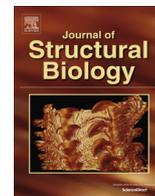




Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: [www.elsevier.com/locate/yjsbi](http://www.elsevier.com/locate/yjsbi)

## Pushing the size limit of *de novo* structure ensemble prediction guided by sparse SDSL-EPR restraints to 200 residues: The monomeric and homodimeric forms of BAX

Axel W. Fischer<sup>a</sup>, Enrica Bordignon<sup>b</sup>, Stephanie Bleicken<sup>c</sup>, Ana J. García-Sáez<sup>c</sup>, Gunnar Jeschke<sup>d</sup>, Jens Meiler<sup>a,\*</sup>

<sup>a</sup> Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, TN 37232, USA

<sup>b</sup> Department of Physics, Freie Universität Berlin, Berlin 14195, Germany

<sup>c</sup> Interfaculty Institute of Biochemistry, Eberhard Karls University Tübingen, Tübingen 72074, Germany

<sup>d</sup> Laboratory of Physical Chemistry, ETH Zürich, Zürich 8093, Switzerland

### ARTICLE INFO

#### Article history:

Received 7 November 2015

Received in revised form 25 April 2016

Accepted 26 April 2016

Available online xxxxx

#### Keywords:

Protein structure prediction

Ab initio

DEER

PELDOR

Protein structure

EPR

### ABSTRACT

Structure determination remains a challenge for many biologically important proteins. In particular, proteins that adopt multiple conformations often evade crystallization in all biologically relevant states. Although computational *de novo* protein folding approaches often sample biologically relevant conformations, the selection of the most accurate model for different functional states remains a formidable challenge, in particular, for proteins with more than about 150 residues. Electron paramagnetic resonance (EPR) spectroscopy can obtain limited structural information for proteins in well-defined biological states and thereby assist in selecting biologically relevant conformations. The present study demonstrates that *de novo* folding methods are able to accurately sample the folds of 192-residue long soluble monomeric Bcl-2-associated X protein (BAX). The tertiary structures of the monomeric and homodimeric forms of BAX were predicted using the primary structure as well as 25 and 11 EPR distance restraints, respectively. The predicted models were subsequently compared to respective NMR/X-ray structures of BAX. EPR restraints improve the protein-size normalized root-mean-square-deviation (RMSD100) of the most accurate models with respect to the NMR/crystal structure from 5.9 Å to 3.9 Å and from 5.7 Å to 3.3 Å, respectively. Additionally, the model discrimination is improved, which is demonstrated by an improvement of the enrichment from 5% to 15% and from 13% to 21%, respectively.

© 2016 Published by Elsevier Inc.

### 1. Introduction

Proteins undergo conformational changes while performing their biological function. Although X-ray crystallography provides snapshots of important conformations, often not all biologically relevant conformations can be crystallized. Nuclear magnetic resonance (NMR) spectroscopy, the premier method to study protein dynamics at atomic detail, suffers from a size limit that complicates a detailed analysis of larger proteins. Electron paramagnetic resonance (EPR) spectroscopy in conjunction with site-directed spin labeling (SDSL) offers an alternative approach to study protein

structure and dynamics. Briefly, typically two cysteine residues are introduced into a *cys*-less variant of the protein and coupled with *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL), which carries an unpaired electron. The dipolar interaction of the two unpaired electrons is inversely proportional to the cubed distance and can be measured with high sensitivity with a pulsed dipolar spectroscopy technique called double electron-electron resonance (DEER) or pulsed electron-electron double resonance (PELDOR) (de Vera et al., 2013; Jeschke, 2012). As for every distance measurement a dedicated protein double mutant needs to be created and tested for functional viability, data obtained from SDSL-EPR measurements are sparse. Thus, such data typically fail to unambiguously determine the structure of a protein at atomic detail. However, it has been demonstrated that in conjunction with *de novo* protein structure prediction algorithms determination of a protein's fold might be within reach (Alexander et al., 2008; Fischer et al., 2015; Hirst

\* Corresponding author at: Vanderbilt University, Departments of Chemistry and Pharmacology, Center for Structural Biology, 465 21st Ave South, BIOSCI/MRBIIL, Room 5144B, Nashville, TN 37232-8725, USA.

E-mail address: [jens.meiler@vanderbilt.edu](mailto:jens.meiler@vanderbilt.edu) (J. Meiler).

URL: <http://www.meilerlab.org> (J. Meiler).

et al., 2011). Whereas previous studies were performed on smaller proteins (Alexander et al., 2008) or mainly based on simulated SDSL-EPR restraints (Fischer et al., 2015), this study evaluates the impact of experimental SDSL-EPR distance restraints on *de novo* protein structure prediction for larger proteins that adopt multiple biologically relevant conformations.

The major challenges of *de novo* protein structure prediction are the vast size of the conformational space that needs to be sampled as well as the discrimination of inaccurate models, *i.e.* the identification of low-energy, biologically relevant states of a protein with a simplified energy function. The simplified macromolecule representations used in *de novo* folding simulations prohibit computation of accurate free energy differences between different conformations. Instead, the approach employed in this study uses knowledge-based energy functions to determine the likelihood of proposed protein models (Woetzel et al., 2012). In parallel, SDSL-EPR distance restraints restrict the sampling space to conformations that are in agreement with the SDSL-EPR data (Bleicken et al., 2014), thus increasing the frequency with which models in agreement with the SDSL-EPR data are sampled. Through incorporation into the scoring function, SDSL-EPR distance restraints also improve the discrimination of inaccurate models. Studying soluble monomeric and homodimeric BAX in this context is especially intriguing due to the large size of the protein and the availability of high-quality experimental SDSL-EPR data sets.

BAX plays a central role in the apoptotic cell death, which is fundamental to the survival of mammals and related to various diseases. Whereas unwanted apoptosis is seen as cause for ischemia and Alzheimer's disease (Bamberger and Landreth, 2002), failure of apoptosis is a key step in developing tumors and autoimmune diseases (Eguchi, 2001; Favaloro et al., 2012; Leber et al., 2010; Strasser et al., 2000). As many different signals for cell death converge on mitochondrial outer membrane (MOM) permeabilization, a better understanding of this mechanism is pivotal for the treatment of diseases related to the apoptotic process (Czabotar et al., 2014). MOM permeabilization is controlled by members of the Bcl-2 family, and the pro-apoptotic protein BAX is described to execute it (Youle and Strasser, 2008). In a healthy cell, BAX is a monomeric, cytosolic protein, whose structure was determined by NMR spectroscopy (Suzuki et al., 2000). Upon pro-apoptotic stimuli, BAX inserts into the MOM, oligomerizes, and creates pores (Czabotar et al., 2014; Youle and Strasser, 2008). Through the pores, cytochrome *c* and other pro-apoptotic proteins are released into the cytosol, initiating a proteolytic cascade leading to cell death. The structure of the membrane-embedded active BAX remains elusive. However, three recent publications have provided valuable new structural insights (Bleicken et al., 2014; Czabotar et al., 2013; Westphal et al., 2014).

Here we apply the BCL::Fold (Karakaş et al., 2012) algorithm to predict the tertiary structure of soluble monomeric BAX and of the dimerization domain of membrane-embedded BAX oligomers. For the solution structure of BAX (Protein Data Bank (PDB) ID 1F16) and the BAX BH3-in-groove dimer (PDB ID 4BDU), high-resolution structures are published (Czabotar et al., 2013; Suzuki et al., 2000) and a number of SDSL-EPR measurements exist (Bleicken et al., 2014). Therefore, this study represents a benchmark test if SDSL-EPR data are sufficient to determine the structure of biologically important states of large, membrane-associated proteins. BCL::Fold is tailored towards assembly of large protein structures from predicted secondary structure elements (SSEs) (Heinze et al., 2015; Karakaş et al., 2012). In a first step, the tertiary structure of soluble monomeric BAX was predicted from twenty-five SDSL-EPR distance restraints (Bleicken et al., 2014), demonstrating the feasibility of the protocol as well as the influence of the limited SDSL-EPR data on *de novo* protein structure prediction. In a second step, the tertiary structure of the dimerization domain of homodimeric BAX ( $\alpha$ -helices 2–5) was predicted from eleven SDSL-EPR distance restraints (Bleicken et al., 2014), demonstrating the applicability of the protocol to oligomeric proteins. In both cases, usage of SDSL-EPR distance restraints significantly improved the accuracy of the sampled models as well as the accuracy with which the models in best agreement with the NMR- and X-ray-derived models could be selected.

imeric BAX ( $\alpha$ -helices 2–5) was predicted from eleven SDSL-EPR distance restraints (Bleicken et al., 2014), demonstrating the applicability of the protocol to oligomeric proteins. In both cases, usage of SDSL-EPR distance restraints significantly improved the accuracy of the sampled models as well as the accuracy with which the models in best agreement with the NMR- and X-ray-derived models could be selected.

