An Efficient Null Model for Conformational Fluctuations in Proteins

Harder, Tim; Borg, Mikael; Bottaro, Sandro; Boomsma, Wouter; Olsson, Simon; Ferkinghoff-Borg, Jesper; Hamelryck, Thomas

Published in:
Structure

DOI:
10.1016/j.str.2012.03.020

2012

Structure
Article

An Efficient Null Model for Conformational Fluctuations in Proteins

Tim Harder,[1] Mikael Borg,[1] Sandro Bottaro,[2] Wouter Boomsma,[2,3] Simon Olsson,[1] Jesper Ferkinghoff-Borg,[2] and Thomas Hamelryck[1,*]
[1]The Bioinformatics Section, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark
[2]DTU Elektro, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[3]Departments of Astronomy and Theoretical Physics, Lund University, SE-223 62 Lund, Sweden
*Correspondence: thamelry@binf.ku.dk
DOI 10.1016/j.str.2012.03.020

SUMMARY

Protein dynamics play a crucial role in function, catalytic activity, and pathogenesis. Consequently, there is great interest in computational methods that probe the conformational fluctuations of a protein. However, molecular dynamics simulations are computationally costly and therefore are often limited to comparatively short timescales. TYPHON is a probabilistic method to explore the conformational space of proteins under the guidance of a sophisticated probabilistic model of local structure and a given set of restraints that represent nonlocal interactions, such as hydrogen bonds or disulfide bridges. The choice of the restraints themselves is heuristic, but the resulting probabilistic model is well-defined and rigorous. Conceptually, TYPHON constitutes a null model of conformational fluctuations under a given set of restraints. We demonstrate that TYPHON can provide information on conformational fluctuations that is in correspondence with experimental measurements. TYPHON provides a flexible, yet computationally efficient, method to explore possible conformational fluctuations in proteins.

INTRODUCTION

Over the past few decades it has become increasingly accepted that proteins are dynamic molecules. Although many proteins adapt unique and specific folds, their inherent flexibility is often essential to the protein's function. However, flexibility can also lead to pathogenesis through misfolding, possibly leading to the formation of aggregates and fibrils (Dobson, 2003; Teilum et al., 2009a).

Computer simulations have emerged as important tools to study the dynamics of proteins, complementing the data obtained from biophysical experiments. A variety of methods are available, ranging from detailed all-atom molecular dynamics (MD) simulations (McCammon et al., 1977; Karplus and McCammon, 2002; Hess et al., 2008) to coarse-grained and approximative methods, such as normal mode analysis (NMA; Levitt et al., 1983), elastic networks (Zheng et al., 2007), tCONCOORD (de Groot et al., 1997; Seeliger and De Groot, 2009), and FRODA (Jacobs et al., 2001; Wells et al., 2005). All methods come with a trade-off between the level of detail and the computational cost for obtaining useful information.

The concept behind MD simulations is to approximate the physical forces acting on a protein and to calculate the motion of particles in the system by applying Newton's laws of motion (McCammon et al., 1977; Karplus and McCammon, 2002; Hess et al., 2008). Because the calculation of these physical forces is computationally expensive, MD simulations are usually limited to short timescales—typically in the range of hundreds of nanoseconds. The high level of detail in MD simulations make general physical conclusions viable (van Gunsteren et al., 1996; Brooks et al., 2009). However, the timescales routinely accessible through MD simulations rarely cover the full dynamic range of proteins. Coarse-grained MD simulations sacrifice certain atomic details to gain a computational advantage, thus allowing longer simulation times or simulations of larger systems. Merging multiple atoms into so-called *beads* or *pseudoatoms* is a common approach to reduce the number of particles in the system (Marrink et al., 2007). Another solution to overcome the computational cost of MD simulations is to use faster computer hardware. Shaw and colleagues were able to achieve a millisecond simulation using custom built special-purpose hardware (Klepeis et al., 2009; Shaw et al., 2010).

Many faster heuristic alternatives to MD have been developed. The idea behind the elastic network (EN) models is that the dynamics of folded, native proteins are rather limited compared to unfolded dynamics and are overall governed by the interresidue contact topology (Bahar and Rader, 2005). Over the past years, the computationally efficient EN models have replaced the original harmonic potentials in many NMA approaches (Bahar and Rader, 2005; Yang et al., 2009). In EN models, the protein's atoms are viewed as point masses that are interconnected by springs. Often, only the backbone C_α atoms are included. Subsequently, a number of conformations are sampled and a principal component analysis is performed on the generated ensemble, yielding the normal modes (Levitt et al., 1983). However, ensembles sampled from EN models can be also used in different scenarios (Zheng et al., 2007); vice versa, normal modes can be also calculated from ensembles generated in MD simulations (Hess et al., 2008).

Other heuristic approaches that include atomic detail have gained popularity over the past years. FRODA (Jacobs et al., 2001; Wells et al., 2005) identifies rigid substructures in the

protein structure to reduce the degrees of freedom for the subsequent simulation. Another widely used heuristic tool is tCONCOORD (de Groot et al., 1997; Seeliger et al., 2007; Seeliger and De Groot, 2009), which has been successfully applied in different contexts (Zachariae et al., 2008; Seeliger and de Groot, 2010). Here, the input structure is analyzed to create a network of constraints. Subsequently, tCONCOORD randomly perturbs the atom coordinates within a box around their initial positions in the native structure. Then, a Monte Carlo procedure changes the perturbed atomic positions until they again satisfy the constraints. In this procedure, the atomic positions are subject to changes sampled from a uniform distribution. Consequently, all the information is encoded in the constraint network; in the absence of constraints, there is no information on how to arrange the atoms.

Here, we present TYPHON, which adopts a probabilistic approach to exploring conformational fluctuations in proteins. TYPHON is based on two recent innovations: TorusDBN (Boomsma et al., 2008) and BASILISK (Harder et al., 2010). TorusDBN and BASILISK are probabilistic models of the conformational space of a protein's main chain and its amino acid side chains, respectively. Both models are formulated as dynamic Bayesian networks (DBNs) and make use of directional statistics (Mardia and Jupp, 2000)—the statistics of angles and directions—to represent protein structure in a natural, continuous space (Hamelryck et al., 2006; Boomsma et al., 2008; Harder et al., 2010). Together, TorusDBN and BASILISK constitute a probabilistic model of protein structure in atomic detail. This model is *generative*; plausible protein conformations can be efficiently sampled. Furthermore, TYPHON incorporates CRISP (Bottaro et al., 2012), an efficient method for applying local modifications to the protein's conformation.

The application of these probabilistic models in TYPHON ensures that the structure remains protein-like on a local length scale throughout the conformational sampling. The long-range structure is maintained by imposing different types of distance-based restraints, which are heuristic representations of nonlocal interactions, such as hydrogen bonds. TYPHON uses Gaussian distributions to implement the restraints, resulting in a valid probabilistic description of the restraint network and the local structure of proteins. This well-justified probabilistic formulation differs from previous ad hoc approaches. TYPHON explores the conformational space accessible to a protein, within the limits imposed by the restraint network. In the absence of a restraint network, sampling is solely guided by the probabilistic models and results in an ensemble of extended conformations with realistic local structure, conceptually reminiscent of an "unfolded state."

In short, TYPHON can be considered a null model of conformational fluctuations, given a set of probabilistic restraints. We again stress that our method is well justified, *given* a chosen set of restraints; the biological relevance of the obtained conformations will necessarily depend on the relevance of the heuristic restraints. However, TYPHON provides default restraints, which typically deliver good results for common applications, as discussed below.

In the following, we compare results obtained from TYPHON with experimental measures describing the native ensemble of folded proteins, including B-factors, nuclear magnetic resonance (NMR) order parameters, and residual dipolar couplings (RDCs). The different measures allow us to investigate how well TYPHON captures the flexibility of a folded protein. We then demonstrate how local unfolding caused by the loss of metal ions is correctly modeled by TYPHON. Finally, we show how fluctuations of local structure can be investigated under the control of the probabilistic models, which is an additional attractive and innovative aspect of our approach.

RESULTS

Overview of TYPHON

TYPHON samples protein structures from a joint probability distribution that includes local and nonlocal interactions (described in more detail in the Experimental Procedures). TYPHON incorporates several sophisticated probabilistic models to maintain the local structure and uses simple Gaussian restraints to maintain relevant nonlocal interactions. Although the choice of these nonlocal restraints is heuristic, the resulting joint probabilistic model is well defined and rigorous. In other words, if a suitable restraint network can be chosen for the problem of interest, TYPHON will typically deliver good results, obtained from a well-defined probability distribution.

By default, TYPHON automatically detects the hydrogen bond network. The geometry of the individual hydrogen bonds is restrained using a simple model based on four distances modeled by Gaussian probability distributions. Disulfide bridges are, by default, treated in a similar way. By default, TYPHON also restrains all distances between C_α atoms that are five or more residues apart in the amino acid chain and within six Å of each other. The latter restraints aim to capture general interactions that stabilize the protein, such as the hydrophobic effect.

