge 2005), or additional translational and distance



**Fig. 5** Alignment tensor reference frames of the 20 best TAP structures ($\chi^2$-scored against the complete scattering curve) out of all $A + B + C$-refined structures (with correct domain orientation) from two different perspectives. LRR domains have been superposed and the alignment tensor frames are placed at the centres of mass of the RRM domains. The centre of mass of the target (crystal) RRM domain is depicted as a small red sphere. The structures on the right-hand side were obtained by a 90° rotation around the vertical axis

information from NOEs (Grishaev et al. 2005), chemical shift perturbations and/or biochemical data (mutational analysis) (Dominguez et al. 2003; Clore and Schwieters 2003), or, for example, paramagnetic relaxation enhancements (Battiste and Wagner 2000; Mackereth et al. 2005). The implementation of the SAS potential in CNS allows that all these different types of data can be readily supplemented for the structure calculation. We also note that in this case a further refinement of the resulting structures considering electrostatic terms in a solvation shell (Linge et al. 2003) is possible and strongly recommended.

There are two general points that limit the accuracy of the target parameters used: firstly, $A$, $B$ and $C$ in Eq. 1 do not consider the presence of a protein solvent shell with different scattering density than the bulk solvent (Svergun et al. 1998). In our case, this effect may well account for a part of the differences between the target values $A_{target}$, $B_{target}$ and $C_{target}$ determined by a fit of the calculated scattering curve (including a hydration shell) and the ones yielding the best results in the MD structural refinement (Supplementary Table 1). Secondly, due to practical reasons, the polynomial fit of the SAS curve is only a truncated form of the correct *sinc* function (Gabel et al. 2006), introducing a certain error in the extracted target parameters.

In contrast to an approach proposed recently (Grishaev et al. 2005), our protocol does not aim to derive relatively high-resolution structural information from the SAS data, but rather provides an efficient rigid body algorithm for combination with complementary NMR data. Therefore, no corrective form factors need to be introduced. Importantly, the determination of the target values $A_{target}$, $B_{target}$ and $C_{target}$ is only required once at the beginning of the MD simulation. As a result our approach is computationally very efficient. For example, a single MD step requires about 1 second of computing time for all SAS potential terms in the case of TAP (235 residues, including hydrogen atoms) on a commercial Pentium computer (3 GHz). Calculation of a single structure, including all SAS terms takes about 1 hour on a single CPU. Presently, a grid search of several target value conditions is necessary so that the effective time per structure calculation is a function of the target parameter conditions explored (see Supplementary Table 2).

An advantage of the protocol presented here is that the user can easily choose and adjust the level and contribution of SAS restraints by selecting the order of SAS terms activated during the simulated annealing. The $A$ potential is related to the radius of gyration which is amongst the most reliable information contained in SAS data and defines the interdomain distance (Gabel et al. 2006). Restraints on the radius of gyration have been used in molecular dynamics calculations for many years (Boczko and Brooks 1995; Hünenberger et al. 1995; Kuszewski et al. 1999). However, the A-term is mathematically distinct from these

since it restrains the distance between the centers-of-mass of two domains. The activation of the B potential (under the precautions mentioned in the Supplementary Material) significantly improves the convergence of structures. If high quality data (good signal/noise, wide angular range) are available an activation of the $C$-term may further refine the geometric distribution of the structures (Fig. 4) even though it has little effect on the $\chi^2$ values of the best structures (Supplementary Table 2).

The coordinate RMSDs between the SAS-refined and the reference structures (Table 1) are rather large. The main reason for this is that, in contrast to other reports (e.g. Grishaev et al. 2005), we have not included any other data than the SAS and RDCs. Especially, no interdomain distance restraints were included, which—if applied—will drastically reduce the coordinate RMSD. In fact, the protocol presented here yields a geometric representation of an ensemble of structures that are virtually equivalent in terms of $\chi^2$ fits with the scattering curve (Figs. 4 and 5), and thus provides a good representation of the conformational space that is consistent with the SAS data. The potential (Eq. 1) allows to represent the accuracy of the refined structures graphically in terms of residual translational degrees of freedom of the domains for a given orientation. Given that the SAS data are in general associated with high structural ambiguities it is important to highlight these uncertainties. Such information is rarely provided by other rigid body modelling approaches.

The results obtained with the rigid-body modelling using SASREF illustrate the utility of the RDC restraints in SAS-based rigid body modelling but also highlight the possibilities of ambiguous reconstructions even in this case. It is important to note that SAS essentially sees the overall particle shape, and if a false positive has a similar shape at low resolution, it may be difficult to discard such solution based on the fit to the "noisy" experimental data (Supplementary Fig. 3, top). This documents the high structural ambiguity of SAS data. It is therefore always necessary to complement SAS-based rigid body models with independent and/or additional structural information.

In the present study we limited the application to a two-domain protein. However, the approach can be easily extended to the structural refinement of two subunits within a multimeric complex with SANS: either by using specific perdeuteration of the two subunits of interest and contrast variation of the solvent ($H_2O$ to $D_2O$ ratio) or by using the natural contrast between protein and RNA/DNA molecules (Timmins and Zaccai 1988).

