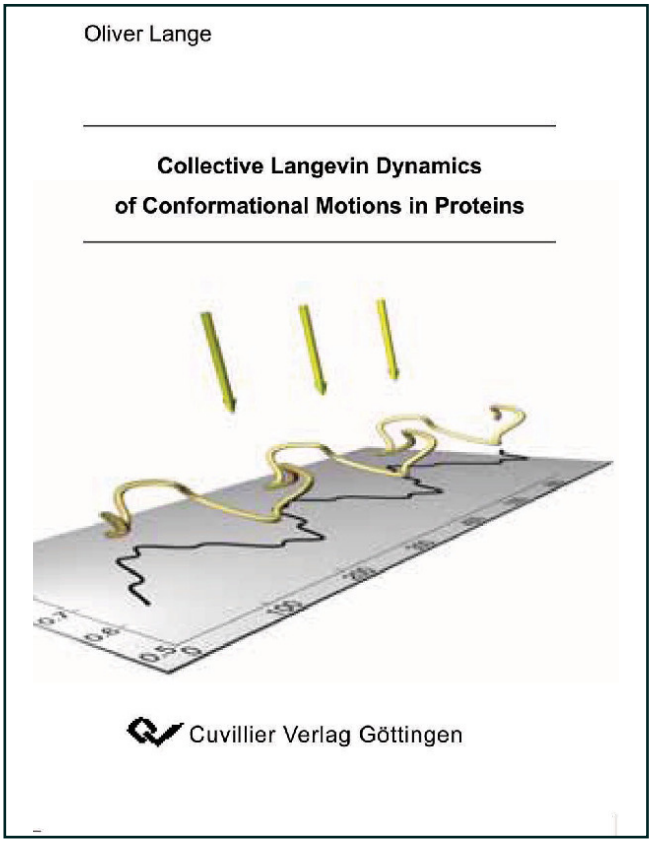
Oliver Lange (Autor)

Collective Langevin Dynamics of Conformational Motions in Protein



https://cuvillier.de/de/shop/publications/2304

The important thing in science is not so much to obtain new facts
as to discover new ways of thinking about them
— Sir William Bragg

Chapter 1
Introduction

Proteins are biological macromolecules, which are mainly composed from polymeric chains of amino acids[1]. They are involved in a diversity of processes in living organisms. Although some play a mere structural role (e.g., collagen in tissues, or α-keratin in hair), the function of most others depends crucially on their dynamics. While for the many examples of motor proteins (e.g., kinesin and F1-ATPase) the connection to dynamics is obvious, the dynamics also plays an important role if primary function is not mobility itself. For example, the ability to change conformation is essential for the function of proteins involved in signal transduction or transport, for molecular recognition, e.g., in the immune system, and for the function of numerous enzymes[1]. In many enzymes, for instance, conformational changes serve to enclose the substrate, thereby preventing its release from the protein and optimally positioning it for the protein to perform its function, as in lysozyme.

To understand the mechanisms of protein function is an intriguing and formidable task. Although remarkable progress has been made in past decades, and despite the number and quality of available methods has been tremendously increased, most mechanisms are not understood on a physical basis, which would require models based on first principles allowing for a quantitative comparison with experimental results.

Experimental techniques made remarkable progress to unravel protein structures (e.g., Xray crystallography[2, 3] and nuclear magnetic resonance spectroscopy (NMR)[4, 5]) and, furthermore, even allow to probe dynamics (NMR relaxation[6], electron paramagnetic resonance (EPR)[7], neutron scattering[8, 9], as well as fluorescence spectroscopy[10]). In some instances different functional states of proteins were structurally characterized by trapping them in certain substates[11]. Furthermore, time-resolved Xray diffraction[12, 13] allows to follow the conformational protein motion with picoseconds time resolution. Wide-spread use of the latter two techniques is impeded, though, by the massive experimental effort involved.