The user can manipulate and verify the restraint network. For example, it is possible to disregard all hydrogen bonds involving side chains or to add or remove restraints between arbitrary atom pairs. In this manuscript, we use different restraint networks to answer different questions. These networks range from involving C_α atoms (see Experimental B-Factors) over hydrogen bonds (see Generating a Native Ensemble) to a small number of disulfide bridges (see Local Structure under the Control of Probabilistic Models).

TYPHON is obviously limited with respect to modeling the formation and dissolution of nonlocal interactions themselves, as the restraint network is fixed throughout the sampling procedure. However, the secondary structure can be, to some extent, put under the control of the probabilistic models (see Local Structure under the Control of Probabilistic Models), allowing for formation and dissolution of certain hydrogen bonds, notably, in helices.

Experimental B-Factors

The Protein Data Bank (PDB; Berman et al., 2000) currently contains over 77,000 solved structures; the majority of them are determined by X-ray crystallography. Experimental B-factors associated with the atoms of a crystal structure often give a first indication of the conformational fluctuations within a protein. The B-factor reflects both the thermal vibrations of single atoms and small structural differences between molecules in the crystal. The latter contribution is of interest for inferring protein flexibility.

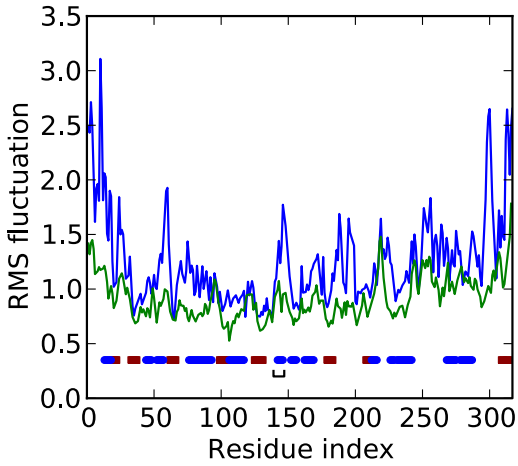

Figure 1. Experimental B-Factors of Candida Antarctica Lipase B
The figure shows root-mean-square fluctuations calculated from the B-factors taken from the crystal structure (PDB: 1tca; green line) and calculated from a TYPHON simulation started from the same crystal structure (blue line). The secondary structure elements are indicated by blue circles for α helices and red squares for β strands. The lid region is indicated by the black bracket.

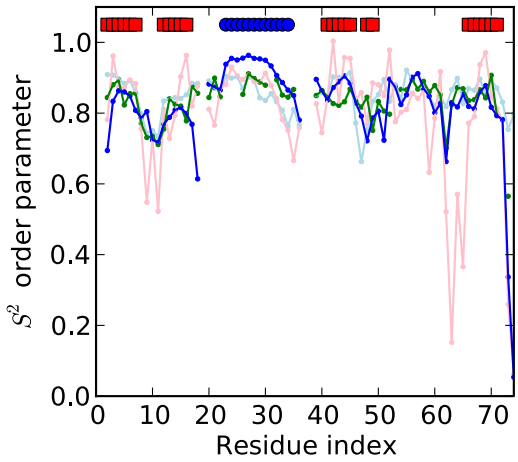

Figure 2. Experimentally Determined S^2 Values (green) versus Values Calculated from a TYPHON Ensemble (blue) for Ubiquitin
The S^2 order parameter is an experimental measure arising from NMR experiments that reflects flexibilities in the protein. It ranges from zero (isotropic motion) to one (no motion). For comparison, the figure also shows S^2 order parameters calculated from an MD simulation (light blue) and a tCONCOORD simulation (light red). The secondary structure is indicated by red squares for β strands and blue circles for α helices. The fragmentation of the lines is due to missing values for Ile23, Glu24, Asn25, Gln31, Ile36, Gly53, Arg72, Arg74, Gly75, and Gly76 in the experimental data and for all proline residues.

In this test, we analyze whether TYPHON is able to reproduce the flexibility that is indicated by the B-factors of a protein.

TYPHON makes it possible to sample an ensemble of structures that is close to the native structure. We illustrate this with the crystal structure of the 317-residue-long protein *Candida antarctica* Lipase B (CalB; PDB: 1tca; Uppenberg et al., 1995). CalB is an enzyme with industrial applications that adopts an α/β fold. A short helix, consisting of residues 139 to 147, is suspected to act as a flexible lid that is important for catalysis, making it a prime subject of dynamics studies (Skjøt et al., 2009). For comparison, we translated the experimental B-factors of the crystal structure into root-mean-square fluctuations (rmsf) using the following relation (Kuzmanic and Zagrovic, 2010):

$$\mathrm{RMSF}_i^2 = \frac{3B_i}{8\pi^2},$$

where B_i is the B-factor for the i-$^{\mathrm{th}}$ residue.

TYPHON used the crystal structure as sole input, from which 581 $C_\alpha \cdots C_\alpha$ Gaussian distance restraints were derived (see the Experimental Procedures). The sampling ran for 50 million iterations. Figure 1 shows RMS fluctuation calculated from the experimental B-factors for the crystal structure and from 1,000 sampled conformations chosen with regular intervals. The overall flexibility along the sequence is well captured. The lid region clearly displays a higher level of flexibility in correspondence with its dynamic nature (Skjøt et al., 2009). The good agreement with the experimental measure is also reflected in the Pearson correlation coefficient, which is equal to 0.71.

Generating a Native Ensemble

Advances in nuclear magnetic resonance (NMR) spectroscopy over the past decades made more detailed studies of dynamics in proteins possible. The S^2 order parameter is a measure arising from NMR experiments describing the amplitude of motion of an N-H vector (Lipari and Szabo, 1982). A backbone segment that is unrestricted in its movement, usually in a region of high flexibility, will have a low S^2 value. For segments in more constrained or rigid regions of the protein, the S^2 value will be higher. Analyzing S^2 order parameters provides a more direct view on the dynamics of a protein compared to the B-factors. In this test, we analyze whether TYPHON is able to capture the fast dynamics of a protein as implied by the S^2 order parameters.

Ubiquitin is a well-studied protein in terms of its dynamics; its relatively small size of 76 amino acids allows for both extensive MD simulations as well as NMR studies. Ubiquitin consists of a five stranded, twisted, and antiparallel β sheet with an α-helix lying across. A number of recent publications discuss the molecular recognition mechanisms using ubiquitin as a model system (Lange et al., 2008; Wlodarski and Zagrovic, 2009; Long and Brüschweiler, 2011).

TYPHON sampling started from a single crystal structure of ubiquitin (PDB: 1ubi; Ramage et al., 1994), with 46 automatically detected hydrogen bonds as restraints, and ran for 50 million iterations. A total of 1,000 structures were sampled in regular intervals. We also generated an ensemble of 1,000 structures using tCONCOORD, starting from the same ubiquitin crystal structure and using default settings. For further comparison, we also included the order parameters calculated from an MD simulation of ubiquitin (Maragakis et al., 2008).

Figure 2 shows the S^2 order parameters calculated from the TYPHON ensemble following Best and Vendruscolo (2004) and order parameters obtained from an experiment by Tjandra et al. (1995). The figure further shows order parameters calculated from a tCONCOORD ensemble obtained with default parameters and from an MD simulation (Maragakis et al., 2008). Overall, the S^2 parameters calculated from the TYPHON

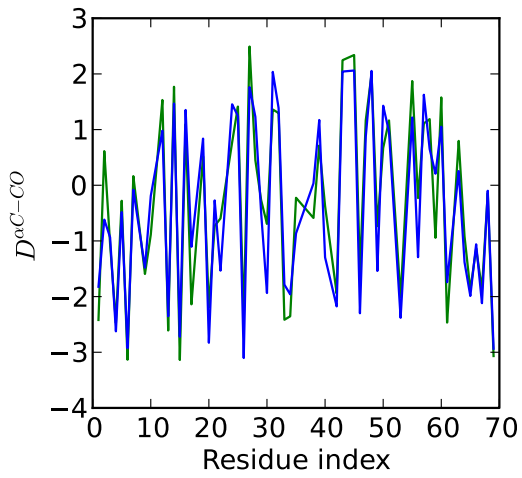

Figure 3. C_α − CO RDC Values for Ubiquitin

The figure shows a comparison between experimentally determined C_α − CO RDCs (green line) and RDCs calculated from a TYPHON ensemble using the procedure described in Showalter and Brüschweiler (2007). (blue line), where the RDCs are plotted on the y axis against the residue index on the x axis. See also Figure S1.

ensemble are in good agreement with the experimental measurements; the correlation coefficient for the two curves is 0.73. The most rigid region is located in the well-ordered α-helix between residues 23 and 33. This region is indeed rigid in the TYPHON ensemble as well though overly so compared to the experimental results (Tjandra et al., 1995). The terminal regions are the most flexible (see Figure 2). Recently, it was found that the increased flexibility in the C-terminus and in loop I between the β1 and β2 strands is of importance for the molecular recognition mechanism of ubiquitin (Lange et al., 2008; Wlodarski and Zagrovic, 2009). The ensemble generated by TYPHON accurately reflects the conformational fluctuations in these regions of interest.

