Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Biophysical Chemistry 272 (2021) 106574

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: www.elsevier.com/locate/biophyschem

A theoretical study on the effects of interdomain flexibility on drug encounter rate for coronavirus nucleocapsid-type proteins

Tatsuhito Matsuo [a,b,c]

[a] Institute for Quantum Life Science, National Institutes for Quantum and Radiological Science and Technology, 2-4 Shirakata, Tokai-mura, Naka-gun, Ibaraki 319-1106, Japan
[b] Laboratoire Interdisciplinaire de Physique (LiPhy), Grenoble-Alpes University, 140 rue de la physique, 38402 Saint Martin d'Hères, France
[c] Institut Laue-Langevin, 71 avenue des Martyrs, CS 20156, 38042 Grenoble Cedex 9, France

ARTICLE INFO

Keywords:
Random walk simulation
Coronavirus nucleocapsid protein
Interdomain flexibility
Drug binding

ABSTRACT

To study the effects of the interdomain flexibility on the encounter rate of nucleocapsid-type protein with drug molecules, where two domains (NTD) are connected by a flexible linker and each NTD has a drug binding site, two-dimensional random walk simulation was carried out as a function of the interdomain flexibility and the drug concentration. NTDs represented as circles undergo random motions constrained by the interdomain flexibility while drug molecules are represented by lattice points. It was found that as the interdomain flexibility increases, the time interval between the drug bindings to the 1st and 2nd NTDs decreases, suggesting that the 2nd drug binding is accelerated. Furthermore, this effect was more significant at lower drug concentrations. These results suggest that the interdomain linker plays a key role in the drug binding process and thus emphasize the importance of characterization of their physicochemical properties to better evaluate the efficacy of potential drugs.

1. Introduction

As of today, it is known that there exist 7 human coronaviruses including the newest SARS-CoV-2, which is the pathogen of coronavirus disease 2019 (COVID-19) pneumonia causing the pandemic [1]. The virion of coronavirus mainly consists of four structural proteins called spike protein, membrane protein, envelope protein, and nucleocapsid protein (N^{pro}). Among them, the nucleocapsid protein is the most abundant viral protein during infection cycle. The major role of N^{pro} is to package the viral RNA genome to form a helical ribonucleoprotein complex, which is essential for replication and transcription of the viral genome. Furthermore, it has been shown that N^{pro} modulates the host innate immune response [2]. Due to its high immunogenicity, N^{pro} has been used in vaccine development and also emerged as a potential drug target [3]. Coronavirus N^{pro} consists of two domains called as N-terminal domain (NTD) and C-terminal domain (CTD), which are connected by a linker region (Fig. 1 (a)). Although both domains (NTD and CTD) have an RNA-binding site, it is known that while NTD is responsible for RNA binding, CTD is responsible for oligomerization [4].

It is well accepted that molecular flexibility in protein molecules such as the one in the active sites and their surrounding regions plays a major role in ligand (or drug) binding by flexibly changing their structure to fit the conformation of the binding partners, which is the so-called "induced-fit" mechanism [5]. In the case of N^{pro}, it is thus important to investigate the binding between a drug molecule and the RNA-binding site on each NTD. Recent molecular dynamics simulation studies on the affinity of various potential drug molecules and the RNA-binding site on the NTD have detected antiviral agents that bind the NTD with high affinity [6–9], suggesting that the new drug development targeting N^{pro} will be accelerated using promising in silico simulation.