## 2. Materials and methods

The tertiary structures of soluble monomeric and homodimeric BAX were predicted using the BCL::Fold (Karakaş et al., 2012) algorithm. A summary of the structure prediction protocol is given in the following section, followed by a section describing how SDSL-EPR distances were translated into structural restraints. The accuracy of the predictions was evaluated by computing a protein-size normalized root-mean-square-deviation of the backbone coordinates (RMSD100, Eq. (2)) (Carugo and Pongor, 2001). Further, we compute the enrichment metric, which quantifies how well the employed scoring function is able to distinguish accurate models from inaccurate models.

### 2.1. Structure prediction protocol

The protocol used to predict the tertiary structure of soluble monomeric BAX and homodimeric BAX is based on the BCL::Fold protocol for soluble proteins (Karakaş et al., 2012). As in the original protocol, a pool containing the secondary structure elements (SSEs) was predicted from the primary structure using the secondary structure prediction algorithms PsiPred (Jones, 1999) and Juko9D (Leman et al., 2013) (Procedure S3). BCL::Fold subsequently uses a Monte Carlo sampling algorithm to assemble the predicted SSEs in the three-dimensional space. BCL::Fold uses the Monte Carlo sampling algorithm in conjunction with the Metropolis criterion (MCM) for energy minimization to search the conformational space for models with a likely overall fold (Procedure S4) (Karakaş et al., 2012). After each Monte Carlo step, models are scored using knowledge-based potentials evaluating different scoring terms like SSE packing, radius of gyration, amino acid exposure, amino acid interactions, loop closure geometry, secondary structure length and content, as well as penalizing potentials for SSE and amino acid clashes (Woetzel et al., 2012). The potential functions for each scoring term were derived from statistics over protein structures deposited in the PDB using the inverse Boltzmann relation (Eq. (1)) (Woetzel et al., 2012).

$$E = -RT \times \ln \frac{P_{obs}}{P_{back}} \quad (1)$$

For each scoring term, the probability of observing a specific feature ( $P_{obs}$ ) was computed from statistics derived from structures deposited in the PDB. This probability is normalized by the probability of observing this feature by chance ( $P_{back}$ ). This normalization ensures that favorable features are assigned negative scores. The term  $RT$  is set to 1 for convenience (Woetzel et al., 2012). For example, one scoring term ( $S_{NC}$ ) evaluates the burial of residues. The degree of burial was quantified using the neighbor count metric (Durham et al., 2009), which assigns a non-negative number – the neighbor count – to each residue. For each amino acid type, statistics over the neighbor count distributions were collected from structures deposited in the PDB. The distributions were binned and the probability of each bin ( $P_{obs}$ ) was computed (Woetzel et al., 2012). After normalization with  $P_{back}$ , Eq. (1) can be used to compute  $S_{NC}$  for each residue in the sampled models. The total score of a protein model – the BCL score – is the weighted sum of the different scoring terms (Woetzel et al., 2012).

Additional scoring terms based on the motion-on-a-cone (CONE) model (Alexander et al., 2008; Hirst et al., 2011) were used to quantify the agreement of the sampled models with the available SDSL-EPR data.

The folding simulation is broken down into five assembly stages. Each stage lasts for a maximum of 2000 MCM steps but is terminated early if a maximum of 400 MCM steps without score improvement in a row is reached. The assembly stages consist of large-scale sampling moves like adding or removing SSEs, flipping and swapping SSEs, as well as large-scale translations and rotations. Over the course of the five assembly stages, the weights for the potentials penalizing SSE and amino acid clashes ramp up to 0, 125, 250, 375, and 500. The weight for scoring the agreement of the model with the SDSL-EPR data remains constant at 50 over all stages. As a result, agreement with SDSL-EPR distance restraints contributes about 45% to the total score, if provided. For previous benchmark studies, various weights for the SDSL-EPR agreement score were evaluated and a weight of 50, which equates to a contribution of 40–50% to the total score, provided the best prediction results (Fischer et al., 2015).

After the assembly stages the model is refined. This process is encapsulated in one stage that consists of small structural perturbations like low-amplitude translations and rotations of SSEs. This stage does not change the overall topology drastically. This stage lasts for a maximum of 2000 MCM steps but is terminated early if a maximum of 400 MCM steps without score improvement in a row are reached. During the refinement stage, the weight for the SDSL-EPR score remains at 50. For homodimeric BAX, the protein structure prediction protocol was slightly altered to assemble and refine the models in C2-symmetry mode (Weiner et al., 2013).

## 2.2. Translating SDSL-EPR distances into structural restraints

Through the DEER/PELDOR experiment, SDSL-EPR spectroscopy measures the distance between two unpaired electrons located in the N-O group of spin labels ( $D_{SL}$ ) that are covalently attached to cysteines in the protein. The DEER experiment consists of microwave pulses at two different frequencies used to measure the dipolar coupling between two electron spins. The pulse sequence at the observer frequency produces an echo. The pulse at the pump frequency flips the coupled spin, thus changing the local field at the observer spin by the dipole-dipole coupling. Variation of the pump pulse delay leads to modulation of the intensity of the refocused echo. The periodicity is a function of the distance dependent coupling between the spin labels (Pannier et al., 2000).

For effective usage of the SDSL-EPR data in a *de novo* structure prediction algorithm that relies on a backbone-only protein model, those distances need to be translated into possible distance restraints for the closest atoms represented in the model, which in our case are the distances between the  $C_{\beta}$ -atoms of the spin labeling sites ( $D_{BB}$ ). In the case of glycine, which lacks a  $C_{\beta}$ -atom, the  $H_{\alpha 2}$ -atom is used instead. The side-chain flexibility of the spin label prevents an unambiguous translation from  $D_{SL}$  into  $D_{BB}$  due to its unknown conformation on the protein. Additionally, the SDSL-EPR experiment is conducted on a double cysteine mutant protein to which spin labels have been covalently bound – a species that is distinct from the wild-type protein, and might have a different structure and dynamics. Lastly, the SDSL-EPR experiment itself and the fitting procedures used to translate the primary DEER data into a distance distribution are accompanied by uncertainties. To quantify the agreement of  $D_{SL}$  with  $D_{BB}$  a knowledge-based potential based on the CONE model was introduced (Alexander et al., 2008; Hirst et al., 2011). The scoring function scores  $D_{SL}$ - $D_{BB}$  ranges of  $-12.5 \text{ \AA}$  to  $+12.5 \text{ \AA}$ , which covers the minimum and maximum difference between  $D_{SL}$  and  $D_{BB}$  (Alexander et al., 2008; Hirst et al., 2011). It assigns a score ranging from 0 (no

agreement) to  $-1$  (optimal agreement) to each  $D_{SL}$ - $D_{BB}$  pair in a protein model. An additional scoring function is used to penalize conformations with  $D_{SL}$ - $D_{BB}$  differences less than  $-12.5 \text{ \AA}$  or greater than  $12.5 \text{ \AA}$  with the purpose of drawing restraints into the  $-12.5 \text{ \AA}$  to  $+12.5 \text{ \AA}$  range (Fischer et al., 2015).

## 2.3. Benchmark setup

To evaluate the influence of SDSL-EPR derived structural restraints on *de novo* protein structure prediction, multiple folding simulations were performed. In a first experiment, the conformational space of soluble monomeric BAX was sampled in the absence of SDSL-EPR restraints. Therefore, the above-mentioned structure prediction protocol was altered so that the SDSL-EPR potential was turned off. Additional folding simulations with the experimentally determined SDSL-EPR distance restraints were performed for soluble monomeric BAX as well as with multiple sets of simulated SDSL-EPR restraints. For each setup, 7500 models were sampled in independent folding trajectories. The sampling accuracy was quantified by computing the RMSD100 (Eq. (2)) (Carugo and Pongor, 2001) with respect to the soluble monomeric BAX structure determined by NMR spectroscopy (PDB ID 1F16, model 8). The discrimination power of the scoring functions was computed using the enrichment metric (see Eq. (3)) (Woetzel et al., 2012). For homodimeric BAX, the same approach was used for the dimerization domain ( $\alpha$ -helices 2–5). RMSD100 computation (see Eq. (2)) was performed with respect to the crystal structure (PDB ID 4BDU).