The order parameters calculated from the MD simulation match the experimental values less well; the correlation coefficient is 0.52. Although the MD ensemble accurately reflects the flexibilities in loop I, it does not well reproduce the fluctuations in the C-terminus. The S^2 order parameters calculated from the tCONCOORD ensemble match the general trend of the experimental curve. The correlation coefficient is 0.53, which is also lower than for TYPHON. The generated ensemble appears to overemphasize the flexibility in certain loops, including the functionally important loop I—around residues 7 to 10. In addition, loop V—around residues 63 to 65—shows considerable discrepancy. Leaving out the flexible C-terminal region, following Lindorff-Larsen et al. (2005), results in correlation coefficients equal to 0.50, 0.55, and 0.28 for the MD, TYPHON, and tCONCOORD ensembles, respectively. In conclusion, TYPHON matches the experimentally determined order parameters, indicating that the fast dynamics—as described by the Lipari-Szabo S^2 parameters—are well captured in the generated ensemble.

Residual dipolar couplings (RDCs) probe the bond vector geometry relative to an external magnetic field. Data acquisition in a nematic phase solvent or in the presence of a paramagnetic center can make measurement of RDCs in the solution state possible (Tjandra et al., 1997; Banci et al., 2004). RDCs are anisotropic quantities and thus average out when molecules undergo isotropic rotational diffusion.

For ubiquitin, Cornilescu et al. (1998) obtained six sets of backbone RDCs in a nematic phase solvent based on phospholipid bicelles. The experimental data was obtained from the Biological Magnetic Resonance Data Bank (BMRB entry: 6457; Ulrich et al., 2008). We used the same TYPHON and tCONCOORD ensembles as in the previous section. Ensemble averages were calculated from these ensembles using the procedure described by Showalter and Brüschweiler (2007; Lindorff-Larsen et al., 2005).

Figure 3 shows experimentally determined C_α − CO RDCs in comparison with RDCs that are calculated from a TYPHON ensemble. Figure S1 (available online) additionally shows correlation plots for all RDCs. In general, there is a good correlation between the values obtained from the TYPHON and the experimental data (see Table 1). The agreement with experiment for the TYPHON ensemble is comparable to the tCONCOORD ensemble and the crystal structure (1UBI). However, Q-factors for the TYPHON ensemble (0.37) are larger than for the tCONCOORD ensemble (0.28) and the crystal structure (0.23), suggesting better qualitative agreement of the tCONCOORD ensemble (Lipsitz and Tjandra, 2004).

Although the reproduction of experimental data, such as residual dipolar couplings and order parameters, serves as a sanity check, it is difficult to make a quantitative assessment of the physical timescales sampled (Showalter and Brüschweiler, 2007). However, collectively, the results suggest that TYPHON samples broader ensembles in some regions of ubiquitin as compared to tCONCOORD. Regions that appear overstabilized may be attributed to the employed restraints, suggesting that TYPHON ensembles can be improved by input of expert knowledge. In view of the excellent structural quality of the generated decoys (compare section Quality of the Sampled Structures), these observations support the interpretation of TYPHON as a suitable "null model" of conformational fluctuations in proteins for a given set of restraints; given the nonlocal

Table 1. Statistics for the RDC Values Obtained from the TYPHON and tCONCOORD Ensembles of Ubiquitin

	N − NH	CO −NH	$C_\alpha - H_\alpha$	N − CO	C_α − CO	$C_\alpha - C_\beta$
Correlation coefficient average RDC TYPHON[a]	0.91	0.90	0.92	0.94	0.93	0.90
Correlation coefficient average RDC tCONCOORD[a]	0.96	0.91	0.96	0.96	0.96	0.97
Correlation coefficient Crystal structure (1UBI)[b]	0.98	0.96	0.93	0.99	0.99	0.97

[a]Correlation coefficients of the TYPHON and tCONCOORD ensembles with the experimental data, respectively.
[b]Correlation coefficient between the crystal structure 1UBI and for all six RDC types.

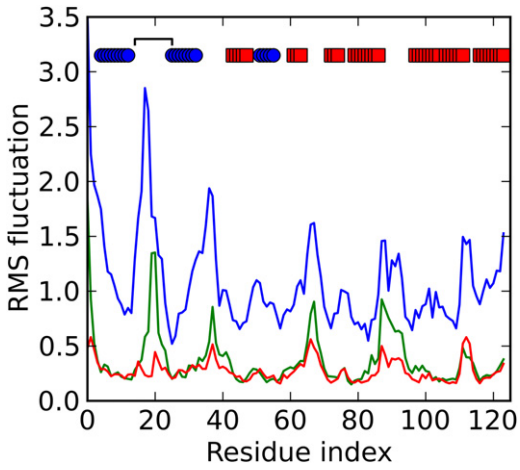

Figure 4. RMS Fluctuations Showing Dynamics of Ribonuclease A
The plot shows the RMS fluctuations measured from a set of PDB structures (see Functional Dynamics of an Enzyme, green line), an ENM analysis (red line), and a TYPHON simulation (blue line). The secondary structure elements are indicated by blue circles for helical residues and red squares for strands. Loop I, including residues 14–25, is indicated by the black bracket.
See also Figure S2, Movie S1, and Table S1.

restraints, the probabilistic models of local structure ensure a thorough exploration of the remaining conformational space.

Functional Dynamics of an Enzyme

Ribonuclease (RNase) A is a pancreatic protein that cleaves single-stranded RNA; its structural dynamics are essential for its enzymatic function (Doucet et al., 2009; Formoso et al., 2010). The protein has 124 residues and adopts an α/β fold that consists of two domains flanking a catalytic site. In this experiment, we analyze whether TYPHON can reproduce the functional dynamics of RNase A. In addition, we compare the TYPHON ensemble to results obtained from NMA.

We initialized TYPHON sampling from the RNase A crystal structure (PDB: 7RSA; Wlodawer et al., 1988) and used the automatically detected hydrogen bond network with default settings, resulting in 76 hydrogen bonds and four disulfide bridges. The sampling was run for 100 million iterations, from which 1,000 structures were retained.

As a measure of the structural flexibility of RNase A, we analyzed 132 experimentally determined structures with a maximum of one point mutation (for a complete list see Table S1). We superimposed the experimental structures using iterative root-mean-square deviation (rmsd) minimization to the average structure and calculated the rmsf of the C_α atoms. We call this set the high-sequence similarity PDB ensemble (Best et al., 2006).

In addition, we compare our result to the dynamics of the enzyme according to the elastic network model (ENM), a coarse-grained model of protein dynamics that has been used to analyze collective motions, residue fluctuations, and conformational changes (Tirion, 1996; Hinsen, 1998; Bahar and Rader, 2005; Ma, 2005; Kimber et al., 2010). In the ENM, the protein structure is approximated as a network of coupled harmonic oscillators between all C_α atoms closer than a specified cutoff radius. The collective motions of the system can be then calculated using NMA. The ENM analysis was performed with the elNémo server and default parameters, using an 8 Å cutoff distance to identify elastic interactions (Suhre and Sanejouand, 2004). The server reports the rmsf calculated from the scaled Eigen vectors of the first hundred modes.

The fluctuations found within the PDB and the TYPHON ensembles (Figure 4) are in good agreement; the correlation coefficient is 0.72. The overall flexibility pattern along the amino acid chain indicates increased mobility in the same regions. The amplitude of the fluctuations is, however, significantly larger for the TYPHON ensemble, indicating that a large volume of the conformational space is sampled. This again confirms the interpretation of TYPHON as a suitable null model of conformational fluctuations for a given set of restraints. Notably, loop I—consisting of residues 14 to 25—has a high degree of flexibility (Figure 4). The dynamics of this loop are especially important for the catalytic activity of the enzyme (Doucet et al., 2009; Formoso et al., 2010). TYPHON sampling started from other crystal structures of RNase A in the PDB yielded similar results (PDB codes 3LXO [Doucet et al., 2010] and 2G8Q [Leonidas et al., 2006]). In contrast, although having only a slightly lower correlation coefficient to the PDB ensemble (0.67), the result from the ENM analysis does not show an elevated flexibility in this loop.

The dynamics of loop I is a requirement for the functional dynamics of RNase A; RNase A has been shown to function through a concerted motion between an open form that can bind substrate and a closed form, where catalysis occurs (Watt et al., 2007). To investigate how the TYPHON ensemble relates to these motions, we performed a principle component analysis on the TYPHON samples and isolated the main modes. The first mode, which contains the most important variations of the ensemble, indeed shows an opening and closing of the catalytic cleft, lending further evidence that the TYPHON ensemble can be used to explore enzyme dynamics. A video of the motion is available online (see Movie S1).

Induced Change in Flexibility

Large-scale motions or major changes in flexibility in proteins are often induced by binding or releasing ligands. These ligands can be as complex as multiatom substrates, inhibitors, or drugs or as simple as single metal ions. In this test we use TYPHON to simulate partial unfolding upon loss of metal ions. This application illustrates how the probabilistic models "step in" to provide information in the absence of restraints.