However, in addition to the intradomain flexibility described above, there is another important factor that determines the drug binding kinetics, which is an interdomain flexibility. According to small-angle X-ray scattering, N^{pro} forms a homodimer, in which the CTDs bind to each other while NTDs are connected to the CTDs through flexible linkers, which are assumed to be in a disordered state [10]. NTDs are thus considered to move freely in solution constrained by the flexible interdomain linker. On the other hand, in the last two decades, it has been shown that intrinsically disordered regions are not completely in an unstructured state but in equilibrium with those in a partially folded state having α-helices [11] and their relative population varies between 10% and 70% depending on proteins [12]. Moreover, some physiological substances such as osmolytes are known to induce helical folding in the intrinsically disordered proteins [13]. Since the chemical

E-mail address: matsuo.tatsuhito@qst.go.jp.

https://doi.org/10.1016/j.bpc.2021.106574
Received 8 January 2021; Received in revised form 19 February 2021; Accepted 2 March 2021
Available online 9 March 2021

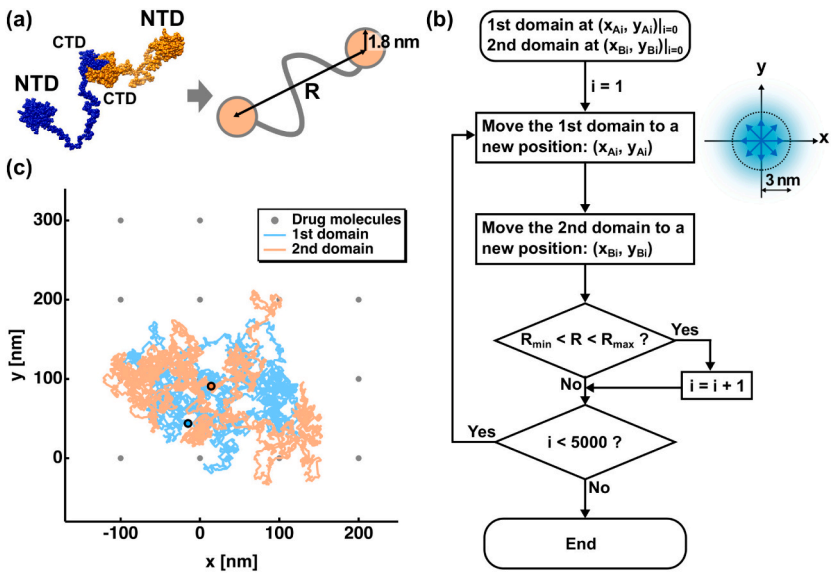

Fig. 1. (a) Approximation of the N^{pro} structure, where monomers are shown in blue and orange (*left*), by two circles with the radii corresponding to the size of each N-terminal domain (NTD). The interdomain distance is denoted as R (*right*), the range of which represents the interdomain flexibility. The structure of N^{pro} was depicted using UCSF Chimera [27]. (b) A flow diagram of random walk simulation employed in this study to generate trajectories of proteins. In the chart, R denotes the interdomain distance as shown in (a), R_{min} and R_{max} denote the minimum and maximum interdomain distances defined by a given interdomain flexibility, respectively (see also Table 1). On the right side, directions that each domain is allowed to move during the simulation are shown in blue arrows. The distance per step was randomly determined based on the gaussian distribution with the standard deviation of 3 nm (see Materials and Methods), and thus has a distribution as shown in marine blue. (c) An example of the trajectory of a two-dimensional random-walk simulation at F = 0.75 and C = 1.7 (For F and C, see Tables 1 and 2, respectively). The initial positions of the 1st domain and of the 2nd domain are shown by cyan and orange circles with their trajectories shown in solid lines in corresponding colors. The drug molecules are represented in grey dots. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

composition inside the cell varies by external stress and during the cell cycle [14–19], the interdomain linker of N^{pro} likely takes various forms with different flexibility, which could change the mobility of each NTD. It is thus essential to gain insights into whether such differences in intramolecular flexibility have any effect on the encounter rate of each NTD with drug molecules or not in order to give further direction in the future research aiming for new drug developments. However, there has been no study addressing this aspect theoretically nor experimentally. Furthermore, it is not straightforward to simulate a three-dimensional system where both the protein and many drug molecules diffuse, and follow the system over a long time until the proteins bind drugs.