## 2.4. Simulation of additional SDSL-EPR distance restraints for soluble monomeric BAX

It seems reasonable to assume that a larger number of SDSL-EPR distance restraints would result in improvements regarding the accuracy of the sampled models as well as the reliability with which accurate models can be selected. To evaluate the influence of the number of restraints on sampling accuracy and model selection, we simulated additional SDSL-EPR distance restraints based on the NMR structure for soluble monomeric BAX (PDB ID 1F16, model 8). The simulation of the additional SDSL-EPR distance restraints consisted of two steps: the selection of pairs of spin labeling sites and the simulation of the spin-spin distance between the two spin labeling sites (Procedure S5). The selection of suitable spin labeling sites was performed using a location selection algorithm that relies on the protein's sequence and predicted secondary structure (Kazmier et al., 2011). It employs Monte Carlo sampling to distribute spin labeling pairs over all SSEs. To avoid buried spin labeling sites, only residues that are predicted to be solvent-exposed were considered. For the resulting set of spin labeling pairs, the spin-spin distance was simulated using the CONE model (Alexander et al., 2008; Hirst et al., 2011). Briefly, the CONE model implicitly models the structure and dynamics of MTSL as a motion-on-a-cone. It yields a probability distribution for the difference between the spin-spin distance ( $D_{SL}$ ) and the  $C_{\beta}$ - $C_{\beta}$  distance ( $D_{BB}$ ) of the spin labeling sites. This model has been successfully evaluated on experimentally determined SDSL-EPR distances for T4-lysozyme and  $\alpha$ A-crystallin (Alexander et al., 2008; Hirst et al., 2011). By adding the predicted distribution to the  $C_{\beta}$ - $C_{\beta}$  distance in the NMR structure of soluble monomeric BAX, the spin-spin distance for a pair of spin labeling sites can be simulated. Using this protocol, three additional sets consisting of 30, 40, and 50 SDSL-EPR distance restraints were simulated for soluble monomeric BAX.

### 2.5. Calculating SDSL-EPR score enrichments

The RMSD100 metric (Carugo and Pongor, 2001) was used to quantify structural dissimilarity between models. The RMSD100 is the protein-size normalized root-mean-square-deviation of the backbone coordinates computed as

$$\text{RMSD100} = \text{RMSD} / (1 + \log \sqrt{L/100}) \quad (2)$$

with  $L$  being the length of the protein chain. The enrichment is used to evaluate how well a scoring function is able to select the most accurate models from a given set of models. The models of a given set  $S$  are sorted by their RMSD100 values and the 10% of the models with the lowest RMSD100 values put into the set  $P$  (positive) the rest of the models will be put into the set  $N$  (negative). The models of  $S$  are then also sorted by their assigned scoring value and the 10% of the models with the lowest (most favorable) score are put into the set  $T$ . The models, which are in  $P$  and in  $T$ , are the models, which are correctly selected by the scoring function, and their number will be referred to as  $TP$  (true positive). The numbers of models, which are in  $P$  but not in  $T$ , are the models, which are not selected by scoring function despite being among the most accurate ones. They will be referred to as  $FN$  (false negative). The enrichment is then calculated as

$$e = \frac{\#TP}{\#P} \cdot \frac{\#P + \#N}{\#P} \quad (3)$$

The positive models are in this case considered the 10% of the models with the lowest RMSD100 values. Therefore,  $\frac{\#P + \#N}{\#P}$  is fixed at a value of 10.0. Consequently, the enrichment can range from 0.0 to 10.0. An enrichment value of 1.0 indicates that the scoring function is unable to discriminate between accurate and inaccurate models and the probability of selecting an accurate model corresponds to random chance. Enrichment values greater than 1.0 indicate that the scoring function is able to select accurate models with a probability that is greater than random chance. Enrichment values smaller than 1.0 indicate that the scoring function selects against accurate models and the probability of selecting accurate models is less than random chance.

### 2.6. Using clustering for model selection

Clustering by RMSD was used for additional model selection trials. A partitioning-based clustering approach was used, which is based on  $k$ -means and implemented in the cluster package (Maechler et al., 2015) in R. Clustering was performed using a maximum average dissimilarity of 3 Å. Clusters were only considered if their population size was at least 1% of all models sampled. The reported RMSD100 values are between the cluster centers (medoids) and the experimentally determined structure.

## 3. Results

In this section, the effect of SDSL-EPR distance restraints on *de novo* protein structure prediction is evaluated under the aspects of sampling accuracy and discrimination power. The features of BAX that complicate *de novo* protein structure prediction in the absence of experimental data are discussed. Subsequently, the effect of SDSL-EPR distance restraints on sampling accuracy and discrimination power are evaluated. Reported results are the accuracies of the models with the lowest RMSD100 values (henceforth labeled as most accurate models) as well as the percentage of models with an RMSD100 value (see Eq. (2)) of less than 8 Å with respect to the corresponding NMR or X-ray crystal structure available. Additionally, the enrichment is reported, which is the per-

centage of the accurate models that can be selected by the scoring function (see Eq. (3)).

### 3.1. Summary of the available SDSL-EPR data for soluble monomeric and homodimeric BAX

The benchmark was performed on the soluble monomeric and the homodimeric states of BAX. Here, we give a summary about the SDSL-EPR data available for both states and how well the respective experimentally determined reference structures (PDB ID 1F16 for soluble monomeric BAX and PDB ID 4BDU for homodimeric BAX) agree with SDSL-EPR data. The latter is important because we evaluate the accuracy of the predicted models based on their structural similarity to the respective experimentally determined structure.

Data taken from the literature where Bleicken et al. measured twenty-five distances for soluble monomeric BAX by Q-band DEER (Table S1) (Bleicken et al., 2014). In their study, the spin labeling sites were selected based on several criteria: While the spin labels should reveal relevant information about the protein structure, their introduction should not change the protein's fold or affect the stability or function of the protein. The spin labeled proteins used in Bleicken's study were shown to retain their fold and the ability to permeabilize large unilamellar vesicles with a composition mimicking the MOM (Bleicken et al., 2014). The structure of soluble monomeric BAX was determined by Suzuki et al. through NMR spectroscopy (PDB ID 1F16) (Suzuki et al., 2000) and was used here as a baseline for comparison. To evaluate the suitability of the available SDSL-EPR distance data for protein structure prediction, all models from the NMR ensemble were scored for agreement with the SDSL-EPR restraints using the CONE model (Alexander et al., 2008; Hirst et al., 2011). The average difference between the observed  $D_{SL}$  and  $D_{BB}$  was 6.3 Å with an average score of  $-0.84$  (Table S1, perfect agreement score is  $-1.00$  whereas the worst possible agreement score is 0.00).

Data taken from the literature where SDSL-EPR distance measurements were performed on membrane embedded, active and homooligomeric BAX by Bleicken et al. (2014). Of the forty-one measured distances, seventeen are within the dimerization domain whereas the remaining twenty-four are within the piercing domain or between dimerization and piercing domain (Bleicken et al., 2014). A crystal structure of a truncated BAX variant covering only the dimerization domain was published by Czabotar et al. (PDB ID 4BDU) (Czabotar et al., 2013). In order to benchmark our algorithm, we consequently opted for predicting the dimerization domain only, for which a reference structure was available. Although the reference structure (PDB ID 4BDU) was crystallized in the absence of the membrane, Bleicken et al. (2014) showed that 4BDU well represents the fold of the dimerization domain as its present in the full length active protein embedded in liposomes and consequently is suitable as a baseline for comparison. This is in agreement with our evaluation, in which we used the CONE model (Alexander et al., 2008; Hirst et al., 2011) to evaluate the agreement of the X-ray crystal structure with the SDSL-EPR data measured by Bleicken et al.: the average difference between  $D_{SL}$  and  $D_{BB}$  was 3.1 Å with an SDSL-EPR agreement score of  $-0.94$  (Table S2), indicating that the crystal structure is in good agreement with the SDSL-EPR data. In this study, we folded residues 54–122, which is identical to the region determined in the crystal structure (PDB ID 4BDU). Of the seventeen published SDSL-EPR distance restraints within the dimerization domain, we only used eleven restraints. The six discarded restraints are between the dimerization domain and residue 126, which is not included in 4BDU.