Cu/Zn superoxide dismutase (SOD1) is a ubiquitous protein in the cytoplasm that is associated with the neurodegenerative disease amyotrophic lateral sclerosis (ALS). ALS results in paralysis and respiratory failure within one to five years from onset (Pasinelli and Brown, 2006). The oligomerization of SOD1 is associated with a gain in toxic function. Experimental evidence suggests that a loss of the two metal ions induces structural changes to the monomeric form of SOD1 and subsequently leads to pathogenic aggregation (Teilum et al., 2009b). However, the exact pathway is still unknown. We used the PDB:2v0a crystal structure as starting point for our experiments (Strange et al., 2007). SOD1 consists of a β barrel with long loops connecting the antiparallel strands. It contains a disulfide bridge and has

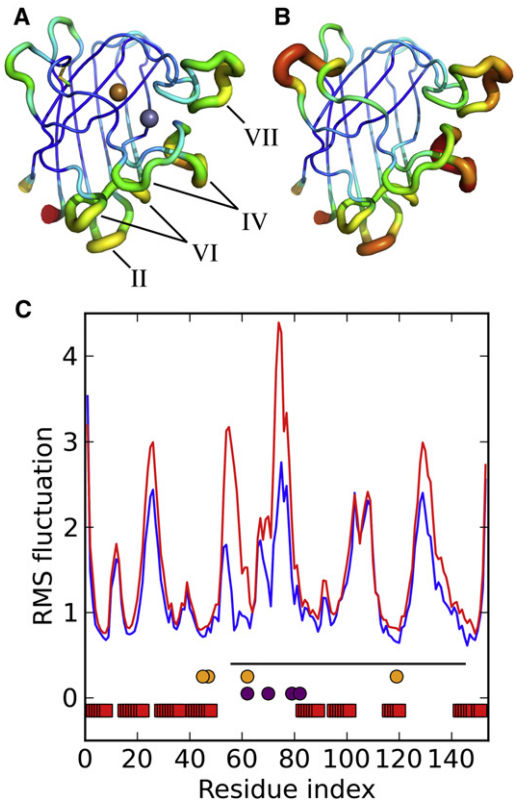

Figure 5. Cu/Zn Superoxide Dismutase

(A) TYPHON ensemble obtained from the native monomer. The Cu and Zn ions are shown as an orange and a purple sphere, respectively. The C57-C146 disulfide bridge is shown as a stick representation. The roman numerals indicate the loop numbers. This ensemble corresponds to the blue line in (C).

(B) TYPHON ensemble obtained without the ions and with the disulfide bridge reduced. This ensemble corresponds to the red line in (C).

(C) Shows the corresponding rmsf curves. The disulfide bridge is indicated as a black line. The residues coordinating the metal ions are marked by orange and purple circles for the copper and zinc ion, respectively.

See also Table S2.

two associated metal ions: a copper ion that is coordinated by four histidines (residues 46, 48, 63, and 120) and a zinc ion that is coordinated by three histidines and an aspartate (residues 63, 71, 80, and 82).

Ding and Dokholyan (2008) performed a discrete MD analysis of the SOD1 monomer. They systematically tested the effect of losing metal ions and/or reducing the disulfide bridge. Each individual event leads to a significant increase in flexibility; the two most affected regions are both located in the long loop IV (Figure 5A). The region around Cys57, which is involved in the disulfide bridge, is primarily affected by the loss of the disulfide bridge. The loss of the metal ions primarily affects the regions adjacent to the ion coordinating histidines. Other parts of the structure seem mostly unaffected by either event. Following this study, we analyzed the mobility of different forms of SOD1, namely, the holo form with the C57-C146 disulfide bridge intact and the apo monomer with the disulfide bridge reduced. Again, we used only a single crystal structure as starting point. We set up two different TYPHON experiments. For the apo experiment, we removed the automatically detected disulfide bridge and did not include any restraints involving the metal ions. For the holo experiment, we added the copper and zinc ions in the form of distance restraints that maintain the mutual distances between the four ion-coordinating atoms (see Table S2) and included the disulfide bridge. The remaining restraint network, consisting of 73 automatically detected hydrogen bonds, was identical in both setups. For each setup we ran three experiments of 100 million iterations each and combined the generated structures for the final evaluation to ensure converged sampling. Note that in the absence of the restraints concerning metal ions and disulfide bridge, the relative influence of the probabilistic models of local structure on the sampled conformations increases.

Figure 5 shows the results of the different experiments in putty representation. The results show that the loss of the metal ions and the reduced disulfide bridge leads to a significant increase in flexibility, especially in the long loop IV between residues 49 and 83 but also in the loops II, VI, and VII. The spike in flexibility around residue 57 can be attributed to the reduced disulfide bridge, which, in the native structure, covalently binds this surface loop. The increased flexibility in other parts of the protein is likely due to the loss of the metal ions. An interesting observation is also the increased flexibility in loop II around residue 25, which is not in direct contact with any of the mutated sites. We speculate that the overall increased mobility in the long loop IV and VI also influenced the flexibility in this region.

The results closely resemble those of Ding and Dokholyan (2008), which were obtained from discrete MD simulations. A TYPHON experiment requires about 20 hours, which would allow scanning of larger sets of clinically known mutations (Andersen et al., 2003). We point out that the increased mobility in loop II was not observed in the MD study of Ding and Dokholyan (2008), which illustrates that TYPHON can deliver results that suggest starting points for new hypotheses or follow-up studies. It should be noted that TYPHON only includes the steric component of the ion loss; changes in electrostatics or solvent accessibility are not directly accounted for. Nonetheless, in this case, modeling the effect of the metal ions as simple Gaussian restraints accurately reproduces the results obtained from much more sophisticated simulations and leads to potentially interesting and new observations.

Local Structure under the Control of Probabilistic Models

The Gaussian restraints obviously do not allow for formation or dissolution of nonlocal interactions; the restraint network is rigorously fixed during the sampling procedure. However, certain nonlocal interactions, such as hydrogen bonds in helices, can be put under the control of the probabilistic models instead. In practice, this means that certain conformational fluctuations of the protein backbone on a local length scale could be investigated. In this application, we explore and illustrate this approach with a small helical protein and investigate helical mobility and $\alpha/3_{10}$-helix transitions.

The Mature T Cell Proliferation Gene 1 (MTCP1) is a known oncogene that is linked to certain types of leukemia (Barthe et al., 2002). The structure of the human $p8^{MTCP1}$ protein has been solved by NMR and consists of three helices.

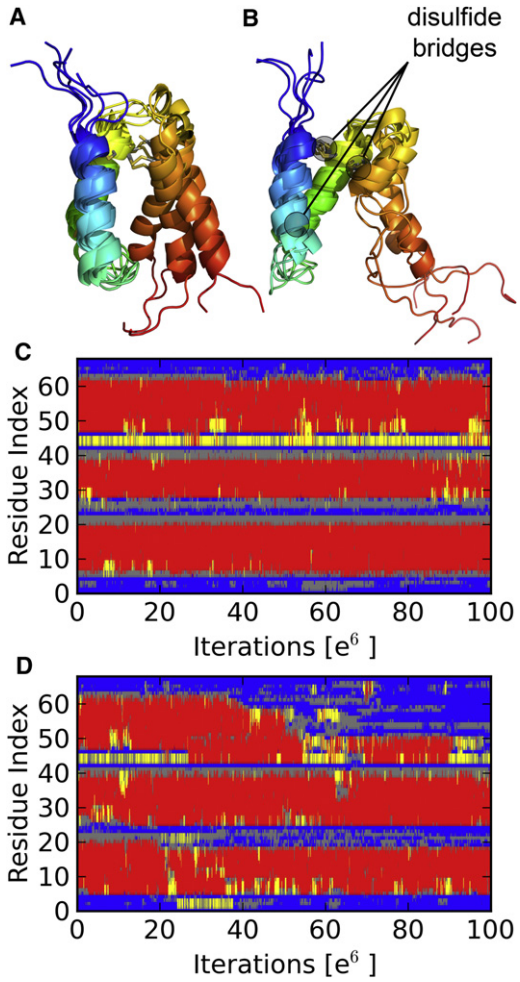

Figure 6. Local Structure under the Control of Probabilistic Models

(A and B) Five representative structures of the simulation (A) with and (B) without fixed secondary structure assignment. The disulfide bridges are shown in stick representation and highlighted in (B).

(C and D) Secondary structure content of the simulation (C) with and (D) without fixed secondary structure input. The secondary structure was measured using DSSP. Color code: red is α-helix; yellow is 3_{10}-helix; gray is β-turn; and blue is random coil.

A stable α hairpin connecting helix I and II is covalently held together by two disulfide bridges between residues 7, 38 and 17, and 28, respectively. A third less restricted and stable helix (helix III) is also connected to helix II with a third disulfide bridge between residues 39 and 50 (Barthe et al., 1997). MD simulations indicate that helix III is fairly flexible with respect to the α hairpin (Barthe et al., 2002).