Based on the above reasons, this study has employed a two-dimensional random walk simulation on N^{pro}-type proteins in order to gain fundamental information on possible effects of the interdomain flexibility on the drug encounter rate of the NTDs when the interdomain flexibility and the drug concentration change. Random walk simulation is a useful tool to investigate the events governed by probability [20]. It has been used to study the diffusion process of entire protein molecules and explain their diffusive behavior observed experimentally [21–23]. However, to the best of the author's knowledge, this technique has not been applied to a protein where two domains are interconnected by a flexible linker such as N^{pro}. Application of this technique to such proteins would be useful to obtain the information on the effects of the interdomain flexibility. The purpose of this study is to see the relative change in time required for drug binding to each NTD caused by increasing interdomain flexibility at different drug concentrations with respect to the case where the interdomain flexibility is low, meaning that the results obtained at F = 0.02 (see Materials and Methods below) serve as a control. This corresponds to comparison of the state where the interdomain linker have different degrees of disordered regions with the extreme state where most of the linker takes on folded structures (F = 0.02). Therefore, in this study, parameters extracted from the trajectories were described after normalization by those obtained at F = 0.02. From the analysis of the simulation results, it was found that as the interdomain flexibility increases, the drug binding to NTDs is generally promoted. Moreover, the time interval between the drug binding to the 1st and the 2nd NTDs decreases with increasing the interdomain flexibility, suggesting that the 2nd drug binding is accelerated. The magnitude of these effects was found to be larger at lower drug concentrations.

2. Materials and methods

2.1. Random walk simulation

Random walk simulation was carried out in two dimensions. Since the stochastic processes occur independently in each direction (x, y, and z axes in the Cartesian coordinate system), reducing one dimension should not affect the features that are focused on in this study.

First, to estimate the overall size of the N^{pro} molecules, a template structures was generated based on the atomic structure of the NTD (PDB ID: 6m3m [24]) and CTD (PDB ID: 7c22 [25]) of N^{pro} of SARS-CoV-2. The interdomain linker was modelled using the Ranch program [26]. One of the obtained models that was randomly selected are shown in Fig. 1 (a), where two NTD domains are separated by about 20 nm. Based on this model, the two NTDs were represented as two identical circles with their radii of 1.8 nm. To carry out the random walk simulation in realistic space- and time-scales, the spatial and time steps in the random walk simulation were determined based on the translational diffusion coefficients (D_T) of the atomistic model of N^{pro} above. The D_T value of this model was calculated using the software HydroPro [28], which was 4.7×10^{-7} cm^2/s. This roughly corresponds to the translational movement of 2 nm during 21 ns according to the relation $D_T = a^2/(4\tau)$, where a is the spatial step and τ is the time step. Furthermore, to take account of the fluctuations of the distance that a particle proceeds at each step, the spatial step was randomly selected based on the gaussian distribution $f(a) = \exp.(-(a/(\sigma\sqrt{2}))^2)/(\sigma\sqrt{(2\pi)})$, where σ is the standard deviation and set to be 3 nm, which leads to the average distance of 2 nm (the a value at 50% of the integrated area of f(a)). The time step of random walk was set to be 21 ns.

During the simulation, the interdomain distance R between the centers of the 2 circles (NTDs) was allowed to change within specified ranges, which reflect the interdomain flexibility. The ranges of the R values used in the simulation are shown in Table 1 with corresponding values of the interdomain flexibility F, which was defined as $F = (R_{max} - R_{min})/R_{dmax}$, where R_{max} and R_{min} denote the maximum and minimum R values at a given condition shown in Table 1, respectively, and R_{dmax} denotes the (R_{max} – R_{min}) value of the highest F value, i.e. 106 (= 109.6–3.6). The movement of each domain is thus restrained according to a given F. The initial positions of the two NTDs were randomly determined so that their interdomain distance falls within the range of the given condition. The procedure to generate trajectories is summarized in a flowchart shown in Fig. 1 (b). At each step of the simulation,

Table 1
The range of the interdomain distance and corresponding values of the interdomain flexibility.