In comparison to this tremendous experimental progress and the enormous variety of available techniques the theoretical treatment of protein dynamics strikes as underdeveloped. Only computer simulation techniques, and especially molecular dynamics (MD) simulations at atomic resolution,

have been applied with noteworthy success to elucidate functional processes in recent years[14]. Therefore, advancements of theoretical methods and concepts are urgently required.

As will be described in Chapter 2, classical MD is an atomistic simulation method, which treats each atom as point mass and describes the interaction between atoms with simple force terms. Trajectories are generated by integrating Newton's equations of motions. It operates in the full $3N$ dimensional configurational space of the protein and the surrounding solvent molecules (where N is the number of atoms). The large number of pair-wise interactions to be evaluated and the short time steps enforced by the fastest motions entail very long computation times, which limits MD at present to systems of $10^5 - 10^6$ atoms and to timescales of several 100ns. Unfortunately, apart from a few exceptions, relevant biological processes, such as the gating of ion channels, allosteric interactions, ligand binding, molecular recognition, chemo-mechanical energy conversion and many more, occur on microsecond to seconds time scales, and thus are currently far out of reach for conventional MD.

This holds true despite considerable efforts to speed up the computations, particularly of the long-range Coulomb forces. Recent developments include efficient methods such as multiple step algorithms[15, 16, 17, 18, 19, 20], fast multipole methods[21, 22, 23, 24], and Ewald summation techniques[25]. Also, the use of constraints[26, 27, 28] helps to increase efficiency. Still, however, processes on time scales of microseconds and beyond can only be studied by resorting to certain 'tricks' to enhance sampling by speeding up conformational motions, as reviewed in Ref. [29]. Unfortunately, this kind of accelerated sampling necessarily implies loss of dynamical information and often loss of thermodynamical accuracy as well[29].

Statistical mechanics is the appropriate theoretical framework to understand the dynamics of many-particle systems such as proteins. One considers a macroscopic state as statistical ensemble of a large number of replica of a microscopic system, which evolve independently from each other. Macroscopic observations of relevant degrees of freedom are obtained by averaging the remaining ones over the statistical ensemble. The applied averaging yields energetics, which are influenced by entropic contributions, and hence free energies need to be considered. Seen from this perspective, protein function and the corresponding highly controlled conformational motions are driven by free energy differences between different substates of the solvated protein.

MD simulation, however, is not intrinsically a statistical mechanics approach, since it describes protein dynamics from a microscopic point of view. Rather, it is used as a 'brute-force' method to generate statistical ensembles. Although a statistical mechanics treatment can be attached to the MD results, this modus operandi impedes profiting from the elegance of this framework. It remains thus challenging to 'go the whole way' and consistently treat relevant degrees of freedom of protein dynamics with statistical mechanics.

In this thesis we advance the methodology beyond conventional 'brute force' MD by applying statistical mechanics to gain a drastic *reduction of the large number of degrees of freedom.* This implies two steps. First, to identify few appropriate slow and relevant degrees of freedom[30], which serve to define a reduced active space within which the dynamics is evolved without explicit treatment

of the remaining orthogonal fast degrees of freedom. Second, to derive suitable equations of motion for these slow degrees of freedom.

As an illustration of our approach consider the well-known dimension reduced treatment of the motion of a Brownian particle, which is also based on a separation of slow and fast degrees of freedom. A Brownian particle is a large solute particle immersed in a fluid of much smaller particles, e.g., water. Its macroscopic erratic movement is the combined result of a large number of collisions with fluid particles. Because the motion of the macroscopic particle is much slower than that of the fluid particles, one can consider the slow and fast motions as uncorrelated. This justifies to treat the solvent coordinates as irrelevant and thus to replace their influence on the slow degree of freedom by a random force, which is memory free due to the separation of timescales. In contrast to our work described below, the trajectory of the Brownian particle is a random walk and can thus be described by a Markov process, i.e., its future evolution does not depend on its past, because (a) the random force is memory free and (b) the motion is overdamped.

