CHAPTER 4

Stochastic Quasi-Newton: application to minimal model for proteins

Knowledge of protein folding pathways and inherent structures is of utmost importance for our understanding of biological function, including the rational design of drugs and future treatments against protein misfolds. Computational approaches have now reached the stage where they can assess folding properties and provide data that is complementary to or even inaccessible by experimental imaging techniques. Minimal models of proteins, that enable the simulation of protein folding dynamics by (systematic) coarse-graining, have provided understanding in terms of descriptors for folding, folding kinetics and folded states. Here, we focus on the efficiency of equilibration on the coarse-grained level. In particular, we applied a new regularized stochastic Quasi-Newton (S-QN) method, developed for accelerated configurational space sampling while maintaining thermodynamic consistency, to analyse the folding pathway and inherent structures of a selected protein, where regularization was introduced to improve stability. The adaptive compound mobility matrix B in S-QN, determined by a factorized secant update (FSU), gives rise to an automated scaling of

all modes in the protein, in particular an acceleration of protein domain dynamics or principal modes and a slowing down of fast modes or 'soft' bond constraints, similar to LINCS/SHAKE algorithms, when compared to conventional Langevin dynamics. We used and analyzed a two-step strategy. Owing to the enhanced sampling properties of S-QN and increased barrier crossing at high temperatures (in reduced units), a hierarchy of inherent protein structures is first efficiently determined by applying S-QN for a single initial structure and $T = 1 > T_\theta$, where T_θ is the collapse temperature. Second, S-QN simulations for several initial structures at very low temperature ($T = 0.01 < T_F$, where T_F is the folding temperature), when the protein is known to fold but conventional Langevin dynamics experiences critical slowing down, were applied to determine the protein domain dynamics (collective modes) towards folded states, including the native state. This general treatment is efficient and directly applicable to other coarse-grained proteins.

4.1. Introduction

Efficient sampling of (free) energy landscapes is important in many physical systems, especially when this landscape is very rugged and/or equilibrium states are unknown. In methods that are based on an intrinsic kinetic description, like Molecular Dynamics (MD), the sampling performance is dictated by the smallest time or length scale in the description. As a result, interesting phenomena like the folding of large proteins, in which the scale associated the fastest/smallest (bond vibrations) and the slowest/largest (formation of α or β domains) modes deviate by several orders of magnitude, are inaccessible by standard MD. Slow processes like the cooperative motion of protein domains remain inaccessible even with the increasing computer power, unless some sort of coarse-graining or averaging over the smallest degrees of freedom is carried out. In recent years, approaches based on smoothing, i.e. equilibration on a coarse-grained level followed by fine-grained refinement [79], were developed to address this problem and applied with some success. Several groups have concentrated on determining representative coarse-grained minimal models of proteins. Nevertheless, this approach suffers from a hereditary property, since now the smallest scale on the coarse-grained level determines the sampling performance. Our starting point is conventional coarse-grained Langevin dynamics (CLD), a widely used stochastic model for effective diffusion on a coarse-grained level. We previously showed how to adapt CLD for improved sampling [80]. The general stochastic Quasi-Newton (S-QN) method applies an automated scaling for different length/time scales in the system, by including curvature or correlation information in the space-dependent mobility, while maintaining thermodynamic consistency. Due to the scaling, all modes in the system are updated roughly equally fast, enabling a significantly larger time step (orders of magnitude) compared to CLD. In addition, within this framework CLD reduces to a stochastic form of the well-known steepest descent method. Since quasi-Newton methods are known for their improved ability to locate saddle points compared to steepest descent, also the sampling pathway is positively affected. In [81], we introduced the fundamentals for the efficient determination of $J(\mathbf{x})$ in

$$\mathrm{d}\mathbf{x} = [-B(\mathbf{x})\nabla\Phi(\mathbf{x}) + k_B T\nabla\cdot B(\mathbf{x})]\mathrm{d}t + \sqrt{2k_B T}\,J(\mathbf{x})\mathrm{d}W(t), \tag{4.1}$$

and considered in detail the performance of S-QN for a quadratic energy potentials Φ. In (4.1), the space-dependent mobility $B(\mathbf{x}) = J(\mathbf{x})J(\mathbf{x})^T$ is determined such that it approximates the inverse Hessian. Moreover, the spurious drift term was omitted in Ref. [81] since we only considered quadratic $\Phi(\mathbf{x})$, k_B is the Boltzmann constant,

T the temperature and dW is a Wiener process. Turning to the discrete S-QN equation [81],

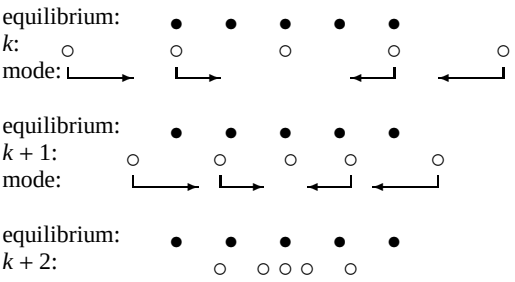
$$\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{2}[B(\mathbf{x}'_k)\nabla\Phi(\mathbf{x}'_k) + B(\mathbf{x}_k)\nabla\Phi(\mathbf{x}_k)]\Delta t + \frac{1}{2}[B(\mathbf{x}'_k)B^{-1}(\mathbf{x}_k) + I]\sqrt{2k_BT}\,J(\mathbf{x}_k)\Delta W_t, \tag{4.2}$$

$$\mathbf{x}'_k = \mathbf{x}_k + \Delta\mathbf{x}^p_k, \tag{4.3}$$

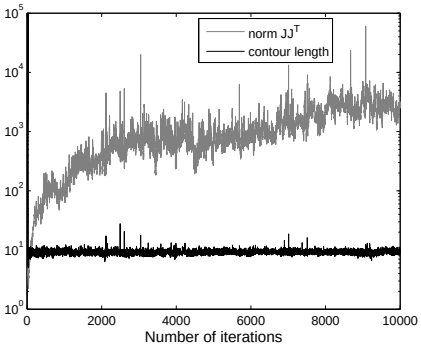
$$\Delta\mathbf{x}^p_k = -B(\mathbf{x}_k)\nabla\Phi(\mathbf{x}_k)\Delta t + \sqrt{2k_BT}\,J(\mathbf{x}_k)\Delta W_t. \tag{4.4}$$

the matrix J in $B = JJ^T$ is determined by a rank-two factorized secant update (FSU) scheme. The inverse B^{-1} is determined via an update method in dual space [81]. As discussed before [81], S-QN can be seen as a real-space analogon of existing Fourier acceleration approaches [56–58].

So far, we did not consider a general solution to the problem of ill-conditioning, i.e. when the condition number $\kappa(H)$ of the Hessian matrix $H = H(\mathbf{x})$ becomes very large or when H is singular. For general energy landscapes, this situation is very likely to occur along the sampling path. A straightforward example of such a situation is that Φ is often invariant under a transformation of the whole system, rendering a singular H and a completely flat energy landscape along this eigenvector of H. As B is determined to increasingly approximate the inverse of H, the condition number $\kappa(B)$ will naturally become very large. For this reason, the update may become prone to numerical errors in the update of B and/or $\nabla\Phi$. An additional complication is that the time step in (4.4) is large and that line searches are not included, in contrast to most standard optimization methods [82]. Moreover, for efficiency, our scheme only corrects B in sampled points, based on local Hessian information. When the energy landscape is almost flat along one of the eigenvectors, the large scaling along this direction will give rise to an updated configuration in a distant and possibly uncorrelated part of the energy landscape. As a consequence, the next few updates may suffer from a complex interplay of numerical errors and a sudden discrepancy between B and the actual inverse Hessian. An illustrative sketch of such an interplay is shown in Figure 4.1a, for a Rouse chain with constant but singular Hessian. Since the Hessian is constant, one could understand the origin of this over-acceleration in terms of the constant time step. Upon applying S-QN for the Rouse chain in three dimensions, however, a distinct correlation can be observed between the sudden deviations from the equilibrium chain length and the peaks in $\|B\|_F$ (Figure 4.1b) that start to appear for rather large $\|B\|_F$. Close examination indicated that these chain length deviations



(a) *Over acceleration of the slow mode.*



(b) $\|JJ^T\|_{\mathrm{F}}$ *and contour length, where the peaks in the contour length coincide with the peaks in* $\|JJ^T\|_{\mathrm{F}}$*; equilibrium chain length is 9.*

Figure 4.1: *Reasons regularization needed: over amplification of the slow modes and erroneous approximation of the eigenvectors.*

are due to numerical errors in updating B. In particular, Φ is invariant under translation of the whole chain, which is the eigenvector $\mathbf{1} = [1...1]^T$ of H with eigenvalue zero. In terms of the (generalized) inverse of H, this eigenvector is associated with the largest eigenvalue (or scaling). After determining the eigenvector associated with the largest eigenvalue for each B, the sudden increase in chain length and norm of B was found to coincide with a slight deviation of this eigenvector from $\mathbf{1}$. Due to large scaling along this direction, deviations that exist for some particles along the chain will be substantially amplified and result in chain extension or compression. Nevertheless, the equilibrium chain length was recovered quickly after chain extension or compression. A way to resolve both discussed issues above simultaneously is to constrain the step size or, equivalent, condition the matrix B by regularization.

Our main goal in this Chapter is to apply the general S-QN method for a system of obvious physical relevance. Minimal models for proteins have the advantage that systematic coarse-graining procedures gave rise to rather simple but accurate energy expressions and that folding pathways and folded states for a number of proteins were determined both experimentally and by simulation [83–91]. Fast and accurate determination of folded states and primary modes along the folding pathway is not only interesting from a scientific point of view, this system also represents a good

benchmark for S-QN because of the complex energy landscape topography and the fact that conventional Langevin dynamics is often employed in current studies [92, 93]. We choose one of these proteins, containing 22 amino acids, as a reference system.
First, we introduce the new regularization parameter ϵ and show that the regularized B^{-1} converges to $H + \epsilon I$, thereby resolving both the problem of conditioning B and the singularity of H. Since regularization gives rise to only a slight modification of the FSU scheme, the efficiency of the original scheme ($8n^2$) is retained. Although we will not consider limited-memory implementation here, the approach can easily be extended to L-FSU [81]. We analyze the effect of regularization and apply the new scheme to the minimal model of the chosen protein in a stepwise manner, by including an increasing number of terms of the total coarse-grained energy expression. We conclude with a detailed analysis of the collective dynamics and sampling behavior for the considered coarse-grained protein.