Additional analysis was conducted to evaluate if bending of secondary structure elements (SSEs) are required to satisfy the SDSL-

EPR restraints. This is important because the complexity of structural sampling does not allow for exhaustive sampling of all possible conformations. On these grounds, BCL::Fold reduces the complexity of the sampling space by assembling the tertiary structure from idealized, straight SSEs, only allowing small deviations from idealized parameters. Therefore, in a second test,  $\alpha$ -helices in the NMR models and the X-ray crystal structure were straightened before scoring in order to quantify the influence of bent SSEs on the agreement with the SDSL-EPR distance restraints. In this context, idealization means setting the dihedral angles ( $\varphi$ ,  $\psi$ ) to  $(-60^\circ, -40^\circ)$  for  $\alpha$ -helices and to  $(-135^\circ, 135^\circ)$  for  $\beta$ -strands. To evaluate the influence of deviations from idealized dihedral angles (bending or kinks) on the agreement with the SDSL-EPR distance restraints, the experimentally determined structures for soluble monomeric BAX (PDB ID 1F16, model 8) and homodimeric BAX (PDB ID 4BDU) were idealized using the BCL software suite (Procedure S6), which sets the dihedral angles of the SSEs to aforementioned idealized values. The agreement of the idealized structures with the SDSL-EPR data was subsequently quantified, showing an average agreement score of  $-0.88$  for soluble monomeric BAX. The resulting agreement is no diminishment from the agreement score for the non-idealized structure of  $-0.88$ . This indicates that a structure with idealized SSEs can achieve agreement with the SDSL-EPR distance data and focusing the sampling on SSEs with idealized dihedral angles won't negatively influence the prediction of the protein's tertiary structure. Based on this analysis, the eighth model of the NMR ensemble for soluble monomeric BAX was selected as reference structure for the benchmark because it had the best agreement with the SDSL-EPR data. Notably, the same model was selected based on the RMSD by Bleicken et al. (2014) between the experimental time domain DEER traces and those simulated with the software MMM2013.2 (Polyhach et al., 2011), based on a rotamer library approach. For homodimeric BAX and straightened SSEs, the average difference between  $D_{SL}$  and  $D_{BB}$  was  $3.8 \text{ \AA}$  with an SDSL-EPR agreement score of  $-0.90$  (Table S2), which again does not constitute a significant diminishment of the SDSL-EPR agreement score for idealized SSEs; indicating that structure assembly from idealized SSEs won't hinder the prediction for homodimeric BAX.

### 3.2. The properties of BAX complicate *de novo* protein structure prediction in the absence of experimental data

BCL::Fold scores protein structures using knowledge-based potentials derived from statistics over properties of protein structures deposited in the PDB (see Section 2 for details). Therefore, if a protein structure significantly deviates from the statistics, an unfavorable score is assigned as compared to alternative conformations, hindering prediction of the protein's tertiary structure with BCL::Fold.

BAX monomers consist of 192 residues, forming nine  $\alpha$ -helices. Due to its ability to interact with membranes, some portions of the soluble monomeric BAX structure are outliers to statistics collected from experimentally determined structures of soluble proteins. Specifically, the exposure of the residues in  $\alpha$ -helix 9 as well as the relative orientation of  $\alpha$ -helix 9 with respect to other  $\alpha$ -helices feature poor agreement with statistics (see Section 2) collected from experimentally determined structures in the PDB (Fig. 1B). Notably,  $\alpha$ -helix 9 is proposedly transmembrane after membrane insertion (Bleicken et al., 2014; Westphal et al., 2014). In consequence, a knowledge-based potential function, as used by many *de novo* folding algorithms, ranks the experimentally determined structure of soluble monomeric BAX poorly compared to alternative arrangements (Fig. 1D, F). BCL::Fold (Karakas et al., 2012), which uses knowledge-based potentials to evaluate the accuracy of a model (Woetzel et al., 2012), is no exception.

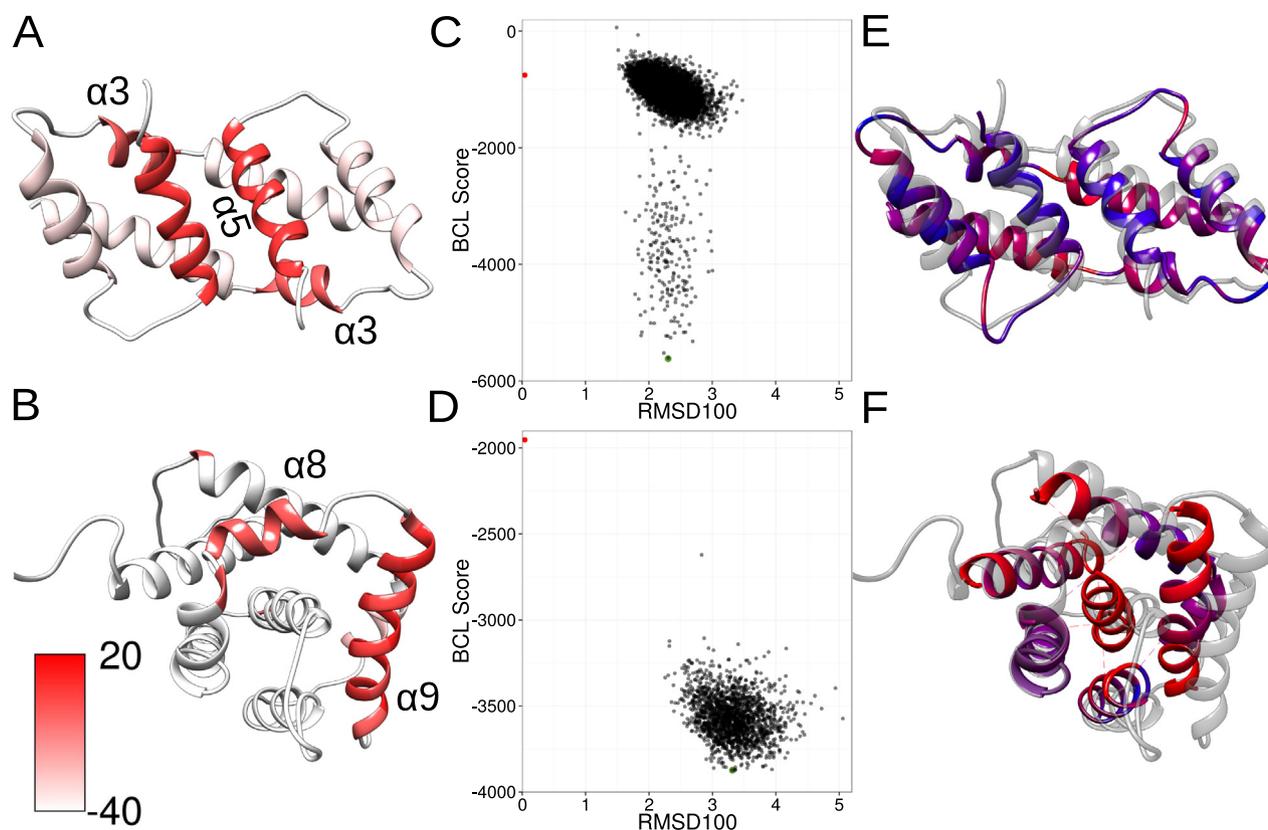
This can be demonstrated by relaxing the experimentally determined structure of soluble monomeric BAX (PDB ID 1F16) in the BCL::Fold force field. During the relaxation, small structural perturbations are applied to the NMR structure. After each perturbation, the resulting structure is scored using the BCL score. This results in a set of models, which structurally deviate from the NMR structure but have a more favorable BCL score. The structures with the lowest score are most likely to be predicted by BCL::Fold as the native structure of soluble monomeric BAX. Fig. 1D shows the BCL scores and dissimilarities to the NMR structure for a set of relaxed models. Soluble monomeric BAX has a local score minimum for conformations with an RMSD100 value of  $3 \text{ \AA}$  to  $4 \text{ \AA}$  (Fig. 1D). The model with the most favorable score shows  $\alpha$ -helix 9 moving closer into a pocket formed by  $\alpha$ -helices 2–5 (Fig. 1F), which reduces the exposure of the residues in  $\alpha$ -helix 9 and results in a more favorable score.