We first investigate to what extent the helices move with respect to each other. We therefore started from the first model of a $p8^{MTCP1}$ NMR ensemble (PDB: 2hp8; Barthe et al., 1997). The experiment ran for 100 million iterations with the three disulfide bridges as only restraints. However, we also imposed the secondary structure of the native structure according to DSSP (Kabsch and Sander, 1983) through TorusDBN (Boomsma et al., 2008). This is a more flexible and "soft" way to restrain the sampling, as the helical regions are allowed to bend or, to a certain extent, form and dissolve hydrogen bonds under the influence of the probabilistic model.

Despite the absence of restraints, besides those involving the three disulfide bridges, all helices remain stable throughout the sampling. Figure 6A shows five representative structures from the ensemble. Helix I and helix II are tightly fixed by the interhelical disulfide bridges, which only allow limited movements. Helix III is only tethered by a single disulfide bond in the beginning of the helix, which results in higher flexibility. As indicated in Figure 6A, helix III slightly tilts away from the other two helices, a behavior that also has been observed in MD simulations (Barthe et al., 2002).

Figure 6C shows the secondary structure content over the course of the first experiment. The consistent red bars show that all three helices remain fully helical throughout the sampling. In the beginning of helix III, we observe transitions between α- and 3_{10}-helix, which is again in agreement with the results of a MD simulations (Barthe et al., 2002).

In the second experiment, we investigate the stability of the helices themselves. We again included restraints concerning the three native disulfide bonds. However, this time we did not provide any secondary structure information to TorusDBN. In other words, this means that TorusDBN still enforces protein-like conformations but does not require them to be helical.

Again helix I and helix II remain stable throughout the sampling as indicated by the consistent red bars in Figure 6D. This is not surprising because both helices are covalently connected near their respective start and end. The entire protein structure is, however, significantly more flexible, expressed by the movement of the helices with respect to each other (compare Figure 6B). In contrast to helices I and II, helix III quickly unfolds up to residue 50, where it is covalently attached to helix II via a disulfide bridge.

In addition to the unfolding helix III, we observe significant differences compared to the first experiment in the loop regions. In particular, for loop II, which connects helix II and III and stretches from residue 39 to 47, we observe a transition to an α-helix. The terminal 18 residues of helix III readily unfold (see Figure 6D), which points to a difference in stability between the first two and the third helix.

This experiment strikingly demonstrates the possibilities of probabilistic models. In the first experiment, which includes the disulfide bridges and secondary structure information, we observed specific movements of the helices with respect to each other and transitions from an α- to a 3_{10}-helix in the beginning of helix III. Both observations concur with the results obtained from MD simulations (Barthe et al., 2002). In the second experiment, which includes the disulfide bridges but not the secondary structure information, we obtained some information on the relative stability of the helices themselves. Helices I and II remain stable, whereas helix III readily unfolds. Again, this difference in stability is in accordance with MD simulations (Barthe et al., 2002).

Quality of the Sampled Structures

To evaluate the quality of the structures, we analyzed 50 random structures from an RNase A ensemble, generated as described previously, using PROCHECK (Laskowski et al., 1993). For comparison, we generated 50 tCONCOORD (Seeliger et al., 2007) samples for the same protein (starting from PDB: 7rsa). The

Table 2. Quality Assessment of Structures Generated by TYPHON

Residues in Regions	tCONCOORD	TYPHON
Most favored (%)	69.8	88.1
Additionally allowed (%)	26.3	10.8
Generously allowed (%)	3.1	0.7
Disallowed (%)	0.7	0.5
ϕ/ψ G factor	−0.93	−0.24
χ_1 G factor	−0.26	0.10
Overall G factor	−0.69	−0.13

The table lists the results of a PROCHECK analysis of a set of TYPHON and tCONCOORD samples. Well-refined structures usually have 90% or more of all residues in the most favored regions. The G factor is a log odds score; higher numerical values denote higher quality.
See also Documents S2 and S3.

detailed PROCHECK reports can be accessed as Documents S2 and S3.

The Ramachandran map divides the main chain's conformational space, as parameterized by the φ and ψ angles, in different regions, some sterically more favorable than others (Ramachandran et al., 1963). Well-refined protein structures are expected to have 90% or more of the backbone dihedral angles in the most favorable regions. The PROCHECK analysis indicates that the TYPHON samples are of good quality; over 88% of all angles are in the most favored regions. In contrast, the tCONCOORD samples have less than 70% of the backbone angles in these favored regions (Table 2).

PROCHECK's G factor is a measure of how well the analyzed structures match the observed distributions of bond lengths, bond angles, and dihedral angles in crystal structures and is expected to be −0.5 or higher for well-refined structures. Also in this respect, TYPHON samples have a higher quality than do tCONCOORD samples; the G-factor is −0.13 versus −0.69. The G factor takes the side-chain quality into account; in this respect, TYPHON undoubtedly benefits from the detailed side-chain modeling in BASILISK (Harder et al., 2010).

Additionally, we performed a WHATIF (Vriend, 1990) packing analysis of the TYPHON and tCONCOORD ensembles of RNase A. The structures generated by TYPHON have an average packing environment score of −1.495. Those generated by tCONCOORD have an average score of −1.944. As well-refined structures have a score around −0.5, both methods might be improved in this respect.

Computational Efficiency

TYPHON is computationally efficient. The ubiquitin experiments used in this study were performed on a regular desktop computer (Intel Core i7, 2.8GHz) and ran for around 10 hr on a single CPU core. The human $p8^{MTCP1}$ protein experiments comprising 100 million iterations ran for about 15 hr. Naturally, the runtime increases as the number of restraints in the network grows, though extensive caching in the calculations minimizes this effect to an extent. With increasing protein size, more iterations will be necessary to achieve a comparable level of convergence. Although a parallelization of a single run onto multiple cores is not possible in the current implementation, it is possible to perform several TYPHON experiments in parallel to obtain better statistics.

DISCUSSION

In this paper we present TYPHON, a probabilistic approach to explore conformational fluctuations in proteins. TYPHON incorporates detailed probabilistic models of the conformational space of a protein's main chain and its amino acid side chains (Boomsma et al., 2008; Harder et al., 2010) and an efficient local backbone resampling algorithm (Bottaro et al., 2012). During sampling by TYPHON, the conformational space is restricted by a set of restraints imposed on the structure. These restraints typically concern nonlocal interactions, such as hydrogen bonds, disulfide bridges, or interactions with metal ions. The protein structure on a local length scale, including main chain and side chains, is controlled by the probabilistic models.

In this study, we show that TYPHON is able to generate structural ensembles that closely resemble native ensembles described by experimental measures. This includes fluctuations as measured by S^2 order parameters, as well as measured by RDC values. The RNase A study shows not only that TYPHON captures the functional dynamics in the correct regions but also that a principal component analysis of the results is feasible to identify large-scale motions. The analysis of the superoxide dismutase results shows that TYPHON can be used to model effects due to the gain or loss of a ligand, including partial unfolding.

Its computational efficiency makes TYPHON a promising tool for larger screening efforts, for example, of known mutations with clinical relevance. Another interesting application lies in generating suitable candidate structure for docking experiments, allowing for some degree of flexibility in the binding pocket (Henzler and Rarey, 2010). The high quality of the generated structures indicates that no irrelevant parts of the conformational space are explored. On the other hand, TYPHON thoroughly samples the relevant conformational space.

The results of the human $p8^{MTCP1}$ protein experiments demonstrate another strength of our approach. With only a minimal set of restraints defined for the system, the effect of the probabilistic models becomes obvious. They control the local structure and maintain the overall secondary structure, while still allowing for significant conformational fluctuations. It should be noted that it is also possible to run TYPHON without explicitly defining the secondary structure, leading to significantly broader sampling.

In the current implementation, TYPHON keeps the constraint network fixed during the sampling. As a next step, it would be advantageous to allow more flexibility in the restraint network, such as the dissolution or formation of arbitrary hydrogen bonds as the sampling progresses. However, this will require the development of a suitable probabilistic model of nonlocal interactions in proteins and its seamless combination with the probabilistic models of local structure. Fortunately, important theoretical progress was recently made in this respect (Hamelryck et al., 2010). Another interesting addition would be to include information from experimental data (Olsson et al., 2011).

