Available online at www.sciencedirect.com
ScienceDirect
Physica A 376 (2007) 394–400
PHYSICA A
www.elsevier.com/locate/physa

The dynamics of inter-residue distances in bovine pancreatic trypsin inhibitor

Wael I. Karain[a,*], Basem Ajarmah[b], Nael I. Qaraeen[b]

[a]Birzeit University, Department of Physics, P.O. Box 14, Birzeit, Palestinian Territory
[b]Birzeit University, Department of Computer Science, P.O. Box 14, Birzeit, Palestinian Territory

Received 12 July 2006; received in revised form 4 October 2006
Available online 15 November 2006

Abstract

Time series for the distance between selected residues surrounding the active site in bovine pancreatic trypsin inhibitor (BPTI) protein are analyzed using a one-dimensional Langevin-type stochastic difference equation. The time series are extracted from molecular dynamics simulations performed on the protein in vacuum and water. The deterministic and stochastic contributions to the time series show interesting behavior for certain residue pairs only, and only in solvated molecular dynamics simulations.

Keywords: Langevin equation; Protein conformations; Time-series analysis

1. Introduction

Proteins are examples of complex stochastic systems [1]. There are indications that the internal protein dynamics are a ‘‘slave ‘‘to the properties of the surrounding environment, most notably the surrounding solvent [2]. In addition, fast random changes in protein structure take place on the pico- and nano-scales [3]. The dynamical behavior of proteins also displays large random fluctuations beginning at a transition temperature of 180–230 K [4,5]. These time scales are easily accessible using molecular dynamics simulations. This offers a way of probing the dynamics involved in structural changes in the protein. And since protein function depends on its structure (conformation), and its ability to switch between different conformations, it is important to understand how exactly this switching takes place. In this work, we analyze a set of time series of inter-residue distances for a selected number of residues surrounding the active site in the protein bovine pancreatic trypsin inhibitor (BPTI). This is done using a recently developed method for extracting the deterministic and stochastic contributions from time series [6,7]. This method has been used to analyze many interesting physical systems [8–13]. It is based on the estimation of the Langevin differential equation [14,15]

$$\frac{\mathrm{d}}{\mathrm{d}t}X(t) = g(X(t), t) + h(X(t), t)\Gamma(t), \tag{1}$$

*Corresponding author. Tel.: +970 972 26717824.
E-mail address: wqaran@birzeit.edu (W.I. Karain).

0378-4371/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2006.10.070

where $X(t)$ is the state of the system at time t, the nonlinear function g provides the deterministic change, the function h gives the amplitude of the stochastic part, and $\Gamma(t)$ provides an uncorrelated white noise term with an average value of zero.

2. Computational methods

The molecular dynamics simulation and analysis were performed using NAMD [16] and VMD[17]. The starting X-ray structure for the simulation (6PTI) was obtained from the Protein Data Bank [18]. The Charmm force field was used [19,20].

Periodic boundary conditions were used. The solvent box volume was 44.9 Å × 40.8 Å × 41.3 Å, and left a 5 Å thick layer between the sides of the water box and the protein edges. The protein was neutralized using the auto-ionize feature in VMD. The simulation contained a total of 1893 water molecules, 882 protein atoms, and 6 Cl^- ions. The MD simulation in the water box was initialized by minimizing the water molecules for 500 fs while keeping the BPTI proteins and ions fixed. Next, the water molecules were heated from 0 to 300 K in 5 ps, and equilibrated for 10 ps at 300 K. The BPTI protein atoms were then released, and the whole system (protein, ions, and water molecules) was minimized for 1000 fs. The system was then heated from 0 to 300 K in 10 ps, and equilibrated for 40 ps at 300 K. The production stage of the 5 ns simulation was started after this equilibration stage. Coordinates of C^α were saved every 50 fs for analysis. The conjugate gradient method was used for minimization. The configurations were fitted and aligned to the first (reference) frame. This was done by removing all translational motion of the center of mass, and then using a least square fit procedure [21]. Each time series consisted of 100,000 points with a time step of 50 fs. The set of time series consisted of inter-residue distances for a selected group of residues surrounding the BPTI active site. These distances were extracted from simulations performed on the protein in vacuum and water.