4.2. Theory

Our factorized secant update (FSU) scheme for updating the factorized term is given by

$$J_{k+1} = J_k + \frac{\alpha_k \mathbf{s}_k \mathbf{y}_k^T J_k - \alpha_k^2 J_k J_k^T \mathbf{y}_k \mathbf{y}_k^T J_k}{\mathbf{y}_k^T \mathbf{s}_k}, \tag{4.5}$$

with

$$\alpha_k^2 = \frac{\mathbf{y}_k^T \mathbf{s}_k}{\mathbf{y}_k^T J_k J_k^T \mathbf{y}_k}, \tag{4.6}$$

where $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$, $\mathbf{y}_k = \nabla\Phi(\mathbf{x}_{k+1}) - \nabla\Phi(\mathbf{x}_k)$ and $B_{k+1} = J_{k+1} J_{k+1}^T$. We have previously shown that (4.5) is equivalent to the standard Davidon-Fletcher-Powell (DFP) update formula [81],

$$B_{k+1} = B_k - \frac{B_k \mathbf{y}_k \mathbf{y}_k^T B_k}{\mathbf{y}_k^T B_k \mathbf{y}_k} + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}. \tag{4.7}$$

For a regular Hessian H, the DFP method is known to possess three important properties [66]: a) positive-definiteness of the update matrix is assured provided that B_0 is chosen to be positive definite, b) for a positive-definite quadratic objective function

$\Phi(\mathbf{x})$, with $\mathbf{x}$ an n-dimensional vector, the algorithm converges to the solution in at most n steps, and c) for a positive-definite quadratic objective function $\Phi(\mathbf{x})$, and if the convergence to the solution requires the full n steps, $B_n = H^{-1}$ where the matrix H is the exact inverse Hessian.

In particular property a) is a prerequisite for our method, due to the presence of the noise term containing J with $JJ^T = B$. Convergence properties b) and c) are very sensitive to the line search accuracy, i.e. the determination of the step size α_k such that Φ is minimal along the search direction or, alternatively, that all search directions $\mathbf{s}_k = \alpha_k B_k \nabla\Phi$ are orthogonal. Since our $\mathbf{s}_k$ contains *two* terms, scaling as α_k and $\sqrt{\alpha_k}$ respectively, and since J is not always available as a matrix, the step size in our method, the physically relevant time step Δt [81], was chosen to be a *constant*. In contrast to QN methods, where the condition for the step size is either an exact line search, a reduction of $||\Phi||$ or the equivalence to Newton's method ($\alpha_k = 1$) [94], the time step in S-QN is also physically constrained. Our time step should be such that the fastest mode in the system is well represented for $M(\mathbf{x}) = B_0 = I$, the often considered initial value for B. For quadratic positive-definite Φ, or for Φ that are locally well approximated by a quadratic function, the fastest mode is associated with the largest eigenvalue $\lambda_{\max}(H)(>> 1)$ of the Hessian H and the time step Δt is thus upper bounded by $\lambda_{\max}(H)\Delta t < 1$ as a rule of thumb. Only when the mobility is a good approximate of the inverse Hessian, the upper bound increases to $\Delta t < 1$ and we obtain improved sampling efficiency.

For a quadratic Φ, we have previously shown that B_k determined by FSU with a constant time step does converge to the generalized inverse Hessian H^- in $O(n^3)$ steps, instead of the theoretical quadratic convergence ($O(n)$) in property c) [81]. However, one should consider that this H was *singular* and that the convergence of the non-extremal eigenvalues of B was actually much faster than $O(n^3)$. The invariance of Φ under certain fully cooperative displacements, such as a translation or a rotation with respect to the center of mass, is a common property and gives rise to a singular H. As a result of $\lambda_{\min}(H) = 0$ and the rank-two update formula, the extremal eigenvalues of B continued to increase/decrease with increasing k and the condition number of B, $\kappa(B) = \lambda_{\max}(B)/\lambda_{\min}(B)$, became arbitrarily large. In general, a *low* condition number $\kappa(B)$ is desirable because of possible round-off errors in the update scheme [95]. As $\lambda_{\max}(B)$ becomes very large, any error in the associated eigenvector is seriously amplified. Moreover, since this displacement is very large, the condition number $\kappa(B_k)$ may be much larger than the condition number of the inverse Hessian $G(\mathbf{x}_{k+1})$ in the updated position for general Φ. It is unclear if the DFP scheme can recover [96].

4.2.1. Existing approaches for conditioning B

Solutions to both problems, the exact line searches and conditioning, have been considered in the QN literature. The attempts to avoid ill-conditioning mainly focus on quadratic and convex objective functions and also the analysis was carried out under conditions which are typical for quasi-Newton methods. Many of the algorithms without line searches are rank-one update formulas that do not always satisfy the required property a), i.e. a positive-definite B, and may even give rise to singular updates [97]. The most successful rank-two approach without exact line searches relies on the convex class of update formulas given by [96]

$$B_{k+1}(\phi) = (1-\phi)B_{k+1}^{\text{DFP}} + \phi B_{k+1}^{\text{dual}} = B_{k+1}^{\text{DFP}} + \phi \mathbf{v}_k \mathbf{v}_k^T \tag{4.8}$$

where B_{k+1}^{DFP} is the DFP update given by (4.7) and

$$\begin{aligned} B_{k+1}^{\text{dual}} &= (I - \frac{\mathbf{s}_k \mathbf{y}_k^T}{\mathbf{s}_k^T \mathbf{y}_k}) B_k (I - \frac{\mathbf{y}_k \mathbf{s}_k^T}{\mathbf{s}_k^T \mathbf{y}_k}) + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{s}_k^T \mathbf{y}_k} \\ \mathbf{v}_k &= (\mathbf{y}_k^T B_k \mathbf{y}_k)^{1/2} \{ \frac{\mathbf{s}_k}{\mathbf{s}_k^T \mathbf{y}_k} - \frac{B_k \mathbf{y}_k}{\mathbf{y}_k^T B_k \mathbf{y}_k} \} . \end{aligned} \tag{4.9}$$

The update B_{k+1}^{dual} is dual to DFP in the sense that it is derived from (4.7) by inversion via the Sherman-Morrison theorem and subsequently interchanging $\mathbf{y}_k$ and $\mathbf{s}_k$. The free parameter ϕ is determined based on update properties and should be chosen $0 \leq \phi \leq 1$ to retain convergence property c) [96, 98]. This class of update formulas was found to be more robust to the update becoming singular or unbounded than the individual methods for $\phi \in \{0, 1\}$ in the case of inexact line searches. Later contributions have adapted (4.8) for optimal conditioning of B (see [99] for details). In *sizing*, the objective function Φ is multiplied with a scalar parameter, based on the recognition that the update scheme is invariant under this scaling. The second method is *switching*, where this switch is between different methods of the convex class (4.8).

Issues of instability and sensitivity to (numerical) noise also occur in *ill-posed* problems in mathematics. Common strategies to deal with these issues include reducing the dimensions of the system, e.g. using different refinements: rigid body, torsion and angle, and removing or adapting the small eigenvalues of the system. A number of established methods, including truncated singular value decomposition, truncated total least square, generalised singular value decomposition, truncated rank-revealing decomposition and Tikhonov regularization, exist for this purpose. Physically-inspired

regularization uses prior knowledge of the system to convert an ill-posed problem (or singular problems) into regular problems. A way to do this is to define additional constraints, for instance by prohibiting atoms from getting too close to each other. In our Rouse chain [81], constraints based on the null space of H could be applied, by subtracting translations and rotations in the displacement of the total chain. Translations are treated by keeping the chain center of mass fixed, and subtracting this part of the displacement in (4.1). Subtracting chain rotation can be done by finding a rotation matrix R, such that the root mean square deviation (RMSD), defined as

$$\text{RMSD} = \min_R \sum_i^N ||R\tilde{\mathbf{r}}_i^k - \mathbf{r}_i^k|| \tag{4.10}$$

is minimal, where $\tilde{\mathbf{r}}_i$ is the position of particle i at iteration k after adding the Langevin displacement and after translation of the chain back to its original center of mass. The new position of particle i becomes $\mathbf{r}_i^{k+1} = R\tilde{\mathbf{r}}_i^k$. Obviously, determining R by straightforward minimization is inefficient, but quaternion algebra provides an efficient alternative [100].

4.2.2. RFSU: a regularized FSU method

Since proper scaling of different modes is key to our approach, we consider the ill-conditioned or singular H and hence the arbitrary large condition number $\kappa(B)$ the most important issue. We further rely on the observation that the convergence for the non-extreme eigenvalues of B for the quadratic Φ in Ref. [81] was relatively fast and also the free parameter ϕ in the convex class of Broyden methods is determined based on conditioning properties. Moreover, by conditioning (4.7) we avoid the nontrivial task of deriving a factorized update algorithm for (4.8) and determining a good value of ϕ for general objective functions Φ (remember that $\phi = 1$ is the DFP method). We could in principle use *sizing* for conditioning B determined by (4.7), but this may not help since H itself is singular. Instead, we regularize the DFP scheme using an approach that is conceptually very similar to Tikhonov regularization.