Range of the interdomain distance (R) $R_{min} < R$ [nm] $< R_{max}$	Interdomain Flexibility (F)
$19.0 < R < 21.0$	0.02
$16.0 < R < 24.0$	0.07
$12.5 < R < 27.5$	0.14
$6.0 < R < 34.0$	0.26
$3.6 < R < 57.6$	0.51
$3.6 < R < 83.6$	0.75
$3.6 < R < 109.6$	1.00

the domains were randomly moved to one of the 8 neighboring points such that they satisfy the distance restraints in Table 1. The directions of the movement in each step were those along the x- or y-axis and in diagonal directions in both plus and minus directions.

For each condition, 6000 runs were carried out, where each run contains 5000 steps of random walk (corresponding to 105 μs), to observe the statistical nature of the system. It was confirmed that even 2000 runs give essentially the same results as those obtained by 6000 runs, which show smaller statistical errors and are presented in this paper, indicating that 6000 runs are enough to extract the statistical features of the current system.

Drug molecules are represented as the two-dimensional lattice points which are evenly spaced. The distance between neighboring drug molecules was determined according to the drug concentrations employed in this study. Since the in vivo drug concentration is generally in the range of 1–5000 nM [29], the corresponding grid intervals are employed in the simulation as shown in Table 2.

2.2. Analysis of the trajectory

Analysis of the trajectories obtained was carried out by superimposing the trajectories described above with the two-dimensional square lattice, where each lattice point represents the position of drug molecules. The trajectories were placed on the lattice such that at the time $t = 0$ [ns], one NTD is randomly positioned in the square area with the side of the lattice spacing defined in Table 2, and the center of the square is at the origin. In the case where the distance between the center of an NTD and a lattice point is within 1.8 nm, the NTD (the 1st domain) was judged to bind the drug molecule at $t = T_1$ [ns]. In the same manner, another NTD (the 2nd domain) binds a drug molecule at $t = T_2$ [ns]. Note that the encounter with a drug molecule is considered to be equal to binding because the affinity between proteins and drug molecules is not taken into account here, which would not affect the conclusions obtained in this study.

During the simulation, there is a possibility that both NTDs bind to the identical drug molecule after some time interval, i.e., the 2nd domain reaches the same lattice point as the one that the 1st domain has already passed some time earlier. It is reasonable to assume that after binding of a drug molecule with the 1st domain, it becomes unavailable to the 2nd domain. On the other hand, from the thermodynamical viewpoint, to keep the drug concentration constant, that lattice point must be occupied again by another drug molecule after some time.

Table 2
The values of the grid interval and corresponding drug concentration.

Grid interval [nm]	Drug concentration (C) [nM]
7	4843
10	1661
15	492
25	106
32	50.7
50	13.3
100	1.7

Because this can be done through diffusion of a drug molecule located on the neighboring lattice points, it is important to estimate the time for this process. The molecular weight of several potential drug molecules for coronavirus N^{pro} is reported to be about 300–400 [6,30], which corresponds to the translational diffusion coefficient of $D_T \sim = 1.3 \times 10^{-5}$ cm^2/s [31]. The average time (t_d) that it takes for a drug molecule to move to the neighboring lattice point is given by $t_d = L^2/(4D_T)$, where L is the distance between the neighboring lattice points at a given drug concentration. Therefore, the binding events between the NTDs and the identical drug that occur within the time t_d were omitted from the analysis of the trajectories.

3. Results and discussion

3.1. Features of the trajectories of the random-walk simulations

In order to see the properties of the trajectories obtained, mean square displacements (MSDs) of the two NTDs were calculated for all the 6000 runs and they were averaged. As shown in Fig. 2 (a), the MSDs are linear functions of time, meaning that the obtained trajectories generated in this study show a typical feature expected for random-walk or Brownian motions. In addition, the interdomain distance between the two NTDs was calculated for each step in each run of the randomly selected 20 runs at each F value and its distribution is shown in Fig. 2 (b). It is clear that as the interdomain flexibility increases, both domains are more separated each other during the simulation time, confirming that the constraint between the interdomain distance is properly incorporated into the current simulation. Then, the effects of the interdomain flexibility on the binding rate with drug molecules were analyzed with changing the drug concentrations.