BPTI is a very well researched and its properties are well documented. It has been called the "hydrogen molecule" of protein dynamics. It is a small and globular protein that lends itself easily to long MD simulations. It is very compact and its shape resembles that of a pear. It binds to and inhibits proteolytic enzymes like trypsin. The reactive site bond in BPTI lies in the narrow edge of the 'pear' between the residues Lys15 and Ala16. The section of the protein that binds to the enzyme during the inhibition is defined by the two anti-parallel strands consisting of residues 12–19, and 34–40 [22–24].

It is the aim of this article to show that the system can be modeled by a difference form of the Langevin equation (1):

$$X(t+\tau) = X(t) + g(X(t);\tau) + h(X(t);\tau)\Psi(t), \tag{2}$$

$$g(x) = \lim_{\tau\to 0}\frac{1}{\tau} < X(t+\tau) - x > \big|_{X(t)=x}, \tag{3}$$

$$h(x)\times h^T(x) = \lim_{\tau\to 0}\frac{1}{\tau} < (X(t+\tau)-x)\times(X(t+\tau)-x)^T > |_{X(t)=x}, \tag{4}$$

where τ is the lag time, $g(x)$ the drift coefficient, and $h(x)\times h^T(x)$ the diffusion coefficient [25]. If small lag times are possible then the limit $\tau\to 0$ can be taken in Eq. (2) to get the differential equation (1), provided that the dependence on τ takes the form $g(X(t);\tau)\approx\tau\, g(X(t))$ [26,27]. The functions g and h are assumed to have no explicit time dependence. However, they could depend on the lag time τ. In this work, this model will be valid for a certain range of values for τ, where the functions $g(X(t);\tau)$ and $h(X(t);\tau)$ are not significantly dependent on the value of τ. Another essential feature is that the system is Markovian, i.e. with no memory effects.

The technique for finding $g(x)$ and $h(x)$ in Eq. (2) using the definitions in Eq. (3) and (4) is the following: the inter-residue distance values in a time series are divided into bins, with the middle value of the bin, x, being the representative value. Given any value $X(t)$ within the bounds of a bin, the future value $X(t+\tau)$ is stored. All future values for a bin form a distribution. The values for $g(x)$ and $h(x)$ are then calculated using Eqs. (3) and (4), respectively.

The Markovian property for each series is checked by comparing the one-step and the two-step conditional probability densities. Agreement of the two would point to a lack of memory effects [7].

3. Analysis and discussion

A typical time series is shown in Fig. 1. This is the inter-residue distance between residues 13 and 39. The series is 5 ns long.

The Markovian property is shown by calculating the one- and two-step conditional probability density functions. As can be seen in Fig. 2, the two closely resemble each other. Simple regression analysis provides a slope very close to 1. This shows that there is no memory dependence in the values of the time series.

The deterministic and random contributions to this data time series, using a lag time value $\tau = 100$ time steps (5 ps) are shown in Fig. 3. The function $g(x)$ in Fig. 3a shows the deterministic contribution. There are three fixed points, where the value of $g(x)$ is zero. The two outside fixed points are stable. The fixed point inside in unstable. Any attempt by the protein to deviate to the left of the first stable fixed point, or to the right of the second stable fixed point, results in a large deterministic contribution that tends to return the protein to

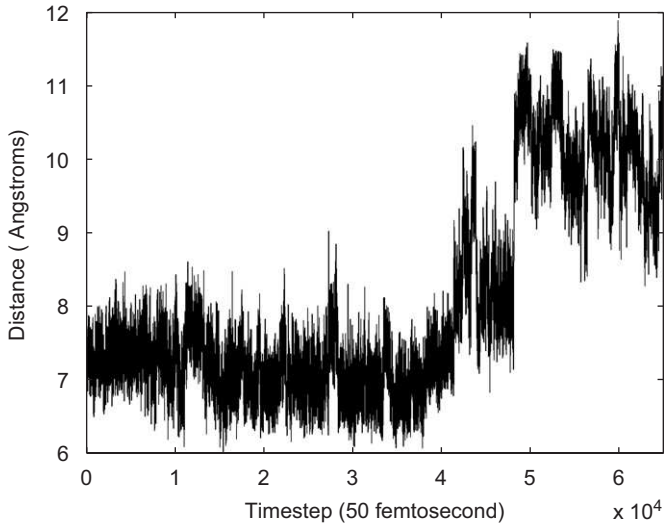

Fig. 1. Time series showing the inter-residue distance between residues 13 and 39.