Availability

TYPHON is available as part of the Phaistos package and can be obtained freely from SourceForge under the GNU public license

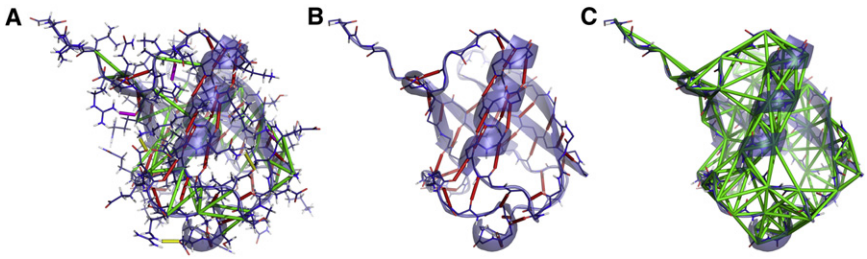

Figure 7. Restraint Network
Depicted are three different calculated networks for ubiquitin (PDB 1ubi).
(A) A network that includes all hydrogen bond types (red: backbone hydrogen bonds; purple: backbone-side-chain hydrogen bonds; and yellow: side-chain-side-chain hydrogen bonds), as well as C_α contacts (green).
(B) A network that includes only the backbone hydrogen bonds.
(C) A network that only includes C_α contacts. The cutoff was 7 Å. The minimum sequence separation between the residues in the chain was two.
See also Figure S3.

(http://sourceforge.net/projects/phaistos/). Currently the Phaistos package is limited to single chain proteins. However, support for multiple chains will be added in the next release.

EXPERIMENTAL PROCEDURES

Overview

The TYPHON network calculation starts from a full atom protein structure, including all of the hydrogen atoms. A restraint network is either loaded from an input file or created in accordance with the protocol described in the following section. In the course of the sampling, the dihedral angles in both the main chain and the side chains are modified under the control of TorusDBN (Boomsma et al., 2008) and BASILISK (Harder et al., 2010), respectively. An efficient local moves method makes subtle movements of the protein backbone possible (Bottaro et al., 2012) and also affects the bond angles in the backbone (see Protein Backbone Move).

Restraint Network Calculation

TYPHON currently supports three classes of restraints involving hydrogen bonds, disulfide bridges, and distance restraints between arbitrary atoms. In the absence of any user input, the program suggests a network using default parameters, which are described in the following paragraphs. This default network is mainly based on biologically relevant restraints, such as hydrogen bonds and disulfide bridges. To stabilize parts of the protein that are naturally stabilized by effects that are not modeled explicitly, TYPHON also connects residues that are far apart in the amino acid sequence but close in space. The user can edit the generated network by adding, removing, or modifying restraints between arbitrary atoms.

We evaluate all potential hydrogen bonds using the DSSP hydrogen bond energy (Kabsch and Sander, 1983). Following Kabsch and colleagues, we discard all candidates with a DSSP energy higher than $-0.5(\mathrm{kcal/mol})$. If an atom has multiple potential hydrogen bonding partners, only the one with the lowest energy is retained. Following the general idea of the DSSP hydrogen bond energy, the hydrogen bond geometry is modeled using four distances. For backbone-backbone hydrogen bonds, these respective distances are explained in more detail in Figure S3. For hydrogen bonds involving side chains, the corresponding standard hydrogen bond acceptors and donors are used; asparagine, aspartate, glutamine, and glutamate can act as hydrogen bond acceptors; arginine, asparagine, glutamine, histidine, lysine, serine, threonine, tryptophan, and tyrosine can act as hydrogen bond donors.

Disulfide bridges are required to have a $S_\gamma \cdots S_\gamma$ distance of 3 Å or less. Similar to hydrogen bonds, the geometry of the disulfide bond is also modeled by four distances consisting of the $S_\gamma \cdots S_\gamma$, $C_\beta \cdots S_\gamma$, $S_\gamma \cdots C_\beta$, and $C_\beta \cdots C_\beta$ distances.

The last class of restraints that are detected by default connects residues that are far apart in the amino acid sequence but close together in space. These restraints stabilize parts of the protein that are naturally stabilized by effects not accounted for explicitly in TYPHON, such as hydrophobic interactions. Residue pairs that are five or more residues apart in the sequence but within six Å ($C_\alpha \cdots C_\alpha$ distance) are modeled by a Gaussian probability distribution on the distance between the two C_α atoms. The distance in the input structure is used as mean μ. The variance σ^2 is set proportional to the square of the distance:

$$\sigma^2 = \left(\frac{\mu}{6}\right)^2.$$

This value was chosen by trial-and-error and produces reasonable results. It allows for more flexibility with increasing distance.

The automatically detected restraints will not always yield the best results, especially when modeling large-scale movements. To keep the framework flexible and utilize the expert knowledge of the researcher, TYPHON allows modifying the restraints and adding distance restraints between arbitrary atom pairs in the structure. In that way, the researcher may additionally stabilize certain parts of the structure or allow more flexibility in other parts. It is also possible to remove automatically detected restraints, for example, when a certain hydrogen bond is known to be weak.

To further simplify this process, TYPHON can generate a PyMOL (Schrödinger, 2010) script that visualizes the restraints. This makes it possible to quickly detect regions that need manual, expert interaction. Figure 7 shows different restraint networks visualized using the generated PyMOL script.

Unstable Hydrogen Bonds

Hydrogen bonds that are in direct contact with solvent molecules are known to be significantly less stable than those that are well shielded. Fernandez and colleagues (Fernández and Berry, 2002; Fernández et al., 2002; Fernández and Scott, 2003; Fernández, 2010) proposed the concept of *dehydrons*, insufficiently shielded hydrogen bonds that are more likely to break. They showed that the number of the carbonaceous groups, CH_n, in a shell around the hydrogen bond is a good estimate of water accessibility. tCONCOORD incorporates this convenient measure to judge the stability of a hydrogen bond (Seeliger et al., 2007). We extended their approach, which was only formulated for backbone hydrogen bonds, to apply to hydrogen bonds involving side chains as well. We therefore moved the centers of the two spheres composing the dehydration shell to the donor nitrogen and the acceptor carbon atoms (Fernández and Berry, 2002; Fernández et al., 2002; Fernández and Scott, 2003; Fernández, 2010). We recalibrated the measure using counts of carbonaceous groups derived from a set of high-resolution crystal structures previously used as training data for BASILISK (Harder et al., 2010). Following Fernandez and colleagues (Fernández and Berry, 2002; Fernández et al., 2002; Fernández and Scott, 2003; Fernández, 2010), we defined the threshold between weak and strong hydrogen bonds as the 4% percentile of the counts. This resulted in thresholds equal to 14, 9, and 7 for backbone-backbone, backbone-side-chain, and side-chain-side-chain hydrogen bonds, respectively. All weak hydrogen bonds are removed from the restraint network by default.

Protein Backbone Move

TYPHON sampling is usually started from the native state of a protein, that is, from a densely packed, compact structure. To capture the subtle movements and flexibilities in compact proteins, it is important to propose local updates of the backbone conformation. A *local move* only affects a limited part of the protein backbone—such as a stretch of five residues—whereas the rest of the protein remains unchanged.

In TYPHON, we use a recently developed type of local move called CRISP (Bottaro et al., 2012). Similar to other methods (Go and Scheraga, 1970; Dodd et al., 1993; Hoffmann and Knapp, 1996; Ulmschneider and Jorgensen, 2003), a local move consists of a concerted rotation of the bond and dihedral angles of the backbone atoms of neighboring residues. Each move involves four elementary steps:

(1) Choose a random stretch in the protein chain.
(2) *Prerotation*: Propose a set of bond and dihedral angle variations in the first $N - 6$ degrees of freedom.
(3) *Postrotation*: Calculate the six remaining degrees of freedom such that the loop closes.
(4) Calculate the bias introduced by performing such a nonrandom modification of the chain. The bias calculation is important when the method is used in a Markov chain Monte Carlo sampling scheme to ensure detailed balance.

This geometrical problem is tackled in an original manner. We derived an analytical solution for the postrotational step, thus avoiding the tedious numerical solution of a system of six equations for the six unknown degrees of freedom. The analytical solution is used to derive an efficient strategy to draw tentative updates of the chain. This scheme makes it possible to continuously control the angular variations of all degrees of freedom involved. The CRISP method thus improves on previous concerted-rotation methods in which, to satisfy all geometrical restraints, tentative updates of the chain are often radically different from the original structure or introduce a suboptimal local structure.

Protein Side-Chain Move

To propose a new side-chain conformation, we use our previously developed probabilistic model of side-chain conformational space, BASILISK (Harder et al., 2010). BASILISK is a dynamic Bayesian network that makes it possible to sample side-chain conformations for all relevant amino acids in continuous space. By default, TYPHON resamples a single, randomly picked residue at a time, proposing an entirely new set of χ angles for the side chain. Both the bond length and the bond angles remain unchanged. To have a roughly equal amount of accepted changes affecting side chains and backbone, TYPHON on average resamples five side chains for every backbone move because a local move affects five backbone residues.