To apply this concept — replacing fast degrees of freedom by a random force — to our case we have to be aware of the differences though. First, it is not at all clear how to select the slow degrees of freedom for the internal motion of a protein. All involved particles, regardless if constituting solvent or protein, are atoms of similar mass, which move with comparable speeds. Second, there will be no clear separation between fast and slow degrees of freedom due to the continuous spectrum of time scales covered by protein dynamics. An important consequence is that the random force contains memory effects. A third difference is that the motion is not overdamped, such that inertia effects matter. Thus, our treatment will have to account for this non-Markovian character of the slow dynamics.

The absence of canonical slow degrees of freedom has led to a diversity of phenomenologically motivated selections of the active space. These include implicit solvent[31], combined atom or bead models[32, 33, 34, 35], and the treatment of polypeptides as chains of stiff 'platelets', for which only ψ-φ backbone angles are retained as explicit degrees of freedom[36, 37]. A somewhat related approach is the gaussian network model[38].

However, by restricting the model to certain atoms or groups of atoms and omitting others, only a very small subset of all possible collective degrees of freedom is considered. One may, therefore, expect to derive improved dimension-reduced descriptions of protein dynamics by dropping this empirical restriction and considering as degrees of freedom m fully general functions $c_i = f_i(\mathbf{x}_1, \ldots, \mathbf{x}_N), \quad i = 1 \ldots m$, of the atomic positions $\mathbf{x}_j$. Linear f_i are widely considered, e.g., within the framework of principal components analysis (PCA)[39], which is often used to systematically derive slow and relevant (essential) collective degrees of freedom from MD simulations or structural ensembles[40]. Here we consider both, linear and non-linear collective degrees of freedom.

The general framework that allows to reduce the full dynamics of all atomic degrees of freedom to dynamics of the selected (collective) degrees of freedom is provided by the projection-operator formalism of Zwanzig and Mori[41, 42]. The resulting generalized Langevin equation (GLE)[43, 44, 45, 46, 47, 48] governs non-Markovian dynamics due to its generalized dissipative term, which is a

convolution of the *memory kernel* with past velocities. We will show how the GLE is derived from Newton's equation of all degrees of freedom by separating the overall motion with the projection-operator into (a) an ensemble-averaged motion on the free energy surface, governed by the *potential of mean force,* and (b) a deviation from these average dynamics.

Combining these two concepts, generalized collective degrees of freedom and dimension reduced dynamics, we here develop the framework of Collective Langevin Dynamics (CLD), which describes protein dynamics in collective coordinates. The projection-operator formalism is used to derive the necessary parameters for the GLE, i.e., an appropriate potential of mean force and memory kernels from short MD simulations. Thereby, all parameters are systematically obtained from first principles, which allows to automate parameter extraction. By construction, there are no general parameters which hold for all proteins, but parameters need to be specifically extracted for the chosen molecular system and the selected set of collective coordinates. The low number of degrees of freedom will allow a computationally efficient generation of trajectories, thereby rendering microseconds timescales accessible.

The main tasks which need to be addressed in this thesis are (1) identification of suitable conformational coordinates, (2) extraction of memory kernels and (3) construction of a suitable free energy landscape from MD simulations, and (4) evaluation of CLD accuracy and performance.

Note that it is a huge task to develop CLD to full maturity, such that we here only attempt the first steps, which we outline below.

(1) Extraction of relevant degrees of freedom

Selection of suitable collective degrees of freedom crucially affects the strength and persistence of memory effects as well as the resolution of conformational states. Thereby, this choice determines the significance of the resulting CLD model for functionally relevant dynamics. Thus, we aim for collective modes which are as slow as possible. Moreover, the active subspace is ideally uncorrelated to those remaining fast degrees of freedom, which are not treated explicitly.

A well established method to identify functional relevant modes in MD trajectories is principal component analysis (PCA)[49, 39, 50, 51, 40]. Therefore, it is a natural choice to consider PCA as a candidate here. It selects those collective degrees of freedom which contribute most to the atomic motion seen in the trajectory by diagonalizing the covariance matrix of atomic displacements.