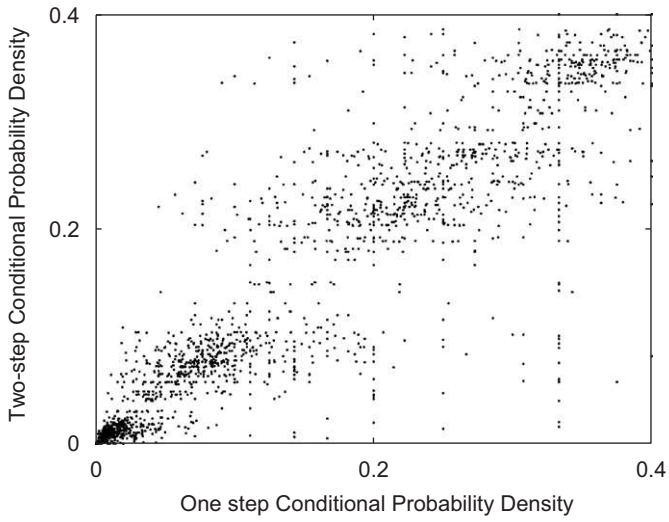

Fig. 2. The relationship between the one- and two-step conditional probability density functions.

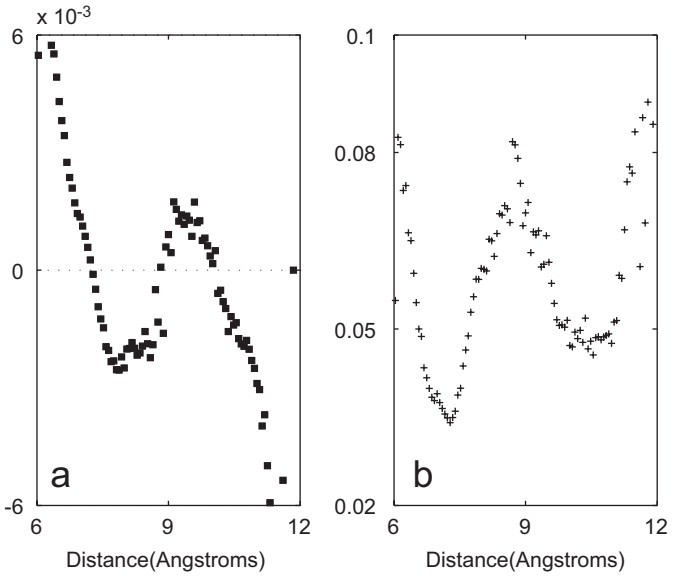

Fig. 3. (a) The deterministic ($g(x)$) and (b) the random ($h(x)$) contribution (in arbitrary units) for the 13–39 inter-residue distance time series. The protein is solvated in water.