The key idea is very simple: we adapt the update scheme such that $\tilde{B}_k^{-1}$ converges to $\tilde{H} = H + \epsilon I$, with ϵ a (small) regularization parameter. It is easy to show that the eigenvalues $\{\lambda_1(H), ..., \lambda_n(H)\}$ shift to $\{\lambda_1(\tilde{H}), ..., \lambda_n(\tilde{H})\} = \{\lambda_1(H) + \epsilon, ..., \lambda_n(H) + \epsilon\}$. For $\epsilon > 0$, all eigenvalues of $\tilde{H}$ are non-zero and the condition number becomes bounded, i.e. $\kappa(\tilde{H}) = \lambda_{\max}(H)/\epsilon + 1$, since $\lambda_{\max}(H)$ is bounded. Similar regularization was previously introduced [101] to deal with negative curvatures, but relied on

compute-intensive explicit determination of eigenvalues of B via LDL^T factorization. Although the key idea is simple, we have to adapt our update formula for B in dual space. Let B_{k+1} of (4.7) be the update. The inverse G_{k+1} of B_{k+1} in the dual space is explicitly given by (4.9) after interchanging $\mathbf{y}_k$ and $\mathbf{s}_k$:

$$G_{k+1} = (I - \frac{\mathbf{y}_k\mathbf{s}_k^T}{\mathbf{y}_k^T\mathbf{s}_k})G_k(I - \frac{\mathbf{s}_k\mathbf{y}_k^T}{\mathbf{y}_k^T\mathbf{s}_k}) + \frac{\mathbf{y}_k\mathbf{y}_k^T}{\mathbf{y}_k^T\mathbf{s}_k} \tag{4.11}$$

It is easy to see that the secant condition in the dual space, $G_{k+1}\mathbf{s}_k = \mathbf{y}_k$, is satisfied. Following an idea introduced by Powell for treating the ill-conditioned case in constrained optimization [102, 103], we note that any matrix $\tilde{G}_{k+1}$ obtained by (4.11) after replacing $\mathbf{y}_k$ by $\tilde{\mathbf{y}}_k$ satisfies the secant condition $\tilde{G}_{k+1}\mathbf{s}_k = \tilde{\mathbf{y}}_k$. Let again $\tilde{G}_{k+1}$ be the regularized matrix $G_{k+1} + \epsilon I$. Since G_k converges to H, $\tilde{G}_k$ converges to $\tilde{H} = H + \epsilon I$. We can now determine $\tilde{\mathbf{y}}_k$ from

$$\tilde{\mathbf{y}}_k = \tilde{G}_{k+1}\mathbf{s}_k = G_{k+1}\mathbf{s}_k + \epsilon\mathbf{s}_k = \mathbf{y}_k + \epsilon\mathbf{s}_k. \tag{4.12}$$

After inversion, we obtain the update scheme

$$\tilde{B}_{k+1} = \tilde{B}_k - \frac{\tilde{B}_k\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k^T\tilde{B}_k}{\tilde{\mathbf{y}}_k^T\tilde{B}_k\tilde{\mathbf{y}}_k} + \frac{\mathbf{s}_k\mathbf{s}_k^T}{\tilde{\mathbf{y}}_k^T\mathbf{s}_k} \tag{4.13}$$

with $\tilde{\mathbf{y}}_k = \mathbf{y}_k + \epsilon\mathbf{s}_k$ that satisfies the secant condition $\tilde{B}_{k+1}\tilde{\mathbf{y}}_k = \mathbf{s}_k$. Since $\tilde{\mathbf{y}}_k^T\mathbf{s}_k = \mathbf{y}_k^T\mathbf{s}_k + \epsilon\mathbf{s}_k^T\mathbf{s}_k$ and $\mathbf{s}_k^T\mathbf{s}_k$ is always positive, a possible cause of numerical errors in the original scheme, i.e. $\mathbf{y}_k^T\mathbf{s}_k$ is positive but very small, is avoided. We only left with the proof that $\tilde{B}_k$ converges to $[H+\epsilon I]^{-1}$. We follow the proof given in Ref. [96] (Section 6) and assume that the objective function is quadratic positive-definite. Setting $K = [H + \epsilon I]^{1/2}\tilde{B}[H + \epsilon I]^{1/2}$ and denoting K_{k+1} and K_k by K^* and K respectively, we can rearrange (4.13) into

$$K^* = K - \frac{K\mathbf{z}\mathbf{z}^TK}{\mathbf{z}^TK\mathbf{z}} + \frac{\mathbf{z}\mathbf{z}^T}{\mathbf{z}^T\mathbf{z}} \tag{4.14}$$

where $\mathbf{z} = [H + \epsilon I]^{1/2}\mathbf{s}_k$ and $\mathbf{y}_k = H\mathbf{s}_k$. In particular, $\mathbf{z}$ is an eigenvector of the first two terms in (4.14) with $\lambda = 0$ and the third term sets this eigenvalue to $\lambda = 1$ while leaving the others unchanged. Since relation (4.14) is exactly the same as in Ref. [96], the convergence of $\tilde{B}$ to $[H + \epsilon I]^{-1}$ is proven. The matrix $\tilde{H}$ is now regular, so $[H + \epsilon I]^{-1} = H^{-1} - \epsilon H^{-1}H^{-1} + O(\epsilon^2)$ exists and is well defined. Finally, we shortly consider the condition number of $\tilde{B}_{k+1}$. Let $\mathbf{v}$ be a eigenvector of G_{k+1} with

eigenvalue λ. It is easy to see that $\mathbf{v}$ is also an eigenvector of B_{k+1} with eigenvalue $1/\lambda$. Moreover, $\mathbf{v}$ is an eigenvector of $G_{k+1} + \epsilon I = \tilde{G}_{k+1}$ with eigenvalue $\lambda + \epsilon$ and of $\tilde{B}_{k+1}$ with eigenvalue $1/(\lambda + \epsilon)$. Hence, the condition number $\kappa(\tilde{B}_{k+1}) = (\lambda_{\max}(G_{k+1}) + \epsilon)/\epsilon$ with ϵ a freedom of choice. In other words, the condition number and the maximum displacements have become bounded when $\lambda_{\max}(G_{k+1})$ is finite. A factorized $\tilde{J}_{k+1}$ such that $\tilde{J}_{k+1}\tilde{J}_{k+1}^T = \tilde{B}_{k+1}$ is found by (4.5) after replacing $\mathbf{y}_k$ by $\tilde{\mathbf{y}}_k$.

4.3. Results and discussion

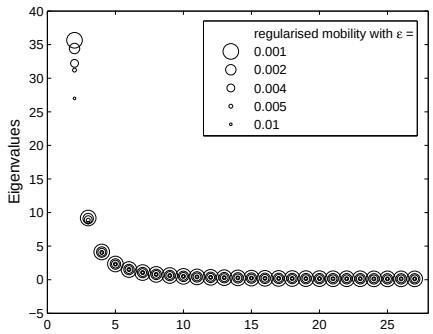
4.3.1. The choice of the regularization parameter

A good choice of ε depends on properties of the system, in particular on the time step Δt and on the smallest eigenvalues of the Hessian H. Since H can easily become singular, a useful rule of thumb is that the step size on the energy landscape is maximized by RFSU, by a factor $\equiv \Delta t/\varepsilon$. Small ε only slow down the largest collective displacements, while for very large ε, the regularization term in B starts to dominate and RFSU will resemble conventional Langevin dynamics with constant friction $1/\varepsilon$. Here, we start by analyzing RFSU for the 1-D chain of our previous paper [81] with a harmonic potential

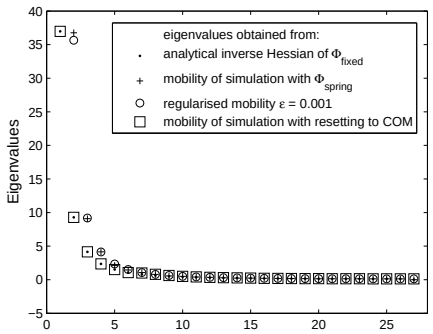
$$\Phi_{\text{spring}} = \sum_{i=1}^{n-1}(x_{i,i+1} - 1)^2; \tag{4.15}$$

and $x_{i,i+1} = |x_i - x_{i+1}|_2$ the distance between particle i and $i+1$. Since the Hessian for this potential is constant, we can analyze the converged eigenvalues of B for different ε. We note that these eigenvalues should converge to $(\lambda(H)+\varepsilon)^{-1}$ (see theory section). Figure 4.2a shows that choosing different ε indeed does not have a noticeable effect on the smallest eigenvalues. As expected, the effect is most significant for the larger eigenvalues ($\equiv$ small $\lambda(H)$) and the largest eigenvalue always converged to $1/\varepsilon$ (remember that $\lambda^{\min}(H) = 0$). In Figure 4.2b we compare the eigenvalue spectra of B_ε ($\circ$) using RFSU for $\varepsilon = 0.001$ and B_Φ (+) using FSU for $n = 27$. We previously showed [81] that B_Φ converges to a generalized inverse of H. After 10000 steps, the largest eigenvalue of B_ε has converged to $1/\varepsilon = 10^3$ while the largest eigenvalue of B_Φ is still increasing [81] (both exceed the vertical axes limit). In particular, B obtained by RFSU converges to a unique $(H + \varepsilon I)^{-1}$, the rate of which depends on the value of ε. Additional symbols in Figure 4.2b denote the eigenvalues spectra obtained

for (4.15) by analytic inversion and for two types of physically-inspired regularizations. First (□), we explicitly constrained FSU for chain translation by resetting the center of mass to the original position at each time step. Second (·), we calculated the analytic eigenvalue spectrum for (4.15) with an additional penalty function for the center of mass, similar to earlier work [81]. It is clear that these two types of regularizations have a very similar effect, since the eigenvalue spectra overlap. However, both spectra differ from RFSU. Removing or penalizing the slowest mode in the displacements apparently slows down *all* modes compared to the non-regularized case. In particular, the eigenvalue spectrum seems shifted (by one) when one compares the spectra for the non-regularized (+) and the physically-inspired regularized results (□ and ·). In contrast, the spectrum for RFSU and non-regularized FSU share many features. In RFSU the acceleration of the slowest modes is maximized and the condition number $\kappa(B)$ is therefore constrained. We also review the effect of