## Conclusions

We present an efficient structural calculation protocol that simultaneously employs SAS translational and RDC

orientational restraints in CNS to define the quaternary arrangement of a two-domain protein. The addition of SAS restraints notably reduces the translational degrees of freedom of the refined structures and partially lifts the orientational degeneracies associated with RDCs from a single alignment medium. Main advantages of our approach are the computational efficiency and that the target potential can be activated at several levels of precision, allowing structural refinement with SAS data from a very limited angular range. The protocol can be combined with additional structural information, for example from NMR chemical shift perturbation, paramagnetic relaxation enhancements or biochemical data. This is required to reduce remaining translational spread of the resulting structures. The approach is generally applicable to multidomain proteins and/or complexes with known single-domain high resolution structures (X-ray or NMR).

# References

Aliprandi P, Sizun C, Perez J, Mareuil F, Caputo S, Leroy JL, Odaert B, Laalami S, Uzan M, Bontems F (2008) S1 ribosomal protein functions in translation initiation and ribonuclease RegB activation are mediated by similar RNA/protein interactions. An NMR and SAXS analysis. J Biol Chem [Epub ahead of print]

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

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

Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. Proc Natl Acad Sci USA 102(47):17002–17007

Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129:5656–5664

Blackledge M (2005) Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. Prog Nucl Mag Res Spectrosc 46:23–61

Boczko EM, Brooks CL 3rd (1995) First-principles calculation of the folding free energy of a three-helix bundle protein. Science 269(5222):393–396

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 54:905–921

Choy WY, Mulder FA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay JD, Kay LE (2002) Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. J Mol Biol 316:101–112

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

Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 120:6836–6837

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6(3):277–293

Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125(7):1731–1737

Gabel F, Simon B, Sattler M (2006) A target function for quaternary structural refinement from small angle scattering and NMR orientational restraints. Eur Biophys J 35(4):313–327

Goult BT, Rapley JD, Dart C, Kitmitto A, Grossmann JG, Leyland ML, Lian LY (2007) Small-angle X-ray scattering and NMR studies of the conformation of the PDZ region of SAP97 and its interactions with Kir2.1. Biochemistry 46:14117–14128

Grishaev A, Wu J, Trewhella J, Bax A (2005) Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. J Am Chem Soc 127(47):16621–16628

Grishaev A, Tugarinov V, Kay LE, Trewhella J, Bax A (2008) Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. J Biomol NMR 40:95–106

Hünenberger PH, Mark AE, van Gunsteren WF (1995) Computational approaches to study protein unfolding: hen egg white lysozyme as a case study. Proteins 21(3):196–213

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

Kozin MB, Svergun DI (2001) Automated matching of high- and low-resolution structural models. J Appl Crystallogr 34:33–41

Kuszewski J, Gronenborn AM, Clore GM (1999) Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. J Am Chem Soc 121:2337–2338

Liker E, Fernandez E, Izaurralde E, Conti E (2000) The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RRM domain and an LRR domain. EMBO J 19(21):5587–5598

Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M (2003) Refinement of protein structures in explicit solvent. Proteins 50:496–506

Mackereth CD, Simon B, Sattler M (2005) Extending the size of protein-RNA complexes studied by Nuclear Magnetic Resonance Spectroscopy. Chembiochem 6:1578–1584

Mareuil F, Sizun C, Perez J, Schoenauer M, Lallemand JY, Bontems F (2007) A simple genetic algorithm for the optimization of multidomain protein homology models driven by NMR residual dipolar couplings and small angle X-ray scattering data. Eur Biophys J 37(1):95–104

Marino M, Zou P, Svergun D, Garcia P, Edlich C, Simon B, Wilmanns M, Muhle-Goll C, Mayans O (2006) The Ig doublet Z1Z2: a model system for the hybrid analysis of conformational dynamics in Ig tandems from titin. Structure 14(9):1437–1447

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

Neylon C (2008) Small angle neutron and X-ray scattering in structural biology: recent examples from the literature. Eur Biophys J [Epub ahead of print]

Petoukhov MV, Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J 89:1237–1250

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

Svergun DI, Koch MH (2002) Advances in structure analysis using small-angle scattering in solution. Curr Opin Struct Biol 12(5):654–660

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

Svergun DI, Richard S, Koch MH, Sayers Z, Kuprin S, Zaccai G (1998) Protein hydration in solution: experimental observation by X-ray and neutron scattering. Proc Natl Acad Sci USA 95(5):2267–2272

Tidow H, Melero R, Mylonas E, Freund SM, Grossmann JG, Carazo JM, Svergun DI, Valle M, Fersht AR (2007) Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. Proc Natl Acad Sci USA 104:12324–12329

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

Yuzawa S, Ogura K, Horiuchi M, Suzuki NN, Fujioka Y, Kataoka M, Sumimoto F, Inagaki H (2004) Solution structure of the tandem Src homology 3 domains of p47phox in an autoinhibited form. J Biol Chem 279(28):29752–29760