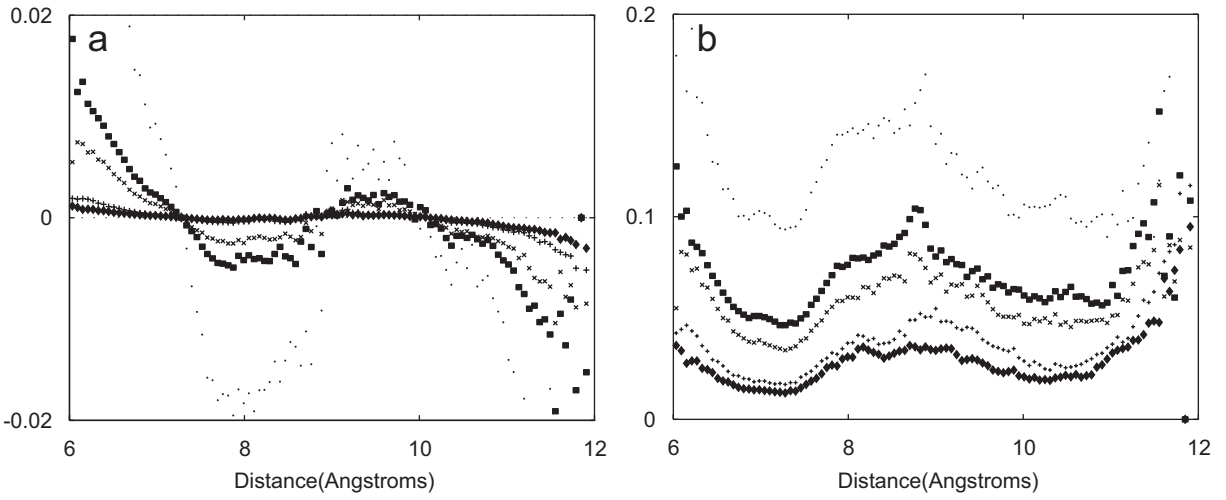

Fig. 4. (a) The time dependence for the deterministic contribution (in arbitrary units) for the time series of the inter-residue distance between residues 13 and 39 (● = 10 steps, ■ = 50 steps, x = 100 steps, + = 500 steps, ♦ = 1000 steps). (b) The time dependence for the stochastic contribution (Å) for the time series of the inter-residue distance between residues 13 and 39 (● = 10 steps, ■ = 50 steps, x = 100 steps, + = 500 steps, ♦ = 1000 steps).

the region defined by the two stable fixed points. The stochastic contributions (Fig. 3b) at these two stable points are small. In comparison, the stochastic contribution at the unstable point is large. Thus, while the deterministic contribution holds the protein at the two stable conformations, the stochastic contribution provides a large random jolt that sends the protein from one stable conformation to another stable conformation.

The dependence of this dynamic behavior on the lag time τ is shown in Figs. 4(a) and (b) for the deterministic contribution and the stochastic contribution, respectively. While it is true that the g and h reconstructions are obtained for the smallest step or as $\tau \to 0$, it is still very interesting to note that the behavior

of g and h are significantly independent off τ in the range between 10 time steps and 1000 time steps. It should be clearly stated that different regions of the dynamics in phase space are being included the larger the value of τ.

To check whether this dynamic behavior truly results from the dynamics of the system, or whether it is an artifact, the time series values were reshuffled randomly. This was done by breaking the time series into segments of equal length and mixing them randomly before applying the analysis discussed above. The original analysis was performed at a lag time of 100 steps. The time series was broken up into segments 10 time steps long (shorter than lag time) and 1000 steps (longer than lag time). The results for $g(x)$ and $h(x)$ are shown in Figs. 5(a) and (b), respectively. It is clear that while breaking up the series into pieces smaller than the time

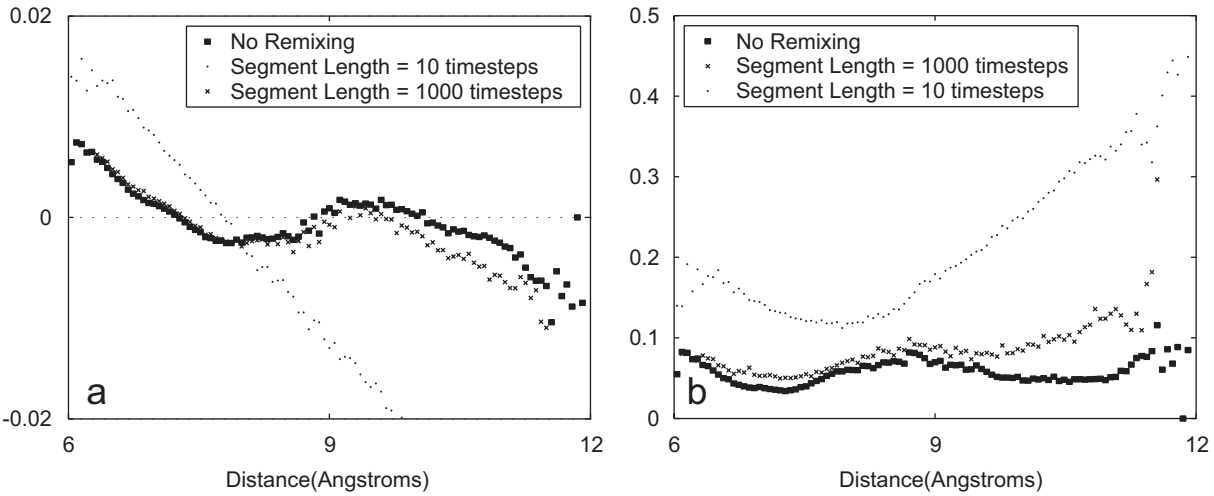

Fig. 5. (a) The effect of shuffling the time series on the value of the deterministic contribution. (b) The effect of shuffling the time series on the value of the stochastic contribution.