(a) *Eigenvalues of mobility constructed with regularized FSU for different values ε.*

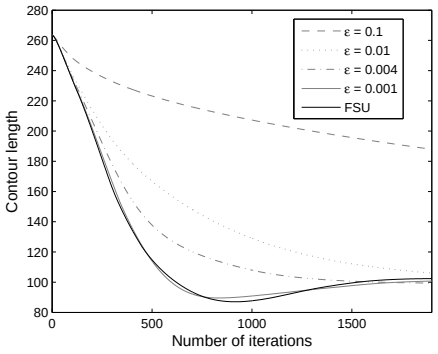


(b) *Eigenvalues of the mobility constructed with different regularization methods.*

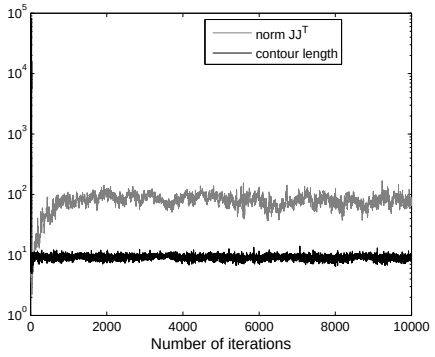
Figure 4.2: *Comparison of the eigenvalues of the analytic inverse Hessian and the mobility as constructed by regularized FSU and other regularization methods with n = 27.*

regularization on the ill-conditioning and over-acceleration issues discussed in the introduction (see Figure 4.1). We consider a 1-D chain (4.15) for $n = 100$ for over-acceleration [81] and the same 3-D Rouse chain as in Figure 4.1 for ill-conditioning. Figure 4.3a shows the evolution of the chain length for different values of ε, including $\varepsilon = 0$ of FSU. Over-acceleration persists in RFSU for small values of ε, as the chain length falls substantially below the equilibrium length $l_{eq} = 99$ before converging.

For the largest considered value, $\varepsilon = 0.1$, convergence to l_{eq} is considerably slowed down and approaches the one obtained for conventional Langevin dynamics [81], since the diagonal regularization matrix starts to dominate. Here, the contribution of collective modes is severely damped by maximizing the largest eigenvalues of B to $1/\varepsilon = 10$, resulting in a much slower cooperative chain contraction than in FSU. Proper behavior is obtained somewhere between these two extremes, in particular for $0.025 < \varepsilon < 0.01$, where the chain length convergence is even better than that of FSU. We found that $\varepsilon = 0.004$ is optimal. Application of RFSU also resolves the problem



(a) *Effect of the regularization on the convergence towards the equilibrium contour length of 3-D Rouse chain with equilibrium length 99.*



(b) *Effect of the regularization on $\|JJ^T\|_{\mathrm{F}}$ and the contour lengths during the simulation of a 3-D Rouse chain with equilibrium length 9.*

Figure 4.3: *Examples of the regularization effect on over acceleration and on amplification of numerical errors.*

of an ill-conditioned B. Upon comparing the matrix norm of B in Figure 4.1b for FSU and Figure 4.3b using RFSU with $\varepsilon = 0.01$, one can observe that the sudden increases in the chain length are completely absent after regularization. Clearly, the effect of amplification of numerical errors in B is no longer present or considerably reduced.

4.3.2. Minimal model for a protein

The development of coarse-grained or minimal models for proteins has always been a very active area of research. Besides the intellectual challenge, a solution to the

protein folding problem will have important implications for the understanding and design of protein function and possible causes of diseases due to protein misfolds. Experiments show that small, single-domain proteins reach their native states on the time scale in the order of 10-1000 milliseconds, thereby setting a lower bound for the representation of the smallest modes. The effect of smoothing the energy landscape by coarse-graining was previously discussed (we refer to existing reviews [104] for details) and coarse-grained models are believed to provide valuable information of folding, assembly and function(s) of biomolecules.

Since the earliest efforts of Flory for heterogeneous polymers [105], several groups have focussed on a coarse-grained off-lattice C_α representation of proteins. Such a representation disregards side chains effects and introduces a three letter particle code for each peptide depending on their nature: *hydrophobic*(B), *polar or hydrophilic*(L) and *neutral*(N). A general expression for the conformational energy potential Φ_p for a string of particles is then given by

$$\Phi_\mathrm{p} = \Phi_\mathrm{bond} + \Phi_\mathrm{bending} + \Phi_\mathrm{dihedral} + \Phi_\mathrm{non}, \tag{4.16}$$

where the terms are the potential for bond length variation, which can be omitted if the bond length is fixed [109], bond angles, dihedral angles and nonbonded interactions. Using this model, several sequence-dependent properties like folding trajectories, metastable fold states and life-times, the effect of mutations and thermodynamic properties like folding and collapse temperatures, in order to discriminate between fast and slow folders, have been considered and compared to existing experimental observations [91].

A realistic kinetic description is vital for these studies and most of them rely on Langevin dynamics, either in the over- or under-damped limit. Our S-QN model considers Langevin dynamics in the overdamped case as a starting point but does not follow a realistic pathway due to the incorporation of curvature information. This is a property shared among many other methods, for instance Monte Carlo (MC) and methods that employ a biasing potential [106], aimed at improved sampling. However, S-QN does allow for an automated determination of collective modes and is much more efficient than Langevin dynamics in finding transition and equilibrium states in a thermodynamically consistent way, as we will show in this section for a particular coarse-grained protein. Due to these unique properties, we believe that this model represents a valuable contribution to the coarse-grained protein modeling toolkit.

4.3.2a. Terms in the energy potential

The bond potential is given by

$$\Phi_{\text{bonded}} = \sum_{i=1}^{N-1} k^b_{ij}(d_{i,i+1} - d_0)^2, \tag{4.17}$$

with N the number of particles, $j = i + 1$, $d_{i,i+1} = \|\mathbf{r}_i - \mathbf{r}_{i+1}\|_2$ is the distance between particle i and $i + 1$, d_0 is the reference bond length and k^b_{ij} is the force (or spring) constant. The angle or bending potential is given by

$$\Phi_{\text{bending}} = \sum_{i=1}^{N-2} k^a_{ijk}(\theta_{ijk} - \theta_0)^2, \tag{4.18}$$

with $j = i + 1$, $k = i + 2$, θ_{ijk} the bond angle defined by the bond vectors $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and $\mathbf{r}_{kj} = \mathbf{r}_k - \mathbf{r}_j$, θ_0 is the reference angle and k^a_{ijk} is the force constant. The torsion potential considered in the next paragraph is very general,

$$\Phi_{\text{torsion}} = \sum_{i=1}^{N-3} k^t_{ijkl}(1 + \cos(n_{ijkl}\phi_{ijkl} - \phi_0)), \tag{4.19}$$

with $j = i + 1$, $k = i + 2$, $l = i + 3$, ϕ_{ijkl} the torsion angle between the normal of the plane through particles i, j, k and the normal of the plane through particles j, k, l. Again k^t_{ijkl} is the force constant and ϕ_0 is the reference angle and n_{ijkl} is the multiplicity. In the simulations of the specific protein, we apply a somewhat different potential [91]

$$\Phi_{\text{dihedral}} = \sum_{i=1}^{N-3} \left[A_i(1 + \cos\phi_i) + B_i(1 + \cos 3\phi_i)\right], \tag{4.20}$$

The non-bonded potential is of Lennard-Jones type and given by

$$\Phi_{\text{non}} = \sum_{i=1}^{N-3} \sum_{j=i+1}^{N} 4\left[C_{ij}\left(\frac{d_0}{d_{i,j}}\right)^{12} - D_{ij}\left(\frac{d_0}{d_{i,j}}\right)^{6}\right], \tag{4.21}$$

4.3.2b. Preliminary calculation of sampling distributions

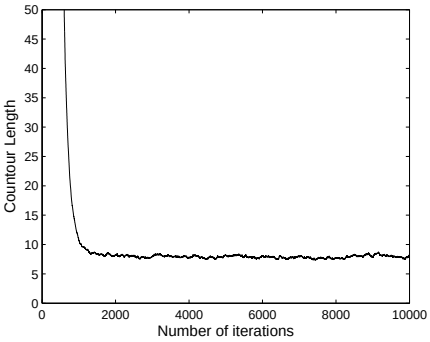
First, we consider the thermodynamically consistency of the sampling obtained by RFSU, which should be the Boltzmann distribution in equilibrium, for the case where only intra-molecular interactions contribute. Sampling distributions for the bond length, angle and torsion, for the torsion potential (4.19), are considered separately using a simulation of 10000 steps in total. We apply RFSU for a string of $N = 9$ equivalent particles, $n = 3 \times N = 27$, and use $k_BT = 10$ and $\Delta t = 10^{-2}$. The force constants are set to $k^b_{ij} = k^a_{ijk} = k^a_{ijkl} = 10^3$ and the reference variables $d_0 = 1$ and $\theta_0 = \phi_0 = \frac{\pi}{2}$. We varied ε and the initial configurations. The presented results are for one initial configuration and rather high $\varepsilon = 0.1$, but representative for all other cases.

The chain length during the simulation, Figure 4.4a, converges to an equilibrium $l_{eq} = (N-1)d_0 = 8$ after roughly 1000 iterations. The analytic Boltzmann distribution was calculated using $\mathcal{N}\exp(-\Phi/k_BT)$, where Φ is the considered potential term. Starting with the bond length, we see in Figure 4.4b that the analytic distribution is reproduced well by the simulated sampling distribution, with mean $d_0 = 1$. From Figures 4.4c and 4.4d we conclude the same for the simulated distribution of the bending angles and dihedrals, respectively. Apparently, S-QN is effective in sampling the equilibrium values for the bond length, angles and dihedrals, in spite of the rather large ε.