Sampling Strategy and Scoring Functions

For sampling, we use a classic Markov chain Monte Carlo (MCMC) approach. According to the Metropolis-Hastings (Metropolis et al., 1953; Bishop, 2006) sampling scheme, a proposed X' structure is accepted with the following likelihood:

$$P_{\mathrm{acc}}(X \rightarrow X') = \min\left(1, \frac{P(X')Q(X' \rightarrow X)}{P(X)Q(X \rightarrow X')}\right),$$

where $P_{\mathrm{acc}}(X \rightarrow X')$ is the probability of accepting to move from structure X to structure X'; $P(X)$ and $P(X')$ are the probabilities of X and X', respectively; $Q(X \rightarrow X')$ and $Q(X' \rightarrow X)$ are the probabilities of proposing to move from X to X' and from X' and X, respectively. $P(X)$ is defined as

$$P(X) \propto P_{\mathrm{R}}(\mathrm{R}) P_{\mathrm{T}}(\mathrm{T}|A) P_{\mathrm{B}}(\mathrm{B}|A) \Delta(X),$$

where A is the amino acid sequence; $P_{\mathrm{R}}(\mathrm{R})$ is the probability density of the restraint network R, consisting of the product of the probability densities of the individual Gaussian restraints; $P_{\mathrm{T}}(\mathrm{T}|A)$ is the density of the backbone angles T according to TorusDBN; $P_{\mathrm{B}}(\mathrm{B}|A)$ is the probability density of the side-chain angles B according to BASILISK; and $\Delta(X)$ is a clash term that is either one or zero. This simple clash function is introduced to avoid close contacts between atoms. We reject every structure with one or more atom pairs below a specific distance cutoff. The exact cutoff distance depends on the atoms involved: 1.5 Å for a hydrogen atom and any other atom; 1.8 Å for S_γ atoms, to allow disulfide bridges; and 2.3 Å for any other atom pair.

The proposal distributions consist of resampling of side-chain conformations using BASILISK (Harder et al., 2010) or local moves using CRISP (Bottaro et al., 2012). To facilitate smooth local perturbations of the backbone chain, CRISP allows for small variations of the backbone bond angles. Each angle is modeled by an atom specific Gaussian distribution with parameters chosen in accordance with the bond-angle term of the OPLS-AA force field (Jorgensen et al., 1996; Kaminski et al., 2001).

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, two tables, one movie, and two documents and can be found with this article online at doi:10.1016/j.str.2012.03.020.

ACKNOWLEDGMENTS

The authors thank Francesco Carbone for help with the p8$^{\mathrm{MTCP1}}$ study. We thank Kaare Teilum (University of Copenhagen, Copenhagen, Denmark), Kresten Lindorff-Larsen (University of Copenhagen), Thomas Poulsen (Novozymes, Bagsvaerd, Denmark), and Leonardo De Maria (Novozymes) for valuable comments and suggestions. T.H. and M.B. are funded by the Danish Council for Strategic Research (NABIIT, 2106-06-0009). W.B. and S.O. are funded by the Danish Council for Independent Research (FNU, 272-08-0315, and FTP, 274-09-0184, respectively). S.B. acknowledges funding from Radiometer, DTU Elektro.

Received: June 27, 2011
Revised: March 8, 2012
Accepted: March 12, 2012
Published online: May 10, 2012

REFERENCES

Andersen, P.M., Sims, K.B., Xin, W.W., Kiely, R., O'Neill, G., Ravits, J., Pioro, E., Harati, Y., Brower, R.D., Levine, J.S., et al. (2003). Sixteen novel mutations in the Cu/Zn superoxide dismutase gene in amyotrophic lateral sclerosis: a decade of discoveries, defects and disputes. Amyotroph. Lateral Scler. Other Motor Neuron Disord. *4*, 62–73.

Bahar, I., and Rader, A.J. (2005). Coarse-grained normal mode analysis in structural biology. Curr. Opin. Struct. Biol. *15*, 586–592.

Banci, L., Bertini, I., Cavallaro, G., Giachetti, A., Luchinat, C., and Parigi, G. (2004). Paramagnetism-based restraints for Xplor-NIH. J. Biomol. NMR *28*, 249–261.

Barthe, P., Yang, Y.S., Chiche, L., Hoh, F., Strub, M.P., Guignard, L., Soulier, J., Stern, M.H., van Tilbeurgh, H., Lhoste, J.M., and Roumestand, C. (1997). Solution structure of human p8$^{\mathrm{MTCP1}}$, a cysteine-rich protein encoded by the MTCP1 oncogene, reveals a new α-helical assembly motif. J. Mol. Biol. *274*, 801–815.

Barthe, P., Roumestand, C., Déméné, H., and Chiche, L. (2002). Helix motion in protein C12A-p8(MTCP1): comparison of molecular dynamics simulations and multifield NMR relaxation data. J. Comput. Chem. *23*, 1577–1586.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The protein data bank. Nucleic Acids Res. *28*, 235–242.

Best, R.B., and Vendruscolo, M. (2004). Determination of protein structures consistent with NMR order parameters. J. Am. Chem. Soc. *126*, 8090–8091.

Best, R.B., Lindorff-Larsen, K., DePristo, M.A., and Vendruscolo, M. (2006). Relation between native ensembles and experimental structures of proteins. Proc. Natl. Acad. Sci. USA *103*, 10901–10906.

Bishop, C.M. (2006). Pattern Recognition and Machine Learning (New York: Springer).

Boomsma, W., Mardia, K.V., Taylor, C.C., Ferkinghoff-Borg, J., Krogh, A., and Hamelryck, T. (2008). A generative, probabilistic model of local protein structure. Proc. Natl. Acad. Sci. USA *105*, 8932–8937.

Bottaro, S., Boomsma, W., Johansson, K.E., Andreetta, C., Hamelryck, T., and Ferkinghoff-Borg, J. (2012). Subtle Monte Carlo updates in dense molecular systems. J. Chem. Theory Comput. *8*, 695–702.

Brooks, B.R., Brooks, C.L., 3rd, Mackerell, A.D., Jr., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., et al. (2009).

CHARMM: the biomolecular simulation program. J. Comput. Chem. *30*, 1545–1614.

Cornilescu, G., Marquardt, J.L., Ottiger, M., and 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.

de Groot, B.L., van Aalten, D.M., Scheek, R.M., Amadei, A., Vriend, G., and Berendsen, H.J. (1997). Prediction of protein conformational freedom from distance constraints. Proteins *29*, 240–251.

Ding, F., and Dokholyan, N.V. (2008). Dynamical roles of metal ions and the disulfide bond in Cu, Zn superoxide dismutase folding and aggregation. Proc. Natl. Acad. Sci. USA *105*, 19696–19701.

Dobson, C.M. (2003). Protein folding and misfolding. Nature *426*, 884–890.

Dodd, L., Boone, T., and Theodorou, D.N. (1993). A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses. Mol. Phys. *78*, 961–996.

Doucet, N., Watt, E.D., and Loria, J.P. (2009). The flexibility of a distant loop modulates active site motion and product release in ribonuclease A. Biochemistry *48*, 7160–7168.

Doucet, N., Jayasundera, T.B., Simonović, M., and Loria, J.P. (2010). The crystal structure of ribonuclease A in complex with thymidine-3′-monophosphate provides further insight into ligand binding. Proteins *78*, 2459–2468.

Fernández, A. (2010). Transformative Concepts for Drug Design: Target Wrapping (Heidelberg: Springer).

Fernández, A., and Berry, R.S. (2002). Extent of hydrogen-bond protection in folded proteins: a constraint on packing architectures. Biophys. J. *83*, 2475–2481.

Fernández, A., and Scott, R. (2003). Dehydron: a structurally encoded signal for protein interaction. Biophys. J. *85*, 1914–1928.

Fernández, A., Sosnick, T.R., and Colubri, A. (2002). Dynamics of hydrogen bond desolvation in protein folding. J. Mol. Biol. *321*, 659–675.

Formoso, E., Matxain, J.M., Lopez, X., and York, D.M. (2010). Molecular dynamics simulation of bovine pancreatic ribonuclease A-CpA and transition state-like complexes. J. Phys. Chem. B *114*, 7371–7382.

Go, N., and Scheraga, H.A. (1970). Ring closure and local conformational deformations of chain molecules. Macromolecules *3*, 178–187.

Hamelryck, T., Kent, J.T., and Krogh, A. (2006). Sampling realistic protein conformations using local structural bias. PLoS Comput. Biol. *2*, e131.

Hamelryck, T., Borg, M., Paluszewski, M., Paulsen, J., Frellsen, J., Andreetta, C., Boomsma, W., Bottaro, S., and Ferkinghoff-Borg, J. (2010). Potentials of mean force for protein structure prediction vindicated, formalized and generalized. PLoS ONE *5*, e13714.

Harder, T., Boomsma, W., Paluszewski, M., Frellsen, J., Johansson, K.E., and Hamelryck, T. (2010). Beyond rotamers: a generative, probabilistic model of side chains in proteins. BMC Bioinformatics *11*, 306.

Henzler, A.M., and Rarey, M. (2010). In pursuit of fully flexible protein-ligand docking: modeling the bilateral mechanism of binding. Mol. Inform. *29*, 164–173.

Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. (2008). Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput. *4*, 435–447.

Hinsen, K. (1998). Analysis of domain motions by approximate normal mode calculations. Proteins *33*, 417–429.

Hoffmann, D., and Knapp, E.W. (1996). Protein dynamics with off-lattice Monte Carlo moves. Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics *53*, 4221–4224.

Jacobs, D.J., Rader, A.J., Kuhn, L.A., and Thorpe, M.F. (2001). Protein flexibility predictions using graph theory. Proteins *44*, 150–165.