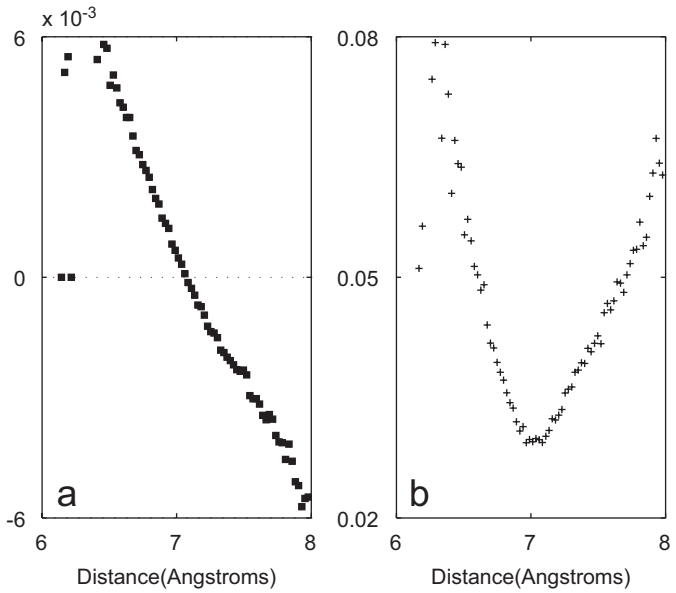

Fig. 6. (a) The deterministic and (b) stochastic contributions (in arbitrary units) for the 13–39 inter-residue distance time series extracted from a simulation with no solvent surrounding the protein.

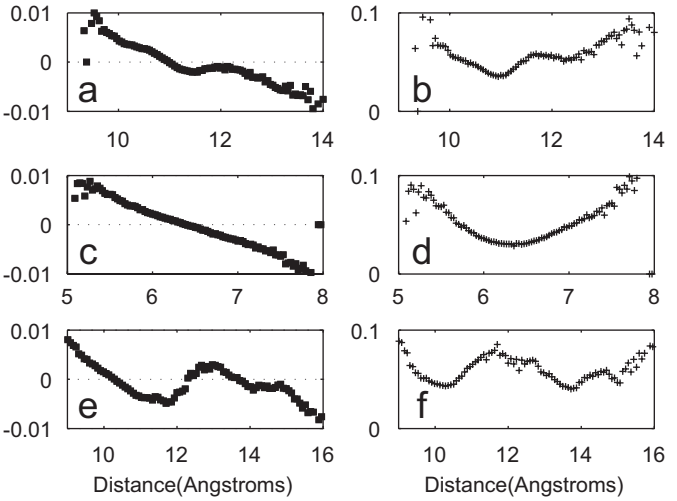

Fig. 7. (a)(c)(e) The deterministic contribution (in arbitrary units) for the 18–38, 16–35, 13–41 inter-residue distance time series, respectively. (b)(d)(f) The random contribution (in arbitrary units) for the 18–38, 16–35, 13–41 inter-residue distance time series, respectively. The protein is solvated in water.

lag removes the dynamic behavior completely, shuffling the series in pieces larger than the time lag keeps the dynamic behavior.

The same analysis was repeated for the time series of the inter-residue distance between residues 13 and 39 extracted from molecular dynamics simulations of the protein in vacuum, with no solvent water present. This was done to check whether the protein dynamics are indeed a 'slave' to the surrounding solvent. As can be seen in Fig. 6, the dynamic behavior exhibited for the solvated protein disappears. This clearly shows that the presence of the solvent plays a very important role in the dynamics of the protein, and the way it switches between conformations.