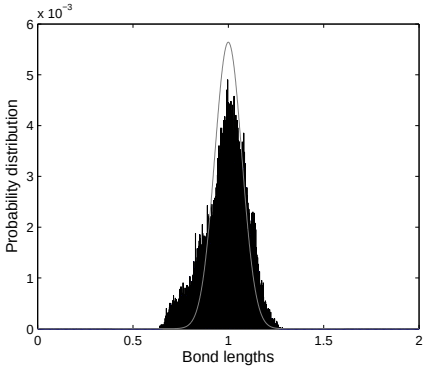
We note that the update is governed by the noise term for the chosen value of $k_BT = 10$ (the noise/drift ratio >> 1). Although slow modes are seriously damped due to regularization, relative to the non-regularized case, the collective contributions of noise and drift enable this effective sampling. As a control, we also performed a simulation using conventional Langevin equation, with a constant mobility $M = 1$ (all other parameters are the same). After 10000 steps, the sampling distribution is far from equilibrium and preferable bond length, bending and torsion angles were not detected.

4.3.2c. Application to the model protein

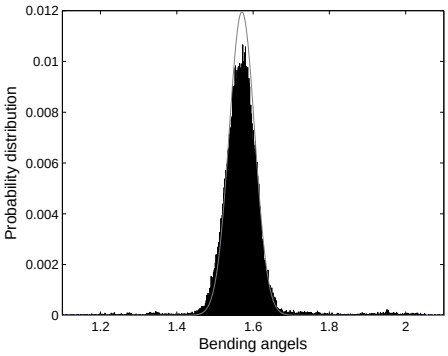
Finally, we apply RFSU to an earlier considered coarse-grained protein G [91], represented by the string $LB_9(NL)_2NBLB_3LB$, where each letter denotes a hydrophobic (B), hydrophilic (L) or neutral (N) coarse-grained peptide. The simulations in Ref. [91] were performed using Langevin dynamics in the under-damped limit for different friction coefficients. Since the value of $\sigma = T_\theta - T_F/T_\theta$ for G is rather low ($\sigma = 0.20$), with $T_\theta = 0.78$ the collapse transition and $T_F = 0.62$ the folding transition temperature (in reduced units of ε_h/k_B, with ε_h a scaling factor in the non-bonded



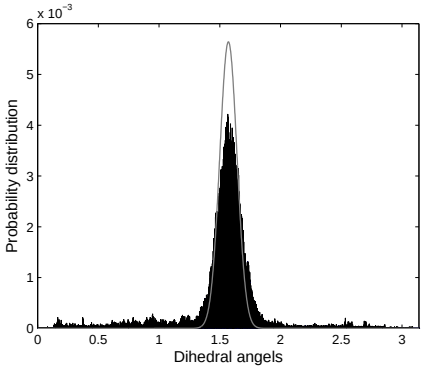
(a) *Total chain length vs iteration index k*



(b) *Distribution of the bond lengths according to the Boltzmann distribution and the calculated distribution using regularized FSU.*



(c) *Distribution of the angles according to the Boltzmann distribution and the calculated distribution using regularized FSU.*



(d) *Distribution of the dihedrals according to the Boltzmann distribution and the calculated distribution using regularized FSU.*

Figure 4.4: *Sampling properties of the contour length, bond length, bending angles and dihedrals using regularized FSU with N = 9 (i.e. n=27),* $\varepsilon = 0.01$, $k_BT = 1$ *and* $\Delta t = 10^{-2}$.

potential), it was argued that the native state has a large basin of attraction [91]. The native confirmation can thus be accessed rapidly and over a rather wide temperature range and would appear to be kinetically two-state-like. For the later comparison, the β-type folded state of protein G is displayed in Figure 1 of Ref. [91]. All three

neutral residues are concentrated in a turn region. The hydrophobic residues in the branches at either side of this turn tend to be in close contact with each other while the hydrophilic residues point outwards. We refer to this state as *native* in the remainder.

Parameters in the potential terms

The parameters in (4.16) are the same as in Ref. [91]: $k^b_{ij} = 100$, $d_0 = 1$, $k^a_{ijk} = 20$ and $\theta_0 = 105^\circ$ in the bond length and angle potentials, $A_i = 0$ and $B_i = 0.2$ if two or more of the four particles are neutral (N) and $A_i = B_i = 1.2$ for all other cases in the torsion potential (4.20). Non-bonded potentials are given by (4.21) with $\varepsilon_h = 1$ and coefficients $C_{ij} = -D_{ij} = \frac{2}{3}$ if one particle is of the L type and the other of type L or B, $C_{ij} = D_{ij} = 1$ if both particles are of type B and $C_{ij} = 0$ and $D_{ij} = -1$ if either one of two particles is of type N.

Distance measure

The distance between a simulated and a reference state is determined using the common structure overlap function χ,

$$\chi = 1 - \frac{2}{N^2 - 5N + 6} \sum_{i=1}^{N-3} \sum_{j=i+3}^{N} \Theta(\epsilon_{\text{tol}} - |d_{i,j} - d^{\text{ref}}_{i,j}|), \tag{4.22}$$

where N is the number of particles, d_{ij} and d^{ref}_{ij} are distances between particle i and j in the simulated and reference state, respectively, and Θ is the Heaviside function. The structure overlap function is insensitive to chain rotation or translation. If $|d_{i,j} - d^{\text{ref}}_{i,j}| \leq \epsilon_{\text{tol}}$, particles i and j are assumed to be in a contact according to the reference state. Unless mentioned otherwise, our reference state is the native state shown in Figure 4.8i. We use $\epsilon_{\text{tol}} = 0.2$ [91].

Regularization parameter and time step

Before reviewing simulation results, we shortly consider the regularization parameter and the time step. In all simulations with RFSU we used either $\varepsilon = 10^{-3}$ or 10^{-2}, a choice that is based on the results of our previous analysis. A large time step provides an enhanced sampling of the potential energy surface. Initially, we considered $\Delta t = 10^{-2}$ for RFSU and found that the simulations are stable. In CLD, such a large time step causes instability and we were forced to simulate with a reduced $\Delta t = 10^{-4}$. Nevertheless, this cause for instabilities is also active in the initial stages of RFSU,

since the initial mobility $B_0 = I$ does not yet include proper scaling for the different modes. This phenomenon is clear in Figure 4.6a where one can observe irregular high peaks during the initial iterations of the simulation. In principle, one can remedy this behavior by preconditioning, i.e. determining a proper choice of $B_0 = J_0 J_0^T \neq I$. This approximate inverse Hessian B_0 can be determined by performing the S-QN method for $k_B T = 0$ (standard QN). Such a B_0 was found to enable stable simulation for an even increased $\Delta t = 10^{-1}$.

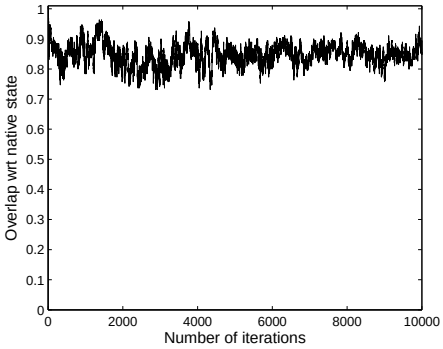
Starting configuration

The simulation pathway over the potential energy hypersurface naturally depends on its complex topography, making the observation of a folding event within the considered limits of simulation time sensitive to the starting state on this folding funnel. The starting state and influence of external factors in biological folding processes remain a matter of speculation, complicating the determination of realistic starting states. Many simulations of protein conformational changes or dynamics begin at the native crystal structure using the atomic coordinates measured from electron density maps obtained by X-ray diffraction methods. Other simulations start from more random conformations that often obey constraints, in that selected interatomic distances in the conformation have particular values. Sometimes this can be done analytically by choosing as random variables either the torsion angles, which always preserve bond length and bond angle values, or the rigid-body rotations and translations, which always preserve the internal structure of the rigid fragment. Methods for generating random structures that satisfy constraints on the distances between pairs of atoms or groups given upper and lower bounds also exist [110, 111]. Weak constraints on bond lengths, angles and torsions are also used in many coarse-grained descriptions, where random (coil) starting configurations are generated at elevated temperatures by presimulation, after disabling non-bonded interactions, or by simulated annealing (SA). The study of Veitshans [91] used a SA procedure to generate 100-300 independent states for protein G and showed that, for $T = 0.41 < T_F$, the fraction of unfolded chains decays roughly exponentially with time and vanishes around $t = 3000$. Here, we select specific starting configurations, the V-shape shown in Figure 4.8a, with upper and lower bounds for otherwise random bond length, angles and torsions. These structures are inspired by our focus on the collective dynamics and based on the identification of the turn region in the native state [91]. We expect that the domains at either sides of the turn point will collectively approach each other in RFSU to reach the native state. The large Φ of this state and the collective behavior that is required

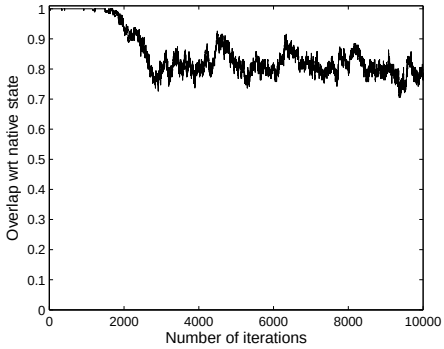
for folding suggest that this state is part of a region that is rather distant from the native state. Moreover, the fraction of folded proteins in CLD will depend on the cumulative Kramer's transitions rates for *all* potentials energy barriers between starting and native states. A direct comparison of our result to the folding rates obtained by Veitshans is therefore rather impossible [91]. Instead, we compare RFSU and CLD (with $B = \gamma I$) for the same (random) starting states.