Whether and to what extent a separation of timescales can be achieved by application of PCA has not yet been systematically assessed. Furthermore, it is not clear, and subject to ongoing discussions[52], whether principal components extracted from short MD simulations can serve to describe protein dynamics at long time scales sufficiently well. In Chapter 2, we will shortly review the theory of PCA and address these questions.

Unfortunately, PCA does not yield fully uncorrelated collective modes[50], because the covariance matrix detects only *linear* correlations. Although the remaining *non-linear* and *multi-coordinate* correlations do not impede using principal components within CLD we might be able to advance the

method by extracting coordinates that are correlated to a lesser extent.

Therefore, we will introduce in Chapter 3 the (Shannon) mutual information[53], which detects *any* correlation. Based on this measure we will develop in Chapter 4 Full Correlation Analysis (FCA), which extracts maximally uncoupled coordinates by minimizing the mutual information of the configurational ensemble.

That the extraction of collective coordinates relies heavily on correlation triggered the wish to find an experimental access to this observable. This would allow a direct verification of collective modes, and possibly an experimental access to collective modes without the requirement of an MD simulation.

Experiments probe correlations in the motion of *atoms* in three dimensional space[54, 55, 56], in contrast to the previously considered correlations between *coordinates*. In Chapter 3 we will show that the established method[57, 58] to quantify such correlations suffers from considerable inconsistencies, and thus misses over 50% of correlations. Since this impedes any meaningful comparison of this observable with experiments, we propose to apply mutual information also to this problem and define a generalized correlation coefficient. In this way we avoid not only the inconsistencies of the previous measure but also detect *non-linear* correlations.

Having then established a solid grasp of correlations on the simulation side, we will compare these with experimental data. A recently reported NMR relaxation experiment promised to probe correlated motions in proteins[59]. Whether the results were really related to correlated motion, however, could not be tested by experiment alone. Therefore, we address this issue by means of MD simulations in Chapter 5.

(2) Extraction of Memory Functions

Extraction of memory kernels from MD simulations is still a challenging problem. Despite considerable efforts, a generally accepted approach has not yet emerged[60, 61, 62, 63]. Thus, we will study different memory extraction schemes, and evaluate their performance within the framework of CLD in Chapter 7.

To our knowledge, all existing algorithms are based on either the Memory equation[62, 64, 65, 66, 67, 68, 61, 69], or on a direct relation of the memory kernel with force autocorrelation functions[63, 70]. We assess both approaches, which have different merits and flaws in the context of CLD. Because exploiting the Memory equation requires solution of an inverse problem, we need to study regularization techniques for its stable and robust numerical solution.

(3) Free energy surface

Free energy surfaces of the conformational coordinates can be estimated by molecular dynamics sampling[71, 72]. More efficient, however, are enhanced sampling techniques[29], for instance, multicanonical methods (e.g., replica exchange MD (REMD)[73]), smart Monte Carlo (SMC)[74], or umbrella sampling[75]. These techniques are complementary to CLD, because they yield canonical

ensembles, but do not yield dynamical information. CLD, on the other hand, yields proper dynamical information, but relies on already known canonical ensembles.

Due to the abundance of available techniques it was not necessary to treat this topic in detail.

(4) Evaluation of CLD models

Assessment of the quality of the obtained dimension-reduced description is non-trivial in itself. Clearly, direct comparison of the observed CLD trajectory with explicit (deterministic) MD simulations is not meaningful, because the underlying GLE governs a stochastic process and because the dynamics is chaotic. Rather, suitable observables such as averages over many realizations of the stochastic process, or time averages such as time correlation functions, transport coefficients, or transition rates should be used[76].