These difficulties in scoring/ranking the sampled models make soluble monomeric BAX an appropriate test case to evaluate if scoring problems can be overcome by incorporating limited structural data from SDSL-EPR experiments. Further, as SDSL-EPR data recently became available for the dimerization domain of homooligomeric BAX, BAX is also a test case for determining a protein's structure in different, biologically relevant conformations. For homooligomeric BAX, similar challenges in the ranking of models in the absence of experimental data can be observed. The radius of gyration of the crystal structure of the dimerization domain (PDB ID 4BDU) significantly deviates from statistics collected from known structures in the PDB. Additional SSE- and residue-based deviations are observed for the exposure of residues and SSE orientations. These deviations are particularly pronounced for the  $\alpha$ -helices 3 and 5 (Fig. 1A). Repeating the relaxation experiment as described above for homodimeric BAX, shows a local score minimum for structures with an RMSD100 value between  $2 \text{ \AA}$  and  $3 \text{ \AA}$  relative to the crystal structure (Fig. 1C). The model with the most favorable BCL score shows a change in the packing of  $\alpha$ -helices 2–5 and a slight reduction of the radius of gyration (Fig. 1E). Comparably to soluble monomeric BAX, these scoring problems make the tertiary structure of homodimeric BAX hard to predict with BCL::Fold because the scoring function does not detect the crystal structure as native-like.

### 3.3. SDSL-EPR distance restraints can overcome *de novo* sampling and scoring problems

By using SDSL-EPR distance restraints, it is possible to overcome scoring and sampling problems, which hinder *de novo* protein structure prediction. As demonstrated in the previous section, the NMR ensemble of soluble monomeric BAX and the X-ray crystal structure of homodimeric BAX score poorly in the BCL::Fold knowledge-based scoring function, which hinders prediction of a model that is in good agreement with the NMR- or X-ray-derived models.

Using SDSL-EPR restraints for soluble monomeric BAX results in a shift of the RMSD100 distributions by around  $1.5 \text{ \AA}$  to models in better agreement with the NMR-derived model (Fig. 2A–D). Whereas without SDSL-EPR data, the most accurate model sampled had an RMSD100 value of  $5.9 \text{ \AA}$ ; by using SDSL-EPR data, the RMSD100 value of the most accurate model could be improved to  $3.9 \text{ \AA}$  (Table 1 and Fig. 2). For further evaluation of the sampling accuracy, the ten best models by RMSD100 were selected and their average RMSD100 value,  $\mu_{10}$ , was calculated (the average RMSD100 values for different numbers of models are shown in Fig. S1). In the absence of SDSL-EPR data, the  $\mu_{10}$  value was  $7.0 \text{ \AA}$ , whereas with SDSL-EPR data the  $\mu_{10}$  value improved to  $5.0 \text{ \AA}$ . Additionally, the percentage of models with an RMSD100 value of less than  $8 \text{ \AA}$ ,  $\tau_8$ , was calculated. For folding without SDSL-EPR data,



**Fig. 1.** The properties of soluble monomeric BAX and homodimeric BAX hinder *de novo* protein structure prediction. (A, B) Due to their properties, parts of homodimeric BAX (A) and soluble monomeric BAX (B) score poorly when evaluated in the BCL::Fold knowledge-based scoring function, hindering prediction of the tertiary structure. Color code: white-red scale with white being good score and red being poor score. (C, D) Relaxing the NMR and X-ray crystal structures in the BCL::Fold force field shows score minima for alternative conformations. Black dots represent alternative conformations with their BCL score (y-axis) and the protein-size-normalized RMSD (RMSD100) relative to the NMR/X-ray crystal structure (x-axis). The NMR/crystal structure is shown as red dot. Green dots are the best scoring structures, which are shown in (E, F). (E) Relaxing the X-ray crystal structure of homodimeric BAX (PDB ID 4BDU) in the BCL::Fold force field results in tighter packing of  $\alpha$ -helices 3 and 5 and a slightly reduced radius of gyration. The relaxed model is shown on a blue-red scale, with blue being structural similarity to the crystal structure (grey) and red being structural dissimilarity. (F) Relaxing the NMR structure of soluble monomeric BAX (PDB ID 1F16, model 8) in the BCL::Fold force field results in  $\alpha$ -helix 9 moving closer into a pocket formed by  $\alpha$ -helices 2–5. The relaxed model is shown on a blue-red scale, with blue being structural similarity to the NMR structure (grey) and red being structural dissimilarity.

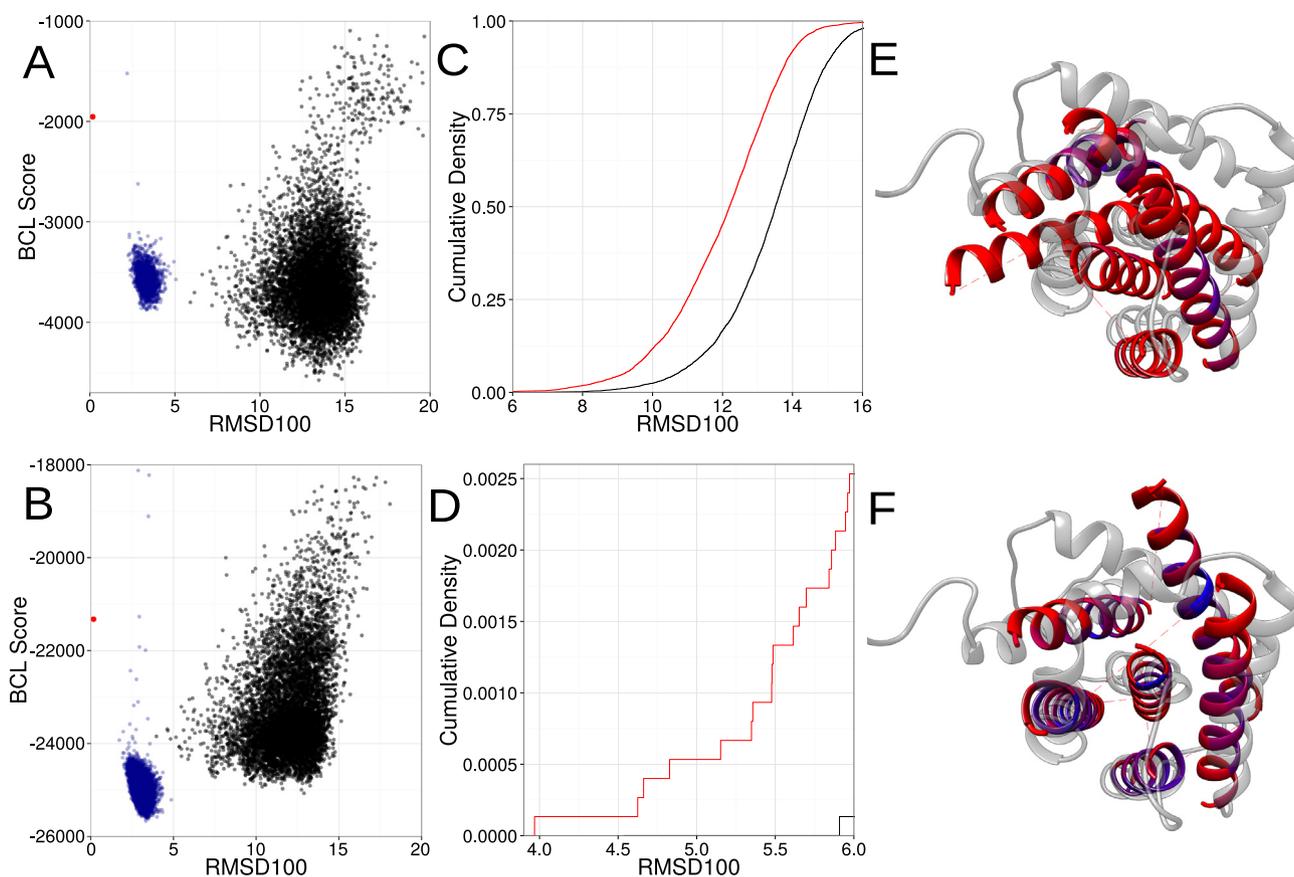
the  $\tau_8$  value was 0.3%, whereas when folding with SDSL-EPR distance restraints the  $\tau_8$  value improved to 1.9%. Using SDSL-EPR restraints for the dimerization domain of homooligomeric BAX improved the RMSD100 value of the most accurate model from 5.7 Å to 3.3 Å. The  $\mu_{10}$  and  $\tau_8$  values improved from 6.8 Å to 3.4 Å and from 0.1 to 16.7%, respectively (Table 1 and Fig. 3). Additional model selection trials were performed using clustering. For soluble monomeric and homooligomeric BAX – in the absence of SDSL-EPR distance restraints, the clusters closest to the experimentally determined structure had an RMSD100 value of 9.2 Å and 11.4 Å, respectively. By using SDSL-EPR distance restraints, clusters with an RMSD100 of 7.1 Å and 4.8 Å could be detected for soluble monomeric and homooligomeric BAX.