Simulations for $T > T_\theta$

We start with simulations for $k_B T = 1$, which reduces to $T = 1$ in reduced units [91]. We note that this temperature is well above the collapse transition temperature $T_\theta = 0.78$. Our reason for this choice is that sampling at elevated temperatures increases the probability of barrier crossing. In Ref. [91], the averaged overlap function $\langle\chi\rangle$ was found to be $\langle\chi\rangle \approx 0.8$ for $T = 1$, where the averaging is carried out over pathways for different starting configurations. As mentioned, our starting configuration is the V-shaped conformation displayed in Figure 4.8a. For this structure, $\chi = 1$ and the domains at either side of the turning point are rather far apart. We used $B_0 = I$ and $\Delta t = 10^{-2}$ for S-QN, where the mobility B_{k+1} was determined by RFSU. For CLD, we used $\gamma = 1$ and $\Delta t = 10^{-4}$.
Figure 4.5 shows the evolution of the structure overlap function χ for both S-QN (a) and CLD (b). Although χ for S-QN is below the one for CLD for practically the complete range of time steps, their final (averaged) values are around the same $\chi \approx 0.8$, in good agreement the results of Veitshans *et al* [91]. Clearly, the chains exhibit in a random coil conformation during its evolution in both methods. The corresponding evolution of the potentials Φ for both simulations are shown in Figure 4.6. Although Φ falls off much faster for S-QN, it fluctuates at a somewhat higher range (values between 40 and 60) than the CLD simulations (between 20 and 40). Although this performance looks very similar, the information contained in Figures 4.5 and 4.6 is incomplete. To analyze the sampling in S-QN and CLD, we explored the potential well structure around the sampling path. In particular, we used conformations obtained at every 500 time steps in S-QN and CLD for quenching into local minima of Φ or *inherent structures* by a standard QN method with a backtracking line search. The rather flat wells and local minima of Φ are displayed in Figures 4.7a and 4.7b. Missing parts of the wells were added symmetrically. In Figures 4.7c (S-QN) and 4.7d (CLD) we show a correlation map of inherent states obtained for S-QN and CLD respectively. Using the overlap function χ as a measure for similarity between different inherent states, white squares denote fully uncorrelated structures ($\chi \approx 1$), while

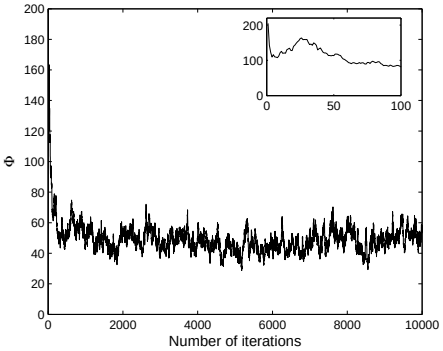


(a) *Overlap function with respect to a configuration from the native class using the regularized FSU with* $\varepsilon = 10^{-3}$ *and*$\Delta t = 10^{-2}$.

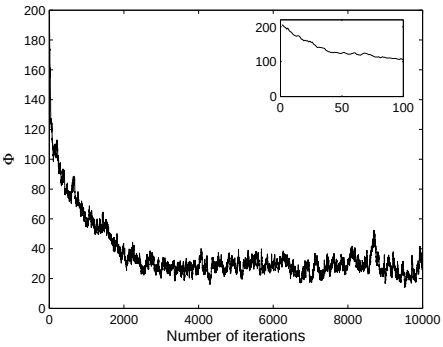


(b) *Overlap function with respect to a configuration from the native class using the conventional Langevin dynamics with* $\Delta t = 10^{-4}$.

Figure 4.5: *Comparison of the structural overlap functions* χ *obtained by regularized FSU and standard Langevin dynamics for simulations of a model protein with* $k_B T = 1$.

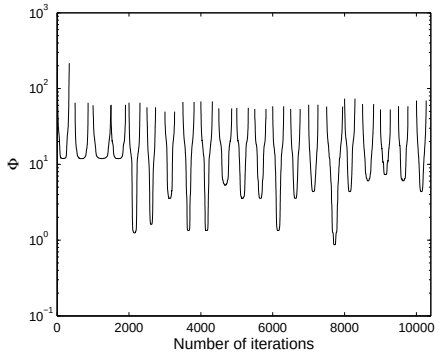


(a) *Evolution of the potentials* Φ *obtained by the regularized FSU with* $\varepsilon = 10^{-3}$ *and*$\Delta t = 10^{-2}$.

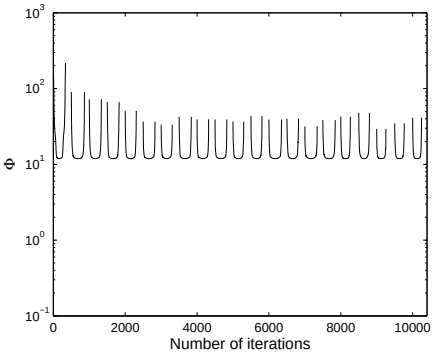


(b) *Evolution of the potentials* Φ *obtained by the conventional Langevin dynamics with* $\Delta t = 10^{-4}$.

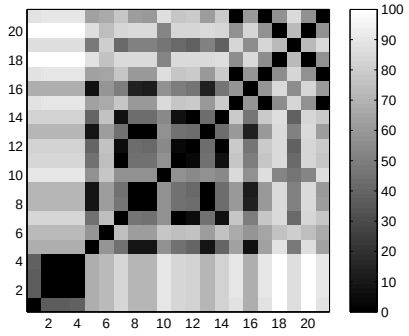
Figure 4.6: *Evolution of the potentials* Φ *for simulations of a model protein with* $k_B T = 1$.



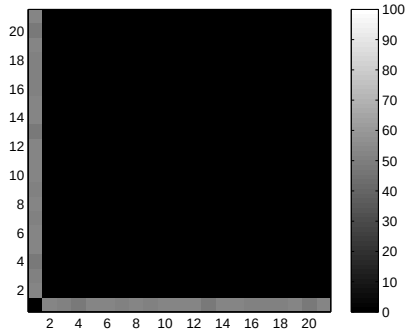
(a) *Local minima found during the simulation with regularized FSU.*



(b) *Local minima found during the simulation with standard Langevin dynamics.*

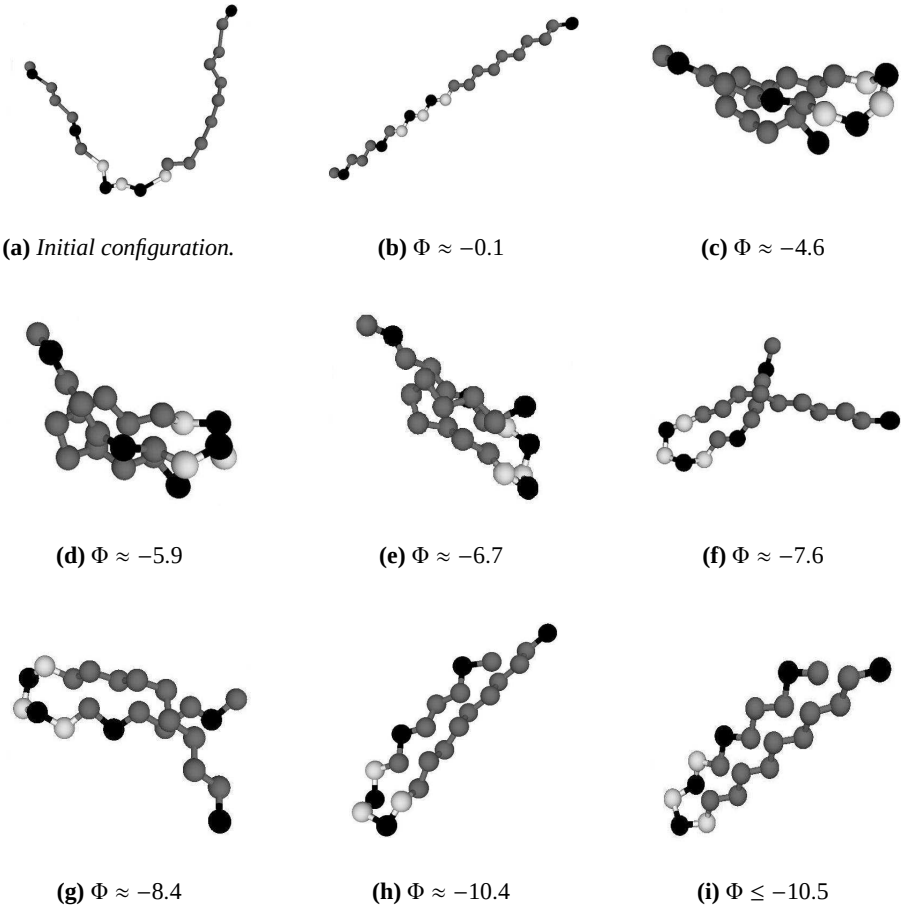


(c) *Correlations of the configurations corresponding to the local minima obtained during sampling with regularized FSU.*



(d) *Correlations of the configurations corresponding to the local minima obtained during sampling with standard Langevin dynamics.*

Figure 4.7: *Comparison of the local minima found by quenching after every 500 steps for the regularized FSU and the standard Langevin dynamics. In order to plot on a log-scale Φ_+ instead of Φ is shown, where $\Phi_+ = \Phi + c_+$, with $c_+ = 12$. The correlation matrices show the correlations between the configurations of the inherent state, the darker the square the more correlated the configurations are.*



(a) *Initial configuration.* **(b)** $\Phi \approx -0.1$ **(c)** $\Phi \approx -4.6$

(d) $\Phi \approx -5.9$ **(e)** $\Phi \approx -6.7$ **(f)** $\Phi \approx -7.6$

(g) $\Phi \approx -8.4$ **(h)** $\Phi \approx -10.4$ **(i)** $\Phi \leq -10.5$

Figure 4.8: *Configurations representing a typical structure from different classes found based on the correlation matrix.*

black squares denote fully correlated or similar structures ($\chi \approx 0$). The first entry in the correlation map is the starting structure shown in 4.8a and auto-correlation gives rise to a back diagonal.