However, one has to take care not to check observables which were used to parameterize the CLD model, rendering the selection of test-observables a delicate choice. Velocity autocorrelation functions, for example, are used to extract memory kernels from MD simulation and, thus, do not represent a rigorous test of CLD. In Chapter 8, we use conformational transition rates as observables, which are fully unrelated to the input - yet statistically meaningful. They are compared to reference rates obtained from a long explicit MD simulation. Additionally, we compare positional autocorrelation functions as a probe of long time correlations, because these are not resolved by the velocity autocorrelation functions used as input.

I do not fear computers. I fear the lack of them.
— Isaac Asimov

Chapter 2
Principal Component Analysis

Principal component analysis (PCA) is a well-established technique for reducing dimensionality. Its applications include data compression, image processing, data visualization, exploratory data analysis, pattern recognition and time series prediction[77]. In this chapter we elucidate whether PCA can be applied to extract from short MD simulations slow slow collective degrees of freedom to treat protein dynamics within the proposed framework of collective Langevin dynamics (CLD).

For analysis of protein dynamics principal component analysis (PCA)[49, 39, 50] is an established method based on the notion that the biggest positional fluctuations occur along collective degrees of freedom. This was first realized by normal mode analyses of small proteins[78, 79, 80]. In normal mode analysis, the potential energy is approximated harmonically and the collective modes are obtained by diagonalizing the Hessian matrix in a local energy minimum. PCA, and the related quasi-harmonic analysis[81, 82, 83, 84] and singular value decomposition[85, 86], have shown that even beyond the harmonic approximation protein dynamics are dominated by few collective modes. In particular, these methods showed that it was generally possible to describe about 90% of the total atomic displacement of a protein with only 5-10% of the collective degrees of freedom[50, 87]. This has led to the concept of the *essential* subspace, which is spanned by a small number of the PCA modes with the highest fluctuational amplitudes. It could be shown that in this way PCA separates protein dynamics in two kinds of modes. The fluctuational distribution of the non-essential (small amplitude) modes is well approximated by a Gaussian. Thus, these modes are called *quasi-harmonic*, and are considered to constitute near constraint degrees of freedoms[88, 89]. For the large amplitude (essential) modes, on the contrary, this approximation is inaccurate, such that they are called *anharmonic*[88, 89]. It was argued that only the latter describe functional relevant motion, since the anharmonicity results from rare transitions between multiple minima, while the motion within the minima is rather quasi-harmonic[88, 89, 90].

As a consequence, the dynamics in the essential subspace, denoted as essential dynamics, are often in the primary focus of computational studies[91, 92, 93, 94, 95], enhanced sampling techniques[96, 97, 98] or simple models of protein dynamics[90, 99, 100, 101]. To investigate whether the essential PCA modes are suitable to serve as conformational coordinates within the CLD framework, we need

to establish that (a) the timescales between essential and non-essential PCA modes must be partially separated, and (b) the essential modes must describe also the long time dynamics sufficiently well.

Essential PCA modes indeed describe slow motion, because via the equipartition theorem the large amplitude modes are connected with a slow effective frequency $\omega_i^{\text{eff}} = \sqrt{\frac{kT}{\langle c_i^2 \rangle}}$[87]. However, the timescale separation needs to be investigated more systematically, because a slow effective frequency does not rule out minor but significant fast contributions to the dynamics of an essential mode. Therefore, we are going to analyze in Sec. 2.4 the power spectra of all principal modes, in order to see whether and under which conditions PCA is able to filter out purely slow motions.

Furthermore, we need to establish that the essential subspace obtained from a short MD simulation describes a considerable and sufficient amount of the overall protein motion observed on long time scales. This question about the convergence of principal modes has led to considerable dispute. Amadei et al. advocated a fast convergence[102], whereas Balsera et al. strongly questioned the suitability of principal modes to describe protein dynamics on long time-scales[52]. This controversy stems from a different perception of convergence of principal components. Amadei et al. found in 2 ns simulations evidence on a remarkably stability of the directions of single eigenvectors[102]. Balsera's rejection of principal modes, however, was mainly based on the slow convergence of the fluctuational amplitudes[52]. Because these amplitudes are not important for the use of the PCA modes within the CLD framework, the findings of Amadei et al. are more relevant to us.