Besides the sampling, a protein structure prediction method must be able to select the most accurate models among the sampled models. To evaluate the ability of the scoring function to select the most accurate models sampled during *de novo* folding, score enrichments were calculated. The enrichment indicates how well the scoring function is able to distinguish between accurate and inaccurate models (see Section 2 for details and Eq. (3)). The term accurate is hereby defined as being among the 10% of the models with the lowest RMSD100 value relative to the experimentally determined structure. For the models generated in the absence of SDSL-EPR data, the enrichment for soluble monomeric BAX was 0.4 (Table 1). The enrichment of less than 1.0 for the BCL::Fold energy function indicates that it actually selects against

topologies in agreement with the X-ray-derived model, presumably due to the poor score of the  $\alpha$ -helix 9 discussed above. With SDSL-EPR distance restraints, the enrichment improved to 1.5. The improvement in enrichment demonstrates that by using SDSL-EPR distance restraints, protein structure prediction methods can overcome model discrimination challenges. For homooligomeric BAX, usage of SDSL-EPR restraints improved the enrichment from 1.3 to 2.1 (Table 1).

### 3.4. A larger number of restraints improves sampling accuracy and selection of accurate models

To evaluate the influence of the number of restraints on the sampling accuracy as well as the algorithm's ability to select accurate models, three additional restraint sets with different numbers of restraints were simulated based on the NMR structure of BAX (PDB ID 1F16, model 8). The spin labeling sites were chosen in order to distribute measurements across all SSEs (see Section 2 for details). The experimentally determined restraint set consisted of twenty-five restraints, whereas the simulated restraint sets for the NMR structure of soluble, monomeric BAX (PDB ID 1F16, model 8) consisted of thirty, forty, and fifty restraints, respectively. To fold soluble BAX with the simulated restraints the same protocol was used as for the experimentally determined restraint set. The number of restraints has a significant effect on the sampling accuracy as well as our ability to select accurate models (Table 1). Whereas the



**Fig. 2.** Structure prediction results for soluble monomeric BAX. (A) Protein structure prediction without SDSL-EPR distance restraints results in a poor correlation between the score of the *de novo* predicted models (black dots) and their accuracy (quantified as RMSD100 relative to the experimentally determined structure). The experimentally determined structure (red dot) and the experimentally determined structure relaxed in the BCL::Fold force field (blue dots) score worse than the *de novo* predicted models. (B) By using SDSL-EPR distance restraints, the score gap between the experimentally determined structure (red dot) and the *de novo* sampled models (black dots) could be reduced. The experimentally determined structure relaxed in the BCL::Fold force field (blue dots) scores better than the *de novo* sampled models. The BCL score of the experimentally determined structure and the relaxed structures includes the EPR agreement score, resulting in lower scores than in (A). (C, D) Using SDSL-EPR distance restraints increased the sampling density of models in agreement with the NMR-derived model (red – with SDSL-EPR distance restraints, black – without). (E, F) In the most accurate model sampled with SDSL-EPR distance restraints (F, blue-red scale, RMSD100 = 3.9 Å), the placement of the SSEs is more similar to the experimentally determined structure (PDB ID 1F16, model 8, grey), than for sampling without SDSL-EPR distance restraints (E, blue-red scale, RMSD100 = 5.9 Å). Color coding: blue-red scale with blue being structural similarity to the experimentally determined structure and red being dissimilarity.

**Table 1**

Sampling accuracy and enrichment are improved by SDSL-EPR distance restraints. By using SDSL-EPR distance restraints in protein structure prediction the sampling accuracy can be improved as it is seen for the RMSD100 values of the most accurate (by RMSD100) model sampled (Best), the average of the ten best models sampled ( $\mu_{10}$ ) and the percentage of the models with an RMSD100 value less than 8 Å ( $\tau_8$ ). A larger number of SDSL-EPR restraints leads to more substantial improvements, which was demonstrated by simulating additional SDSL-EPR restraint sets consisting of 30, 40, and 50 restraints. More restraints constantly result in more pronounced improvements in the sampling accuracy. SDSL-EPR distance restraints also improve the ability to select the accurate models among the sampled models, which is shown by improved enrichment values ( $e$ ).

Protein	Restrains	RMSD100 [Å]		$\tau_8$ [%]	$e$
		Best	$\mu_{10}$		
Monomer	Without	5.9	7.0	0.2	0.4
Monomer	25 experimental	3.9	5.0	2.5	1.5
Monomer	30 simulated	4.2	4.9	7.5	4.1
Monomer	40 simulated	4.1	4.4	11.4	4.2
Monomer	50 simulated	3.9	4.2	11.5	4.5
Dimer	Without	5.7	6.8	0.1	1.3
Dimer	11 experimental	3.3	3.4	13.7	2.1

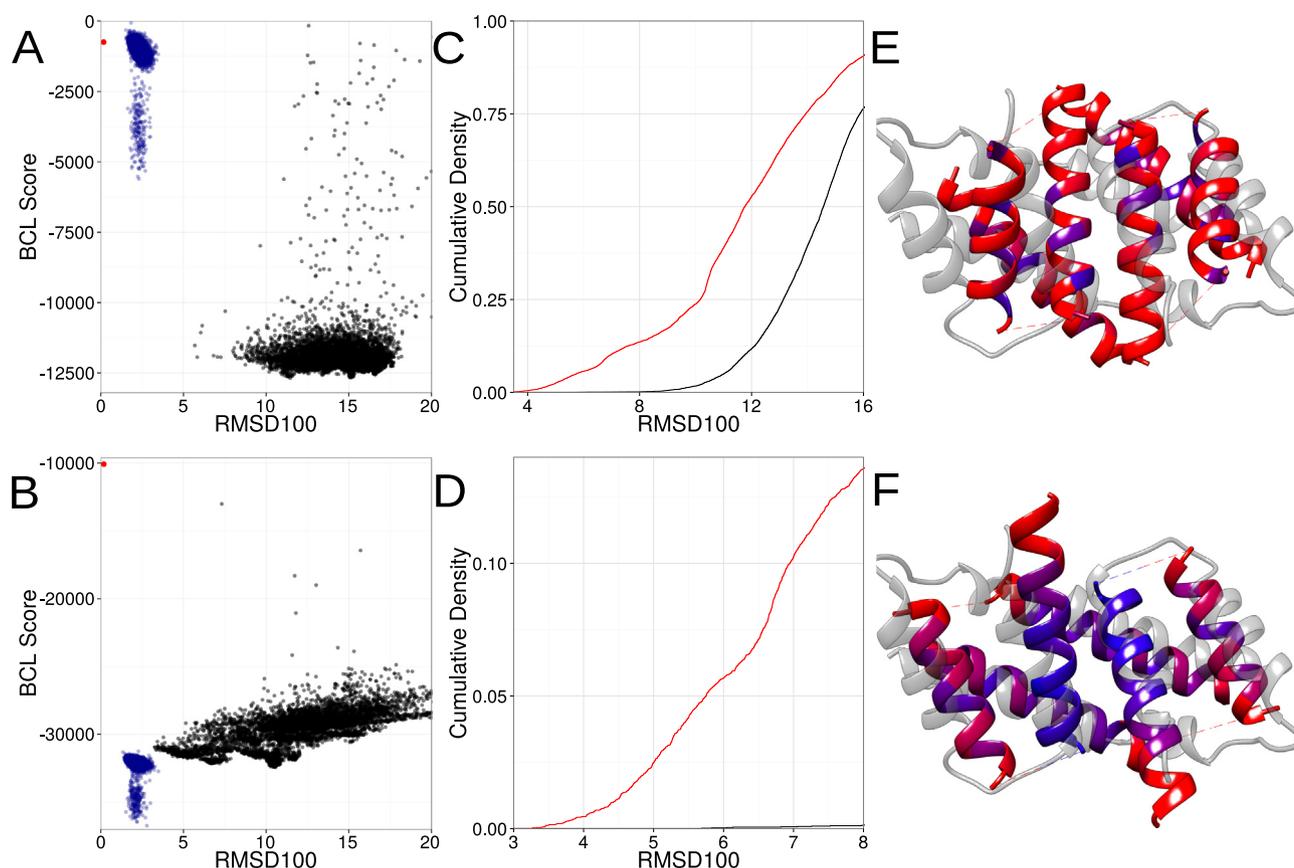
$\mu_{10}$  value for the twenty-five experimentally determined restraints was 5.0 Å, it was 4.9 Å for thirty restraints, 4.4 Å for forty restraints and 4.2 Å for fifty restraints. For folding with twenty-five

restraints, the  $\tau_8$  value was 1.9%, with thirty restraints 7.5%, with forty restraints 11.4%, and with fifty restraints 11.5%. The enrichment of 1.5 for folding with twenty-five restraints improves to 4.1 for thirty restraints, 4.2 for forty restraints, and 4.5 for fifty restraints (Table 1).