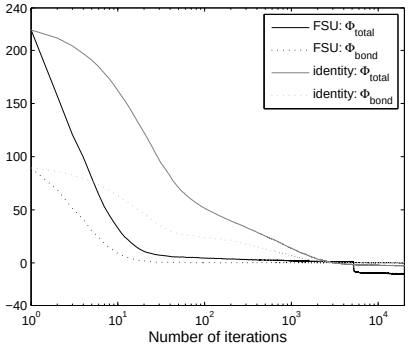
Correlated inherent structures were found to map one-to-one to distinct local minima of the potential energy Φ, ranging from $\Phi = -0.1$ to $\Phi = -11.1$. Based on this analysis, we classified the inherent structures for S-QN into several categories: $\Phi \approx -0.1$ ($k = 500, 1000, 1500$), $\Phi \approx -4.6$ ($k = 9000$), $\Phi \approx -5.9$

($k = 8500, 9500$), $\Phi \approx -6.7$ ($k = 4500$), $\Phi \approx -7.6$ ($k = 7000, 8000, 10000$), $\Phi \approx -8.4$ ($k = 3000, 5000, 5500, 6500$), $\Phi \approx -10.4$ ($k = 2500$) and $\Phi \leq -10.5$ ($k = 2000, 3500, 4000, 6000, 7500$). Representative structures are displayed in Figures 4.8b- 4.8i in the order of decreasing Φ. For CLD all local minima are the same, $\Phi \approx -0.1$ in Figure 4.7b, and only the inherent structure in Figure 4.8b was identified by the quenching procedure. These results show that a much larger part of the potential energy landscape is sampled by S-QN compared to CLD, which samples only very local, despite the apparently very similar χ in Figure 4.5. The structure in Figure 4.8i ($\Phi = -10.75$) is visually equal to the native structure of protein G in Figure 1 of Ref. [91]. We note that we identified a slightly different structure with even lower $\Phi = -11.13$. However, this structure belongs to the same class.

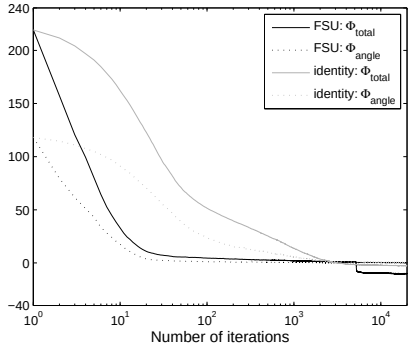
Simulations for $T < T_F$

We focus on the multiscale nature of our method by considering a temperature $T = 0.01$, much lower than the folding temperature. For such a low temperature, folding may become very slow in conventional Langevin dynamics [91]. To anticipate, we selected several V-shaped random initial configurations that are not extremely far from the native state but still have minimal overlap: χ is close to one. We carried out a preconditioning step for the determination of B_0 in RFSU, which prevents the irregularity at the beginning of the simulation (see insert Figure 4.6a). This preconditioning enables stable simulations with $\Delta t = 10^{-1}$. For a genuine comparison of folding rates in S-QN and CLD, an appropriate constant mobility γ in CLD should be determined. We can either determine this mobility based on physical properties, i.e. the effective friction coefficient, or relate to the inverse Hessian in the initial $\mathbf{x}_0$. The fastest mode for a quadratic potential, the one that sets the maximum time step in CLD, is associated with the largest eigenvalue of H. Using the calculated largest eigenvalue $\lambda_{\max}(B_0^{-1}) \sim O(10^5)$ of the approximation B^{-1} of $H + \epsilon I \sim H$ in $\mathbf{x}_0$ and taking into account the time step in S-QN, we obtain an equivalent time step Δt for CLD of approximately 10^{-6}. Since we found the scheme to be stable for $\Delta t = 10^{-4}$, we use this value instead. For each method (S-QN and CLD) we performed 10+ simulations using different starting configurations. From each of these two sets, we selected and analyzed the one that reaches the lowest Φ.

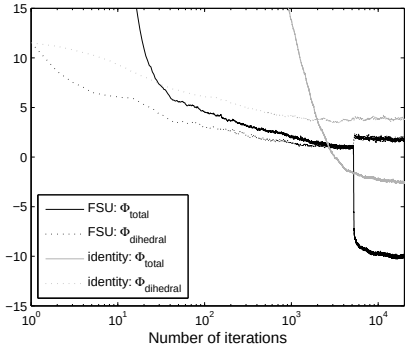
From Figure 4.9a, we observe a much faster decrease of the potential energy Φ_{total} for S-QN than for CLD, despite the fact that we have chosen a larger time step in CLD. For better understanding of the folding kinetics, we considered the evolution of the different contributions to the energy potential separately (see Figures 4.9a, 4.9b, 4.9c



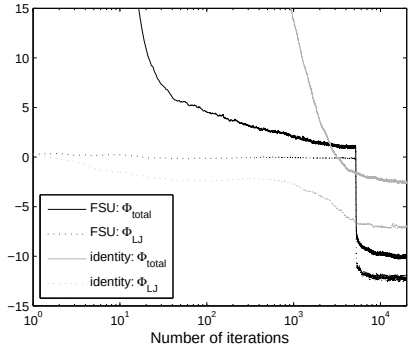
(a) *Comparision of the evolution of* Φ_{bond} *during simulations with the regularized FSU and the standard Langevin dynamisc.*



(b) *Comparision of the evolution of* Φ_{angle} *during simulations with the regularized FSU and the standard Langevin dynamisc.*



(c) *Comparision of the evolution of* $\Phi_{dihedral}$ *during simulations with the regularized FSU and the standard Langevin dynamisc.*



(d) *Comparision of the evolution of* Φ_{LJ} *during simulations with the regularized FSU and the standard Langevin dynamisc.*

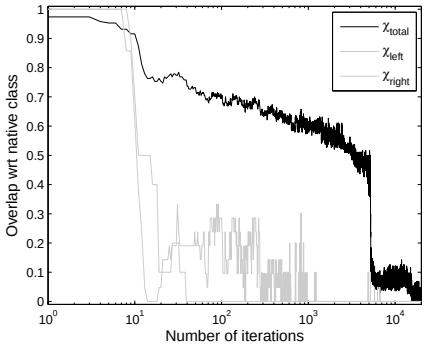
Figure 4.9: *Comparision of the evolution of the different potentials during simulation with the regularized FSU and the standard Langevin dynamisc.*

and 4.9d). Collective modes in S-QN are typically first observed in the potentials for intra-chain interactions (Φ_{bond}, $\Phi_{bending}$ and $\Phi_{dihedral}$) while the contribution of the non-bonded potential remains rather insignificant to later stages. The bond length and angles converge in $O(10)$ steps to their equilibrium values, see Figures 4.9a and 4.9b. From Figure 4.9c, one can observe that after $\sim$ 500 iterations the curves for $\Phi_{dihedral}$ and Φ_{total} have the same decreasing nature and eventually start to overlap

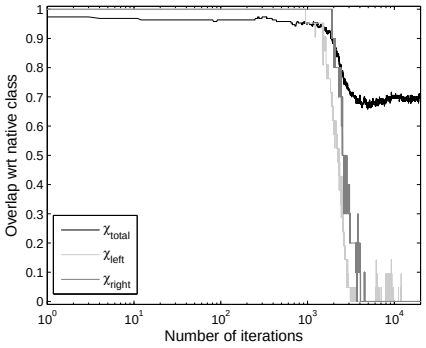
after $k \sim 2000$. At this stage, the equilibration of the dihedral angles is clearly the main contribution to the decreasing potential Φ_{total}, an observation that is supported by the other intra-chain potentials fluctuating around zero. Following this stage, there is a sudden drop in Φ_{total} after 5000 iterations, reflecting an equivalent drop in Φ_{LJ}, Figure 4.9d. The torsion potential responds to this sudden drop by a small increase followed by equilibration. We conclude that this reflects a collective chain collapse where different chain domains at either side of the turn move coherently, a process that is considerably accelerated when the long-range Lennard-Jones interactions become more significant. In the rearrangement process that follows, most probably concentrated in the turn region, the bond lengths and bond angles remains constant but the torsion angles have to adapt.

For CLD we see a gradual decrease of the intra-chain potentials and the non-bonded potentials, a signature of diffusion. The intra-chain potentials decrease much slower than for S-QN and the bonds and angles reach equilibrium values only after $O(10^3)$ steps. We note that large force constants in the bond terms, introduced by Veitshans *et al* [91] as an alternative to RATTLE or SHAKE, ensure some collectivity. Unlike S-QN, one can observe an immediate but small decrease of the non-bonded potential, Figure 4.9d, indicative of a very slow and sequential collapse of non-bonded interaction sites along the chain. From the values of torsion and non-bonded potentials at $k = 20000$ it is clear that this zippering-like process continues beyond the end of the simulation, and that the chain has not reached a stable state.