Other types of dynamical behavior were exhibited when the same analysis was performed on different residue pairs with the data extracted from the solvated protein simulation. A sampling representing different types of results are shown in Figs. 7(a)–(e).

The behavior in Fig. 7(a) and (e) shows non-linearity in the deterministic contribution with one stable fixed point and two stable fixed points, respectively. This is in contrast to the behavior shown in Fig. 7(c), which shows no nonlinearity. The stochastic contributions in Figs. 7(b) and (f) show a rich collection of local maxima as opposed to Fig. 7(d). This sampling of results points to the fact that different residue pairs perform correlated motions, jumping between different conformations.

4. Conclusion

The deterministic and stochastic contributions to time series of inter-residue distance in the active region of the BPTI protein show a rich variety of dynamic behavior. This behavior is dependent on the presence of solvent around the protein, offering further evidence that proteins are a 'slave' to the surrounding environment. The results can pinpoint residues that collaborate to perform jumps between different conformations, and others that show no such correlation. This analysis technique can further clarify how proteins perform their different functions by adapting their structure accordingly. The analysis of time series extracted from longer simulations and performed at different temperatures, should give a clearer picture of the dynamics involved in concerted protein motions using this analysis technique.

References

[1] P.W. Fenimore, H. Frauenfelder, B.H. McMahon, R.D. Young, Physica A 351 (2005) 1.
[2] A. Paciaroni, S. Cinelli, G. Onori, Biophys. J. 83 (2002) 1157.

[3] J.C. Smith, Q. Rev. Biophys. 24 (1991) 1.
[4] J. Fitter, Biophys. J. 76 (1999) 1034.
[5] H. Frauenfelder, G.A. Petsko, D. Tsernoglou, Nature 280 (1979) 558.
[6] S. Siegert, R. Friedrich, J. Peinke, Phys. Lett. A 234 (1998) 275.
[7] R. Friedrich, S. Siegert, J. Peinke, St. Lueck, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, G. Pfister, Phys. Lett A 271 (2000) 217.
[8] J. Gradisek, I. Grabec, S. Siegert, R. Friedrich, Mech. Syst. Signal Process. 16 (2002) 831.
[9] T.D. Frank, P.J. Beek, R. Friedrich, Phys. Lett. A 328 (2004) 219.
[10] P. Sura, S.T. Gille, J. Mar. Res. 61 (2003) 313.
[11] S. Kriso, J. Peinke, R. Friedrich, P. Wagner, Phys. Lett. A 299 (2002) 287.
[12] T. Kuusela, T. Shepherd, J. Hietarinta, Phys. Rev. E 67 (2003) 061904.
[13] W.I. Karain, N.I. Qaraeen, B. Ajarmah, Phys. Lett. A 312 (2006) 497.
[14] I. Nobuyuki, W. Shinzo, Stochastic Differential Equations and Diffusion Processes, North-Holland, Amsterdam, 1981.
[15] M. Kijima, Stochastic Processes with Applications to Finance, Chapman & Hall, 2002.
[16] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, J. Comput. Phys. 151 (1999) 283.
[17] W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graph. 14 (1996) 33.
[18] A. Wlodawer, J. Nachman, G.L. Gilliland, W. Gallagher, C. Woodward, J. Mol. Biol. 198 (1987) 469.
[19] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, J. Comput. Chem. 4 (1983) 187.
[20] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, J. Phys. Chem. B 102 (1998) 3586.
[21] W. Kabsch, Acta. Cryst. A 32 (1976) 922.
[22] M. Karplus, J.A. McCammon, Nat. Struct. Biol. 9 (2002) 646.
[23] P. Ascenzi, A. Bocedi, M. Bolognesi, A. Spallarossa, M. Coletta, R. De Cristofaro, E. Menegatti, Curr. Protein Pept. Sci. 4 (2003) 231.
[24] M.H. Yu, J.S. Weissman, P.S. Kim, J. Mol. Biol. 249 (1995) 388.
[25] M. Siefert, A. Kittel, R. Friedrich, J. Peinke, Europhys. Lett. 61 (2003) 466.
[26] N.G. Van Kampen, Stochastic Processes in Physics and Chemistry, North-Holland, New York, 1981.
[27] H. Risken, The Fokker–Planck Equation, Springer, Berlin, 1984.