3.2. Assumption on the positions of drug molecules

In this study, three parameters were analyzed based on the

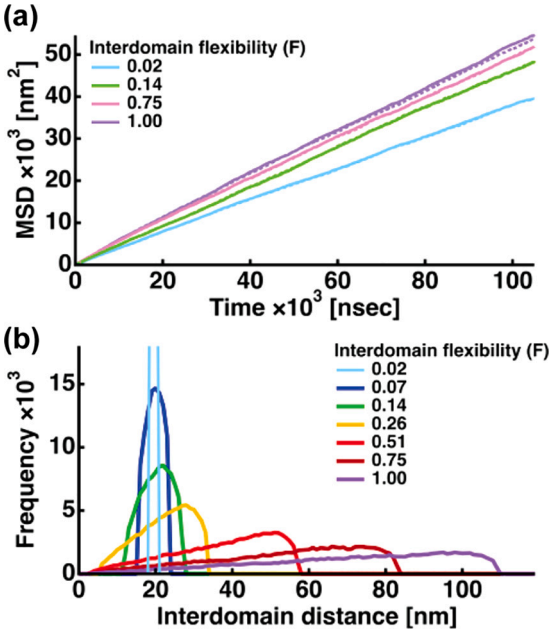

Fig. 2. (a) Mean square displacements (MSD) of the 1st and the 2nd domains during simulations as a function of time. The MSDs shown are those averaged over 6000 runs. Those at different interdomain flexibility (F) are shown. The solid lines denote the MSDs of the 1st domain and the dotted lines denote the MSDs of the 2nd domain. Except for F = 1.00, the solid and dotted lines are almost overlapping. (b) Distributions of the interdomain distance during the simulation (20 runs randomly selected, i.e. 10^5 steps) at all the F values employed in this study.

trajectories obtained. Those parameters are the time required for binding of the 1st domain to a drug molecule (T_1), the time required for binding of the 2nd domain to a drug molecule (T_2), and the time interval between the 1st and the 2nd drug bindings ($\Delta T = T_2 - T_1$) (For the definition of T_1 and T_2, also see the Materials and Methods). Regarding the drug molecules, it is assumed that they are positioned at square lattice points with their intervals changing according to a given drug concentration. Compared with the real system where both the proteins and drug molecules undergo Brownian motion, this assumption would definitely affect the absolute values of T_1, ΔT, and T_2. However, when focused on the ratios of these parameters obtained at, for instance, two different values of interdomain flexibility at the same drug concentration, the relative changes in these parameters would not be affected by the current assumption: In the real system, all drug molecules undergo random diffusive motions and form ensembles of various trajectories. Statistically, these ensembles of trajectories of the drug molecules are common among different systems (e.g. between two systems which contain the drug molecules and NTDs at either of two different F values). For each trajectory of drug molecules, the only differences between the two systems are the positions and the movements of the 1st NTD and the 2nd NTD. In this case, even if the drug molecules are fixed on their average coordinates estimated from the corresponding drug concentration, the effects of the difference in interdomain flexibility would be reflected in the ratio of the parameters extracted from the two systems. After all, the changes in parameters obtained at different F values relative to those at F = 0.02 would not be affected. Therefore, in the following, the T_1, ΔT, and T_2 values are shown after normalization by those at F = 0.02, and their absolute values are in Fig. S1 of the supplementary material only for reference.