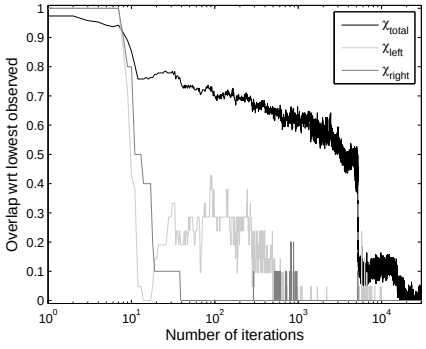
Next, we consider the structure overlap function $\chi = \chi_{\text{total}}$, using the structure of Figure 4.8i from the native class as a reference state. Although principal modes can in principle be determined as eigenvectors of B along the simulation pathway, domains that play a role in collective behavior can be anticipated from the start and the native state. We define partial overlap functions χ_{left} and χ_{right} that only consider two disjunct sub-domains LB_8 and BLB_3LB at either side of the turn region, respectively. We note that the partial overlap functions can both vanish even when the structure is not native. Sub-domains are chosen such that the turn region is excluded. In Figure 4.10a all structure overlap functions are combined for S-QN. It is clear that the left and the right domain assemble into a native conformation very fast. However, after χ_{left} and χ_{right} have vanished, they increase again at later stages and fluctuate between $0 - 0.3$ due to internal reorganizations. While the sub-domains rearrange (at least 70% stays conform to the native state), χ_{total} is gradually decreasing. After 1000 iterations, the rearrangements in the left and right domains damp out. The sudden drop in χ_{total} coincides with the drop in the non-bonded potential (see Figure 4.9d), when the LJ-



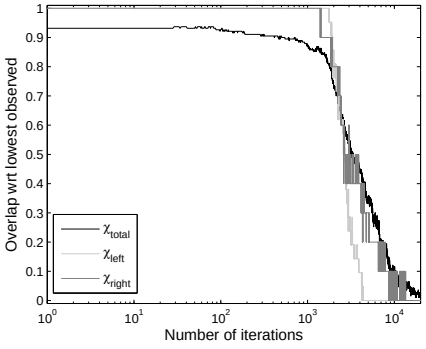
(a) *Evolution of the structural overlap with respect to a configuration from the native class as reference configuration during the simulation with regularized FSU.*



(b) *Evolution of the structural overlap with respect to a configuration from the native class as reference configuration during the simulation with standard Langevin dynamics.*



(c) *Evolution of the structural overlap with respect to the configuration corresponding to the lowest observed Φ during the simulation with regularized FSU.*



(d) *Evolution of the structural overlap with respect to the configuration corresponding to the lowest observed Φ during the simulation with standard Langevin dynamics.*

Figure 4.10: *Evolution of the structural overlap with respect to the configuration corresponding to the lowest observed Φ and with respect to the configuration from the native class during the simulation with regularized FSU and the standard Langevin equation.*

interactions become significant, and small rearrangements in subdomains. Once the native state is found (around $k = 15000$) only conformations in the native basin are sampled, since collective modes seize to contribute and the noise amplitude is low

for $T = 0.01$. For a comparison, Figure 4.10b contains the overlap functions for conventional Langevin dynamics. It is clear that χ_{left} and χ_{right} decrease at least one order of magnitude slower than in S-QN. Although the sub-domains eventually reach their native conformation, rearrangement of the chain as a whole is much slower and the chain does not reach the native conformation within simulation time.
The analysis of the overlap function with respect to the native class seems arbitrary for comparing the dynamics, since conventional Langevin dynamics does not necessarily converge to the same class as S-QN. Therefore, we introduce different reference states, in particular the state with the lowest observed Φ during each of the simulations (the lowest observed state) by S-QN or CLD. In Figures 4.10c and 4.10c $\chi_{\text{total}}, \chi_{\text{left}}$ and χ_{right} for S-QN and CLD are shown, respectively. The analysis for the S-QN is very similar to the one discussed earlier, since the lowest observed state is in the class of native states. For CLD, we see that both sub-domains rearrange into their lowest observed structure very slowly and that the signature of χ_{total} is highly correlated with these partial overlap functions. In particular, the formation of optimal structure (almost) coincides with the formation of optimal partial structure in the two sub-domains. This is a clear sign of non-cooperative behavior. As a reference, partial structure forms at an earlier stage and moves collectively as the result of long range interactions in S-QN.

4.4. Discussion

Force constants

Fixed bond length are often enforced in molecular modeling by computationally demanding SHAKE or RATTLE algorithms [107, 108]. An alternative constraint method, LINCS (Linear Constraint Solver) developed in 1997 [109], directly resets the constraints rather than the derivatives of the constraints and is up to four times faster than SHAKE. In Ref. [91], an alternative approach is employed to circumvent the computational demands associated with these methods. Stiff bonds are enforced by setting the force constants to a high $k = 100$. A trade-off of this approach is that the time step should be reduced. The automated scaling provided by S-QN gives rise to a fast convergence to the equilibrium chain length l_{eq} and 'soft' constraints on the bond length, as we showed for Rouse chains in one and three dimensions, even for a very small force constant $k = 1$. From this protein study, we found that the minimization of the intra-chain energy potentials, and also Φ itself, is rather insensitive to

a scaling of the potential

$$\Phi_\alpha = \alpha(\Phi_{\text{bond}} + \Phi_{\text{bending}} + \Phi_{\text{dihedral}}) + \Phi_{\text{non}}, \tag{4.23}$$

where α was varied between 0.1 and 1. Although they may vary in details, the signatures of both the intra-molecular potentials and the (partial) overlap functions do not depend on the considered values of α. In particular, the overlap functions is of the same form as illustrated in Figure (4.10a): first the partial structures (left and right part) are established, followed by a drop in χ_{total} indicating the increasing influence of long range interactions. Hence, the bond length is *automatically* constrained and the force constants in the potential can be significantly reduced. This observation provides tools for further optimization of performance.

Different starting configurations and regularization parameter choice

We accounted for the kinetic arrest that may occur at a very low temperature $k_BT = 0.01$ by considering a sharper V-shaped starting conformation than shown in Figure 4.8a for $k_BT = 1$. Moreover, the regularization parameter ε used in RFSU for $k_BT = 0.01$ is one order of magnitude larger ($\varepsilon = 10^{-2}$) than in the simulation with $k_BT = 1$ ($\varepsilon = 10^{-3}$) to avoid over-acceleration. Since the value of regularization parameter ε determines the upper bound to the acceleration of the collective modes, i.e. the damping with respect to the non-regularized case, a reduced ε speeds up the slowest collective modes that are important for bringing the structure of Figure 4.8a to the native configuration. We have also performed simulations with V-shaped starting configurations that have minimal overlap at the beginning of the simulations, $\chi \approx 1$, and found very similar signatures in the overlap functions and the contribution of the potential terms.

Partial structures

The partial structures LB_8 and BLB_3LB, important for the analysis of collective modes, were chosen based on the native chain conformation shown in Figure 4.8i, or alternatively Figure 1 of Ref. [91]. The question may arise whether our analysis is sensitive to this particular choice. For this reason, we systematically varied the partial structures. The evolution of the partial overlap functions for the left wing, LB_i with $i = 3..7$, converges to zero in the same way as for LB_8 in Figure 4.10a. These partial structures, however, remain in their native configuration after this stage, unlike LB_8 that was shown to fluctuate between 0 and 0.3. For the right partial structure, we also

calculated the partial overlap function for LB_3LB, B_3LB and B_2LB. Shortening the sub-domain leads to increased convergence to the native partial structure, although all convergence rates are very similar to the one for BLB_3LB ($O(10)$). Extending partial structures beyond the chosen ones affects the evolution of overlap functions. The fluctuations in the overlap function of LB_8 are caused by one of the B beads, that is part of the turn region. Including the turn region completely, by dividing the chain in two equally large sub-domains LB_9N and $(LN)_2BLB_3LB$, slows down the convergence rates of the partial overlap functions, since the turn region has to facilitate collective rotation and contraction by local rearrangements. Nevertheless, for RFSU these partial overlap functions posses the same features as in Figure (4.10a): the partial overlap functions vanish before the chain as a whole finds it native state.

4.5. Conclusions

We applied a regularized stochastic Quasi-Newton method to study a protein model. The new form of regularization was incorporated in the factorized update (FSU) algorithm used for the determination of the curvature-dependent mobility B in S-QN and resolves often occurring problems associated with a singular or ill-conditioned local curvature H. The regularized FSU (RFSU) scheme is based on determining an estimate of local curvature (B^{-1}) that converges to $H + \varepsilon I$ instead of H, and does not affect the efficiency of the original FSU scheme [81].

This work is part of a series of articles towards the development of a general efficient and stable method for thermodynamically consistent accelerated equilibration. In Ref. [80], we introduced the principles of S-QN and proved enhanced sampling performance and barrier crossing for 1-and 2-dimensional examples. In the original S-QN, standard Davidon-Fletcher-Powell (DFP) and Choleski decomposition was used for the determination of B and the noise amplitude, respectively. In Ref. [81], we introduced cost effective FSU and limited-memory (L-)FSU algorithms for determining L with $B = LL^T$ and analyzed the sampling performance and multi-scale nature for a simple but physically relevant Rouse chain with quadratic potential. Here, we regularized S-QN via RFSU and applied the resulting method for a previously developed coarse-grained model of a rather short protein with a β-type folded state. We developed an efficient two-step approach: the enhanced sampling gives rise to many inherent states, i.e. local minima on the energy landscape, when simulating above the collapse temperature, and, below the folding temperature, the cooperative modes of a protein can be efficiently determined and considerably speed up the folding process

when compared to conventional Langevin dynamics.
A detailed analysis of the S-QN results shows that the folding pathway can be divided in local and collective parts. Bond length, bond angles and torsions are equilibrated in this order, a feature that is rather insensitive to the force constants used in the potentials. After this equilibrium is established, a sudden drop in the total potential occurs, corresponding to the non-bonded interactions becoming more significant. This sudden drop coincides with the sudden drop in the total overlap function, that provides a distance measure with respect to the native state. The overlap function also shows that partial structures, i.e. carefully chosen sub-domains of the protein, form prior to this drop and move collectively to reach the native state. This collective dynamics is absent in conventional Langevin dynamics, where native contacts form as a result of (local) diffusion.
The considered coarse-grained protein is rather short and was previously determined to be a good folder [91]. Moreover, the native structure was known from this earlier study, although we identified a slightly different structure with even lower potential energy. It is tempting to use the two-step approach for the determination of inherent structure and principle mode analysis of proteins with direct biological function. Crystal structures can serve as input for the determination of inherent structures of biological relevance, and inherent structures can be analyzed for the class of proteins that are not easily crystallized, for instance membrane proteins. Realistic coarse-grained parameters for these proteins can in principle be determined by systematic coarse-graining procedures and external factors (for instance, chaperones) can be explicitly included in the S-QN method. We leave such study for future work.