Jorgensen, W.L., Maxwell, D.S., and Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. *118*, 11225–11236.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577–2637.

Kaminski, G.A., Friesner, R.A., Tirado-Rives, J., and Jorgensen, W.L. (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J. Phys. Chem. B *105*, 6474–6487.

Karplus, M., and McCammon, J.A. (2002). Molecular dynamics simulations of biomolecules. Nat. Struct. Biol. *9*, 646–652.

Kimber, M.S., Yu, A.Y.H., Borg, M., Leung, E., Chan, H.S., and Houry, W.A. (2010). Structural and theoretical studies indicate that the cylindrical protease ClpP samples extended and compact conformations. Structure *18*, 798–808.

Klepeis, J.L., Lindorff-Larsen, K., Dror, R.O., and Shaw, D.E. (2009). Long-timescale molecular dynamics simulations of protein structure and function. Curr. Opin. Struct. Biol. *19*, 120–127.

Kuzmanic, A., and Zagrovic, B. (2010). Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. Biophys. J. *98*, 861–871.

Lange, O.F., Lakomek, N.A., Farès, C., Schröder, G.F., Walter, K.F., Becker, S., Meiler, J., Grubmüller, H., Griesinger, C., and de Groot, B.L. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science *320*, 1471–1475.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thorton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. *26*, 283–291.

Leonidas, D.D., Maiti, T.K., Samanta, A., Dasgupta, S., Pathak, T., Zographos, S.E., and Oikonomakos, N.G. (2006). The binding of 3′-N-piperidine-4-carboxyl-3′-deoxy-ara-uridine to ribonuclease A in the crystal. Bioorg. Med. Chem. *14*, 6055–6064.

Levitt, M., Sander, C., and Stern, P.S. (1983). The normal modes of a protein: native bovine pancreatic trypsin inhibitor. Int. J. Quantum Chem. *24* (*Suppl 10*), 181–199.

Lindorff-Larsen, K., Best, R.B., Depristo, M.A., Dobson, C.M., and Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. Nature *433*, 128–132.

Lipari, G., and Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. J. Am. Chem. Soc. *104*, 4559–4570.

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

Long, D., and Brüschweiler, R. (2011). In silico elucidation of the recognition dynamics of ubiquitin. PLoS Comput. Biol. *7*, e1002035.

Ma, J. (2005). Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure *13*, 373–380.

Maragakis, P., Lindorff-Larsen, K., Eastwood, M.P., Dror, R.O., Klepeis, J.L., Arkin, I.T., Jensen, M.Ø., Xu, H., Trbovic, N., Friesner, R.A., et al. (2008). Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. J. Phys. Chem. B *112*, 6155–6158.

Mardia, K.V., and Jupp, P.E. (2000). Directional Statistics (New York: John Wiley and Sons).

Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., and de Vries, A.H. (2007). The MARTINI force field: coarse grained model for biomolecular simulations. J. Phys. Chem. B *111*, 7812–7824.

McCammon, J.A., Gelin, B.R., and Karplus, M. (1977). Dynamics of folded proteins. Nature *267*, 585–590.

Metropolis, N., Rosenbluth, A.W., et al. (1953). Equations of state calculations by fast computing machines. J. Chem. Phys. *21*, 1087–1092.

Olsson, S., Boomsma, W., Frellsen, J., Bottaro, S., Harder, T., Ferkinghoff-Borg, J., and Hamelryck, T. (2011). Generative probabilistic models extend the scope of inferential structure determination. J. Magn. Reson. *213*, 182–186.

Pasinelli, P., and Brown, R.H. (2006). Molecular biology of amyotrophic lateral sclerosis: insights from genetics. Nat. Rev. Neurosci. *7*, 710–723.

Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. J. Mol. Biol. *7*, 95–99.

Ramage, R., Green, J., Muir, T.W., Ogunjobi, O.M., Love, S., and Shaw, K. (1994). Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. Biochem. J. *299*, 151–158.

Schrödinger, L. (2010). The PyMOL molecular graphics system (version 1 1.3r1), Schrödinger, LLC.

Seeliger, D., and De Groot, B.L. (2009). tCONCOORD-GUI: visually supported conformational sampling of bioactive molecules. J. Comput. Chem. *30*, 1160–1166.

Seeliger, D., and de Groot, B.L. (2010). Conformational transitions upon ligand binding: holo-structure prediction from apo conformations. PLoS Comput. Biol. *6*, e1000634.

Seeliger, D., Haas, J., and de Groot, B.L. (2007). Geometry-based sampling of conformational transitions in proteins. Structure *15*, 1482–1492.

Shaw, D.E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R.O., Eastwood, M.P., Bank, J.A., Jumper, J.M., Salmon, J.K., Shan, Y., and Wriggers, W. (2010). Atomic-level characterization of the structural dynamics of proteins. Science *330*, 341–346.

Showalter, S.A., and Brüschweiler, R. (2007). Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints. J. Am. Chem. Soc. *129*, 4158–4159.

Skjøt, M., De Maria, L., Chatterjee, R., Svendsen, A., Patkar, S.A., Ostergaard, P.R., and Brask, J. (2009). Understanding the plasticity of the α/β hydrolase fold: Lid swapping on the *Candida antartica* Lipase B results in chimeras with interesting biocatalytical properties. Chembiochem *10*, 520–527.

Strange, R.W., Yong, C.W., Smith, W., and Hasnain, S.S. (2007). Molecular dynamics using atomic-resolution structure reveal structural fluctuations that may lead to polymerization of human Cu-Zn superoxide dismutase. Proc. Natl. Acad. Sci. USA *104*, 10040–10044.

Suhre, K., and Sanejouand, Y.H. (2004). *ElNémo*: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res. *32* (*Suppl 2*), W610–W614.

Teilum, K., Olsen, J.G., and Kragelund, B.B. (2009a). Functional aspects of protein flexibility. Cell. Mol. Life Sci. *66*, 2231–2247.

Teilum, K., Smith, M.H., Schulz, E., Christensen, L.C., Solomentsev, G., Oliveberg, M., and Akke, M. (2009b). Transient structural distortion of metal-free Cu/Zn superoxide dismutase triggers aberrant oligomerization. Proc. Natl. Acad. Sci. USA *106*, 18273–18278.

Tirion, M.M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys. Rev. Lett. *77*, 1905–1908.

Tjandra, N., Feller, S.E., Pastor, R.W., and Bax, A. (1995). Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation. J. Am. Chem. Soc. *117*, 12562–12566.

Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M., and Bax, A. (1997). Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat. Struct. Biol. *4*, 732–738.

Ulmschneider, J., and Jorgensen, W. (2003). Monte Carlo backbone sampling for polypeptides with variable bond angles and dihedral angles using concerted rotations and a Gaussian bias. J. Chem. Phys. *118*, 4261–4272.

Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., et al. (2008). BioMagResBank. Nucleic Acids Res. *36* (Database issue), D402–D408.

Uppenberg, J., Ohrner, N., Norin, M., Hult, K., Kleywegt, G.J., Patkar, S., Waagen, V., Anthonsen, T., and Jones, T.A. (1995). Crystallographic and molecular-modeling studies of lipase B from Candida antarctica reveal a stereospecificity pocket for secondary alcohols. Biochemistry *34*, 16838–16851.

van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P., and Tironi, I.G. (1996). Biomolecular Simulation: The GROMOS96 Manual and User Guide (Zürich, Switzerland: Verlag der Fachvereine Hochschulverlag AG an der ETH Zurich).

Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. J. Mol. Graph. *8*, 52–56, 29.

Watt, E.D., Shimada, H., Kovrigin, E.L., and Loria, J.P. (2007). The mechanism of rate-limiting motions in enzyme function. Proc. Natl. Acad. Sci. USA *104*, 11981–11986.

Wells, S., Menor, S., Hespenheide, B., and Thorpe, M.F. (2005). Constrained geometric simulation of diffusive motion in proteins. Phys. Biol. *2*, S127–S136.

Wlodarski, T., and Zagrovic, B. (2009). Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. Proc. Natl. Acad. Sci. USA *106*, 19346–19351.

Wlodawer, A., Svensson, L.A., Sjölin, L., and Gilliland, G.L. (1988). Structure of phosphate-free ribonuclease A refined at 1.26 A. Biochemistry *27*, 2705–2717.

Yang, L., Song, G., and Jernigan, R.L. (2009). Protein elastic network models and the ranges of cooperativity. Proc. Natl. Acad. Sci. USA *106*, 12347–12352.

Zachariae, U., Schneider, R., Velisetty, P., Lange, A., Seeliger, D., Wacker, S.J., Karimi-Nejad, Y., Vriend, G., Becker, S., Pongs, O., et al. (2008). The molecular mechanism of toxin-induced conformational changes in a potassium channel: relation to C-type inactivation. Structure *16*, 747–754.

Zheng, W., Brooks, B.R., and Hummer, G. (2007). Protein conformational transitions explored by mixed elastic network models. Proteins *69*, 43–57.