3.3. Drug binding to the 1st domain

First, T_1 values were calculated for all the runs and then its statistical feature, i.e. the average and the standard error of the mean of the distribution of the T_1 values were calculated. Then, to see the effects of the interdomain flexibility, the T_1 values at each drug concentration were normalized by that at F = 0.02 and their values are shown in Fig. 3 (a) as a function of F. It is found that the T_1 values decrease as F increases. The decreases in T_1 except for C = 4843 [nM] are statistically significant according to the Steel-Dwass test for nonparametric multiple comparisons [32,33] (in particular, $p < 0.001$ for F ≥ 0.51), suggesting that the 1st drug binding is promoted. This would be due to that the less restrained NTD can diffuse more freely, facilitating finding a drug molecule. This promotion effect is enhanced as the drug concentration decreases. At the highest concentration of C = 4843 [nM], the decrease in T_1 compared with that at F = 0.02 is not statistically significant, suggesting that the interdomain flexibility does not play a role on the drug binding. Under this condition, there are many drug molecules around the NTD, and thus it can bind to the nearest drug molecule within a relatively short time regardless of the degree of the interdomain flexibility. This result also implies that at high drug concentrations, it may be possible to estimate the drug binding kinetics of the proteins properly without considering the effect of flexible linkers. At the lowest concentration of C = 1.7 [nM], the T_1 values do not decrease as much as at the higher concentrations though the decrease in T_1 compared with that at F = 0.02 is statistically significant ($p < 0.001$ for F ≥ 0.26), suggesting that there exists an environmental condition that maximizes the binding-promotion effects of the interdomain flexibility depending on the size of the protein and drug concentrations.

3.4. Drug binding to the 2nd domain

Next, ΔT ($= T_2 - T_1$) values were calculated at each value of F and C, and their distributions after normalization by the value at F = 0.02 are shown in Fig. 3 (b). It is found that at C = 4843 [nM], the ΔT values do not change significantly as F increases ($p > 0.05$), indicating that the

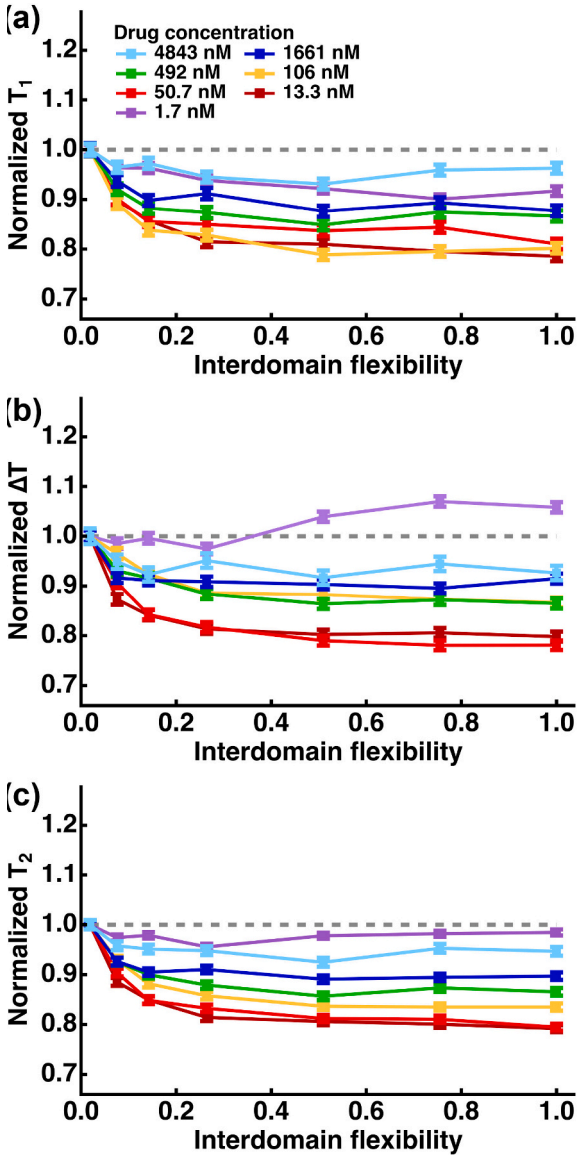

Fig. 3. Summary of the extracted parameters from the trajectories of the random-walk simulation. The time required for the 1st drug binding (T_1), the time interval between the 1st and the 2nd drug binding ($\Delta T = T_1 - T_2$), and the time required for the 2nd drug binding (T_2), are shown as a function of interdomain flexibility in (a), (b), and (c), respectively. These profiles are normalized by the corresponding values at F = 0.02. Error bars show the standard error of the mean. The profiles of T_1, ΔT, and T_2 before normalization are shown in Fig. S1.