Nevertheless, both antagonistic studies are based on short simulations — due to the limited computer power at their time — rendering the judgment of the suitability of principal modes for description of protein motion on long time scales a precarious extrapolation. Therefore, we resolve this question by analyzing in Sec. 2.5 how well principal components computed from short MD simulations can describe the dynamics observed in a much longer (i.e., 450ns) MD simulation of crambin. Besides, we depart from Amadei's work not only by means of much longer simulation time, but also by adopting a new measure of stability (cf. Sec. 2.3.3) that is particular suited to answer our question.

In the subsequent section we introduce PCA as maximization of fluctuational amplitude and report its basic properties. Since the following investigations are based on extended MD simulation, we sketch its principles in Sec. 2.2 and use the opportunity to introduce in Sec. 2.3.1 all simulation systems used within this work.

2.1 Theory of principal component analysis

We shortly review the most common derivation of PCA to illustrates its basic properties. PCA is applied to ensembles of protein structures $\left\{\mathbf{r}^{(k)}\right\}_{k=1\ldots M}$, where $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)^{\mathrm{T}}$ denotes the positions of its N atoms in three dimensional space and angular brackets denote the ensemble average $\langle f(\mathbf{r}) \rangle = M^{-1} \sum_{k=1}^{M} f\left(\mathbf{r}^{(k)}\right)$. PCA aims at finding linear orthogonal projections $c_i = \mathbf{a}_i^{\mathrm{T}} (\mathbf{r} - \langle \mathbf{r} \rangle)$, where the $\mathbf{a}_i$ are unit-vectors, such that the cumulative variances of the projections to the first m modes, $\sigma_m^2 = \left\langle \sum_{i=1}^{m} c_i^2 \right\rangle$, are maximized for all $m = 1 \ldots 3N$. The c_i are then called *principal*

components.

Now we show that the mode vector $\mathbf{a}_i^{\mathrm{T}}$ corresponds to the normalized eigenvector associated with the i-th largest eigenvalue of the covariance matrix of atomic displacements,

$$\mathbf{C} = \langle (\mathbf{r} - \langle \mathbf{r} \rangle) (\mathbf{r} - \langle \mathbf{r} \rangle)^{\mathrm{T}} \rangle.$$

Without loss of generality it is assumed that $\langle \mathbf{r} \rangle = 0$, i.e., $\mathbf{C} = \left\langle \mathbf{r}\mathbf{r}^{\mathrm{T}} \right\rangle$. First, the variance of the first principal component c_1 is maximized, i.e., $m = 1$, and, subsequently, the other principal components are obtained by a simple repetition of the steps with $m > 1$.

A maximizer $\mathbf{a}_1$ of the variance $\sigma_1^2 = \left\langle \mathbf{a}_1^{\mathrm{T}} \mathbf{r} \left(\mathbf{a}_1^{\mathrm{T}} \mathbf{r} \right)^{\mathrm{T}} \right\rangle = \left\langle \mathbf{a}_1^{\mathrm{T}} \mathbf{r}\mathbf{r}^{\mathrm{T}} \mathbf{a}_1 \right\rangle = \mathbf{a}_1^{\mathrm{T}} \left\langle \mathbf{r}\mathbf{r}^{\mathrm{T}} \right\rangle \mathbf{a}_1$ under the constraint $\|\mathbf{a}_1\|^2 = 1$ must solve the normal equation

$$0 = \left\langle \mathbf{r}\mathbf{r}^{\mathrm{T}} \right\rangle \mathbf{a}_1 - \lambda \mathbf{a}_1, \tag{2.1}$$

which is obtained by differentiation with respect to $\mathbf{a}_1$ of the Lagrangian function