## 4. Discussion

### 4.1. Interpretation of the reported sampling accuracies and enrichments

It should be noted that comparison to 1F16 and 4BDU is somewhat limited: The RMSD100 values between the twenty individual models in 1F16 ranges from 1.7 Å to 4.7 Å with an average of 3.0 Å. The relatively low precision of the NMR-derived models represents an upper limit for the accuracy of 1F16. In result, any model that approaches this accuracy limit is in agreement with 1F16 within its accuracy limits. Additionally, in the case of the dimeric structure, deviations may be caused by 4BDU being derived from a protein crystal with a reported resolution of 3.0 Å and in absence of membranes or membrane mimics, whereas the SDSL-EPR measurements were completed on full-length BAX variants inserted into large unilamellar vesicles mimicking the mitochondrial outer



**Fig. 3.** Structure prediction results for homodimeric BAX. (A) Protein structure prediction without SDSL-EPR distance restraints results in a poor correlation between the score of the *de novo* sampled models (black dots) and their accuracy (quantified as RMSD100 relative to the experimentally determined structure). The experimentally determined structure (red dot) and the experimentally determined structure relaxed in the BCL::Fold force field (blue dots) score significantly worse than the *de novo* sampled models. (B) Protein structure prediction with SDSL-EPR distance restraints results in an improved correlation between the score of the sampled models (black dots) and their accuracy. Whereas the experimentally determined structure (red dot) scores worse than the sampled models, the relaxed experimentally determined structure (blue dots) scores better than the sampled models. The BCL score of the experimentally determined structure and the relaxed structures includes the EPR agreement score, resulting in lower scores than in (A). (C, D) Using SDSL-EPR distance restraints significantly improves the sampling density of models in agreement with the NMR- and X-ray derived models, result in a shift of the distribution of about 6 Å (red – with SDSL-EPR distance restraints, black – without). (E) Without SDSL-EPR distance restraints the placement of the SSEs of the most accurate model sampled (blue-red scale, RMSD100 = 5.7 Å) is dissimilar to the experimentally determined structure (grey). (F) By using SDSL-EPR distance restraints the SSE placement of the most accurate mode sampled (blue-red scale, RMSD100 = 3.3 Å) resembles the experimentally determined structure (grey). Color coding: blue-red scale with blue being structurally similar to the experimentally determined structure and red being dissimilarity.

membrane lipid composition (MOM-LUVs), *i.e.* in a more native-like environment (Bleicken et al., 2014). Arguably, a comparison to the SDSL-EPR relaxed version of 4BDU and 1F16 could provide a more accurate measure of success of the folding simulation. As such models are however biased by the BCL::Fold scoring function we opted for comparison with the original PDB entries.

#### 4.2. Energy function and sampling limitations hinder *in silico* protein structure prediction

The major obstacle and challenge of *in silico* determination of a protein's tertiary structure is the vast conformational search space combined with the complicated models needed to compute an accurate estimate of a protein's free energy. These obstacles are overcome by simplifications in the scoring function and sampling space that are often coupled to a simplified representation of the protein. In concrete terms, simultaneous and exhaustive sampling of the  $\phi$ - and  $\psi$ -angles in the protein backbone and  $\chi$ -angles in the protein side-chains is prohibitive. BCL::Fold drastically reduced the search space by eliminating all  $\chi$ -angles – side-chains are represented as 'superatoms', eliminating  $\phi$ - and  $\psi$ -angles in flexible loop regions by not explicitly modeling loop regions, and assembling predicted SSEs starting from idealized  $\phi$ - and  $\psi$ -angles allow-

ing only for limited deviations. Additionally, explicit simulation of the protein's environment, like the membrane or the solvation water molecules, is circumvented by implicit models. Still, enumeration of all possible folds within an acceptable timeframe remains prohibitive for larger proteins. As shown in Figs. 2A and 3A, in the absence of any experimental data neither are models in agreement with the NMR- and X-ray-derived models sampled in a frequent manner, nor is it possible to distinguish more accurate models from less accurate models. For soluble monomeric BAX or the dimerization domain of membrane-embedded homooligomeric BAX, the experimentally determined structures both score poorly in the BCL scoring function. Even after relaxing the experimentally determined structures in the BCL::Fold force field to find a conformation in agreement with the NMR- and X-ray derived models in a score minimum, the relaxed structures score worse than models that are not in agreement with the NMR- and X-ray-derived models (Figs. 2A and 3A).

#### 4.3. SDSL-EPR measurements can overcome the limitations of *de novo* protein structure prediction

SDSL-EPR distance measurements can be performed in a native-like environment and provide experimental data that can be inter-

interpreted as structural restraints, thus compensating for the algorithm's limitation in sampling the large conformational space and estimating the free energy of these conformations accurately. Direct incorporation of the SDSL-EPR distance data into the BCL::Fold scoring function reduces the complexity of the energy function by removing local minima in the scoring function that are inconsistent with the experimental SDSL-EPR distance data, reinforcing conformations that are. Therefore, incorporation of SDSL-EPR distance restraints can overcome limitations in sampling and scoring. This was demonstrated by relaxing the experimentally determined structures in the BCL::Fold force field using SDSL-EPR restraints (Figs. 2B and 3B). The relaxed structures are similar to the NMR- and X-ray derived models and have a more favorable score than most of the sampled models. As a direct result of the improved pseudo-energy landscape, the Monte Carlo Metropolis algorithm favors conformations that are in agreement with the SDSL-EPR data, leading to the sampling of models that are in better agreement with the NMR- and X-ray-derived models. Significant shifts of the accuracy distributions are observed for soluble monomeric BAX as well as the dimerization domain of homo-oligomeric BAX (Figs. 2C and 3C, and Table 1). For soluble monomeric BAX, the accuracy distribution improves by about 1.5 Å, whereas for homo-oligomeric BAX the improvement is about 4 Å. Additionally, using of SDSL-EPR distance data mitigates the problem of distinguishing accurate models from inaccurate models (Figs. 2B and 3B, and Table 1). This effect is more pronounced for homooligomeric BAX, for which the enrichment improves from 1.3 to 2.1. The score-accuracy plots in Fig. 3B show an improved correlation between score and RMSD100. Although the best scoring model is still not in perfect agreement with the X-ray-derived model, a high model density exists in the 3 Å to 5 Å range, which could be detected through clustering. The results of this study demonstrate that SDSL-EPR distance restraints can mitigate the limitations of *de novo* protein structure prediction algorithms, by increasing the sampling frequency of the models that are in agreement with the SDSL-EPR data and by complementing the energy evaluation with structural restraints.

## 5. Conclusion

This study demonstrates that even a limited number of SDSL-EPR distance restraints are able to introduce score minima for conformations, which have better agreement with the structural models derived from NMR or crystallography. Therefore, challenges in conformational sampling and model discrimination in *de novo* protein structure prediction can be overcome through incorporation of sparse SDSL-EPR distance restraints. This was demonstrated by the improved accuracy of the models as well as the improved enrichment of accurate models. In conclusion, a combined approach of *de novo* protein structure predictions methods and SDSL-EPR distance restraints is able to predict the fold of larger proteins that adopt multiple conformations.

## Availability

The BCL software suite is available at <http://www.meilerlab.org/bclcommons> under academic and business site licenses. The BCL source code is published under the BCL license and is available at <http://www.meilerlab.org/bclcommons>.