$$L_\lambda (\mathbf{a}_1) = \mathbf{a}_1^{\mathrm{T}} \left\langle \mathbf{r}\mathbf{r}^{\mathrm{T}} \right\rangle \mathbf{a}_1 - \lambda \left(\|\mathbf{a}_1\|^2 - 1 \right),$$

where λ denotes a Lagrange multiplier. Eq. (2.1) yields the necessary condition that the maximizer $\mathbf{a}_1$ has to be an eigenvector of the covariance matrix $\left\langle \mathbf{r}\mathbf{r}^{\mathrm{T}} \right\rangle$ corresponding to an eigenvalue λ. Moreover, $\lambda = \mathbf{a}_1^{\mathrm{T}} \left\langle \mathbf{r}\mathbf{r}^{\mathrm{T}} \right\rangle \mathbf{a}_1 = \sigma_1^2$, i.e., the maximum variance is given by the largest eigenvalue and its corresponding eigenvector.

After the first $d-1$ projection vectors $\mathbf{a}_i$ have been identified, the subsequent vector $\mathbf{a}_m^{\mathrm{T}}$ is obtained by maximizing the cumulative variance $\sigma_m^2 = \left\langle \sum_{i=1}^{d} c_i^2 \right\rangle = \sigma_{m-1}^2 + \left\langle \mathbf{a}_m^{\mathrm{T}} \mathbf{r} \left(\mathbf{a}_m^{\mathrm{T}} \mathbf{r} \right)^{\mathrm{T}} \right\rangle$ keeping the first $m-1$ vectors, and thus the variance σ_{m-1}^2, fixed. This yields, repeating the steps above, that $\mathbf{a}_m^{\mathrm{T}}$ is the eigenvector of $\mathbf{C}$ corresponding to the m-largest eigenvalue.

An illustrative alternative is to define principal components as the projections c_i for which the reproduction error $\left\langle \|\mathbf{r} - \hat{\mathbf{r}}_m\|^2 \right\rangle$ is minimized. The m-dimensional reproduction $\hat{\mathbf{r}}_m = \mathbf{A}^{\mathrm{T}} (c_1, c_2, \ldots, c_m, 0, \ldots, 0)^{\mathrm{T}}$, where the rows of $\mathbf{A}$ are formed by the vectors $\mathbf{a}_i^{\mathrm{T}}$, is the motion in the original $3N$-dimensional space, which can be described using only m degrees of freedom c_i. The minimization of the reproduction error is equivalent to maximization of σ_m^2

$$\left\langle \|\mathbf{r} - \hat{\mathbf{r}}_m\|^2 \right\rangle = \left\langle \mathbf{r}^{\mathrm{T}}\mathbf{r} - \mathbf{r}^{\mathrm{T}}\hat{\mathbf{r}}_m - \hat{\mathbf{r}}_m^{\mathrm{T}}\mathbf{r} + \hat{\mathbf{r}}_m^{\mathrm{T}}\hat{\mathbf{r}}_m \right\rangle,$$

which yields with $\mathbf{r} = \mathbf{A}^{\mathrm{T}} (c_1, c_2, \ldots, c_{3N})^{\mathrm{T}}$ due to $\left\langle \mathbf{r}^{\mathrm{T}}\hat{\mathbf{r}}_m \right\rangle = \left\langle \hat{\mathbf{r}}_m^{\mathrm{T}}\mathbf{r} \right\rangle = \left\langle \hat{\mathbf{r}}_m^{\mathrm{T}}\hat{\mathbf{r}}_m \right\rangle = \left\langle \sum_{i=1}^{m} c_i^2 \right\rangle = \sigma_m^2$ that

$$\left\langle \|\mathbf{r} - \hat{\mathbf{r}}_m\|^2 \right\rangle = \left\langle \mathbf{r}^{\mathrm{T}}\mathbf{r} \right\rangle - \sigma_m^2.$$

Thus, using PCA the dimension reduced description of the protein dynamics has the smallest reproduction error that is possible to achieve with a given number m of collective degrees of freedom.