## Author contributions

A.W.F, E.B, G.J., J.M, S.B., and A.J.G-S. wrote this manuscript. A.W.F. performed the computational experiment and the data analysis.

## Acknowledgements

We want to thank Benjamin Mueller for thorough proofreading of the manuscript.

Work in the Meiler laboratory is supported through NIH (R01 GM080403, R01 GM099842, R01 DK097376) and NSF (CHE 1305874). This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

A.J.G-S. and S.B. were supported by the Max Planck Society and the German Ministry for Education and Research (BMBF, grant N. 0312040).

Parts of the data analysis were performed using the R package with ggplot2 (Wickham, 2009) and cluster (Maechler et al., 2015). The renderings of the models were created using Chimera (Pettersen et al., 2004). The composite figures were created using Inkscape.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jsb.2016.04.014>.

## References

- Alexander, N., Al-Mestarihi, A., Bortolus, M., Mchaourab, H., Meiler, J., 2008. De Novo high-resolution protein structure determination from sparse spin-labeling EPR data. *Structure* 16, 181–195. <http://dx.doi.org/10.1016/j.str.2007.11.015>.
- Bamberger, M.E., Landreth, G.E., 2002. Inflammation, apoptosis, and Alzheimer's disease. *Neuroscientist* 8, 276–283. <http://dx.doi.org/10.1177/1073858402008003013>.
- Bleicken, S., Jeschke, G., Stegmueller, C., Salvador-Gallego, R., García-Sáez, A.J., Bordignon, E., 2014. Structural model of active Bax at the membrane. *Mol. Cell* 56, 496–505. <http://dx.doi.org/10.1016/j.molcel.2014.09.022>.
- Carugo, O., Pongor, S., 2001. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci.* 10, 1470–1473. <http://dx.doi.org/10.1110/ps.690101.of>.
- Czabotar, P.E., Lessene, G., Strasser, A., Adams, J.M., 2014. Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat. Rev. Mol. Cell Biol.* 15, 49–63. <http://dx.doi.org/10.1038/nrm3722>.
- Czabotar, P.E., Westphal, D., Dewson, G., Ma, S., Hockings, C., Fairlie, W.D., Lee, E.F., Yao, S., Robin, A.Y., Smith, B.J., Huang, D.C.S., Kluck, R.M., Adams, J.M., Colman, P. M., 2013. Bax crystal structures reveal how BH3 domains activate Bax and nucleate its oligomerization to induce apoptosis. *Cell* 152, 519–531. <http://dx.doi.org/10.1016/j.cell.2012.12.031>.
- de Vera, I.M.S., Blackburn, M.E., Galiano, L., Fanucci, G.E., 2013. Pulsed EPR distance measurements in soluble proteins by site-directed spin labeling (SDSL). *Curr. Protoc. Protein Sci.* 1–29. <http://dx.doi.org/10.1002/0471140864.ps1717s74>.
- Durham, E., Dorr, B., Woetzel, N., Staritzbichler, R., Meiler, J., 2009. Solvent accessible surface area approximations for rapid and accurate protein structure prediction. *J. Mol. Model.* 15, 1093–1108. <http://dx.doi.org/10.1007/s00894-009-0454-9>.
- Eguchi, K., 2001. Apoptosis in autoimmune diseases. *Intern. Med.* 40, 275–284. <http://dx.doi.org/10.2169/internalmedicine.40.275>.
- Favaloro, B., Allocati, N., Graziano, V., Di Ilio, C., De Laurenzi, V., 2012. Role of apoptosis in disease. *Aging* 4, 330–349 (Albany, NY).
- Fischer, A.W., Alexander, N.S., Woetzel, N., Karakas, M., Weiner, B.E., Meiler, J., 2015. BCL::MP-Fold: membrane protein structure prediction guided by EPR restraints. *Proteins Struct. Funct. Bioinf.* 83, 1947–1962. <http://dx.doi.org/10.1002/prot.24801>.
- Heinze, S., Putnam, D.K., Fischer, A.W., Kohlmann, T., Weiner, B.E., Meiler, J., 2015. CASP10-BCL: fold efficiently samples topologies of large proteins. *Proteins Struct. Funct. Bioinf.* 83, 547–563. <http://dx.doi.org/10.1002/prot.24733>.
- Hirst, S., Alexander, N., Mchaourab, H.S., Meiler, J., 2011. ROSETTA-EPR: an integrated tool for protein structure determination from sparse EPR data. *Biophys. J.* 100, 216a. <http://dx.doi.org/10.1016/j.bpj.2010.12.1390>.
- Jeschke, G., 2012. DEER distance measurements on proteins. *Annu. Rev. Phys. Chem.* 63, 419–446. <http://dx.doi.org/10.1146/annurev-physchem-032511-143716>.
- Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202. <http://dx.doi.org/10.1006/jmbi.1999.3091>.
- Karakas, M., Woetzel, N., Staritzbichler, R., Alexander, N., Weiner, B.E., Meiler, J., 2012. BCL: De Novo prediction of complex and large protein topologies by assembly of secondary structure elements. *PLoS One* 7, e49240. <http://dx.doi.org/10.1371/journal.pone.0049240>.
- Kazmier, K., Alexander, N.S., Meiler, J., Mchaourab, H.S., 2011. Algorithm for selection of optimized EPR distance restraints for de novo protein structure

- determination. *J. Struct. Biol.* 173, 549–557. <http://dx.doi.org/10.1016/j.jsb.2010.11.003>.
- Leber, B., Lin, J., Andrews, D.W., 2010. Still embedded together binding to membranes regulates Bcl-2 protein interactions. *Oncogene* 29, 5221–5230. <http://dx.doi.org/10.1038/onc.2010.283>.
- Leman, J.K., Mueller, R., Karakas, M., Woetzel, N., Meiler, J., 2013. Simultaneous prediction of protein secondary structure and transmembrane spans. *Proteins Struct. Funct. Bioinf.* 81, 1127–1140. <http://dx.doi.org/10.1002/prot.24258>.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2015. *Cluster: Cluster Analysis Basics and Extensions*.
- Pannier, M., Veit, S., Godt, A., Jeschke, G., Spiess, H.W., 2000. Dead-time free measurement of dipole-dipole interactions between electron spins. *J. Magn. Reson.* 142, 331–340. <http://dx.doi.org/10.1016/j.jmr.2011.08.035>.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. <http://dx.doi.org/10.1002/jcc.20084>.
- Polyhach, Y., Bordignon, E., Jeschke, G., 2011. Rotamer libraries of spin labelled cysteines for protein studies. *Phys. Chem. Chem. Phys.* 13, 2356–2366. <http://dx.doi.org/10.1039/c0cp01865a>.
- Strasser, A., O'Connor, L., Dixit, V.M., 2000. Apoptosis signaling. *Annu. Rev. Biochem.* 69, 217–245. <http://dx.doi.org/10.1146/annurev.biochem.69.1.217>.
- Suzuki, M., Youle, R.J., Tjandra, N., 2000. Structure of Bax: coregulation of dimer formation and intracellular localization. *Cell* 103, 645–654. doi: 10.1016/S0092-8674(00)00167-7.
- Weiner, B.E., Woetzel, N., Karakas, M., Alexander, N., Meiler, J., 2013. BCL::MP-fold: folding membrane proteins through assembly of transmembrane helices. *Structure* 21, 1107–1117. <http://dx.doi.org/10.1016/j.str.2013.04.022>.
- Westphal, D., Dewson, G., Menard, M., Frederick, P., Iyer, S., Bartolo, R., Gibson, L., Czabotar, P.E., Smith, B.J., Adams, J.M., Kluck, R.M., 2014. Apoptotic pore formation is associated with in-plane insertion of Bak or Bax central helices into the mitochondrial outer membrane. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4076–E4085. <http://dx.doi.org/10.1073/pnas.1415142111>.
- Wickham, H., 2009. *ggplot2: Elegant Graphics for Data Analysis*.
- Woetzel, N., Karakas, M., Staritzbichler, R., Müller, R., Weiner, B.E., Meiler, J., 2012. BCL::score-knowledge based energy potentials for ranking protein models represented by idealized secondary structure elements. *PLoS One* 7, e49242. <http://dx.doi.org/10.1371/journal.pone.0049242>.
- Youle, R.J., Strasser, A., 2008. The BCL-2 protein family: opposing activities that mediate cell death. *Nat. Rev. Mol. Cell Biol.* 9, 47–59. <http://dx.doi.org/10.1038/nrm2308>.