interdomain flexibility does not play a role on drug binding as observed in T_1. On the other hand, the ΔT values decrease at other drug concentrations (p-values were in the range of 0.001–0.004 for F ≥ 0.26) except for C = 1.7 [nM], suggesting that the flexible interdomain linker promotes the second drug binding. As a result, as shown in Fig. 3 (c), the time required for the second drug binding (T_2) is also found to decrease except for C = 4843 and 1.7 [nM] as the interdomain flexibility increases ($p < 0.001$ for F ≥ 0.07). These results also imply that the total number of drugs that the NTDs encounter during a given time increases as well.

At C = 1.7 [nM], the ΔT values are found to increase as F increases ($p < 0.001$ for F ≥ 0.51) while the T_2 values do not decrease as the interdomain flexibility increases unlike at other drug concentrations. To ascertain the origin of this unique behavior, the distribution of the positions of the 2nd domain at a time step when the 1st domain has bound

to a drug molecule was calculated for all the runs as shown in Fig. 4. In this figure, the coordinates of the 1st domain are translated to the origin. It is found that when the interdomain flexibility is the lowest (F = 0.02), the 2nd domains are circularly distributed near the origin because of the restraints imposed on the R values. On the other hand, at F = 1.00, the 2nd domains are distributed in much larger areas than at F = 0.02. Thus, on average, while the 2nd domains at F = 0.02 tend to bind to a drug molecule at the origin, those at F = 1.00 need to move a longer distance to bind to a drug. This explains why the T_2 and hence ΔT do not decrease as the interdomain flexibility increases at the lowest drug concentration. Since the binding events which take place against the identical drug molecule within the time interval of t_d (see Materials and Methods) are omitted, it appears that the results obtained at the lowest drug concentration might also occur in the real situation.

3.5. Separation of contributions of interdomain flexibility and interdomain length

In the current simulation, higher F values result in both a broader distribution of interdomain distances in addition to a significant increase in maximum interdomain distance. Although in the real system, partial folding/unfolding in secondary structures in the linker, which could occur in different chemical environments, brings about these both effects, it is intriguing to separate the role of flexibility and length. For this purpose, additional simulations were carried out as follows: For each condition on interdomain flexibility in Table 1, the range of the interdomain distance was significantly narrowed down by setting $R_{min} = 0.95R_{max}$. Under these conditions representing very "rigid" molecules, the same procedures as above were conducted. Because the molecules in these conditions are stretched and hence almost preserve their maximum interdomain distance during the simulation, comparison of the results obtained for these conditions and those obtained for the conditions in Table 1 (i.e. those described above) would provide information on the effect of the interdomain flexibility on the T_1, ΔT and T_2 parameters with minimizing the effect of the differences in maximum interdomain distance. The T_1, ΔT and T_2 values obtained for the "rigid" conditions are denoted as T_{1L}, ΔT_L and T_{2L}, respectively, while those obtained for the original calculation described above are denoted here as T_{1H}, ΔT_H and T_{2H}, respectively. Fig. 5 shows the ratios of T_{1H}/T_{1L}, $\Delta T_H/\Delta T_L$ and T_{2H}/T_{2L} at each drug concentration. It is seen that all these parameter values are lower than 1.0 ($p < 0.001$ according to the one-sample Wilcoxon signed rank test [34]) at all the drug concentrations except for two T_{2H}/T_{2L} values at (F, C) = (0.02, 1.7) and (1.00, 1.7), suggesting that it is the interdomain flexibility rather than the differences in maximum interdomain length that facilitates the drug binding of each domain. At the extreme drug concentrations at C = 4843 and 1.7 [nM], the parameter values are found to be closer to 1.0 than at other drug concentrations, suggesting that the impact of the interdomain flexibility depends on drug concentrations. Moreover, combined with the results shown in Fig. 3 that the promotion effects of drug binding are reduced at these two drug concentrations compared with at other concentrations, these results support the notion of the interdomain flexibility promoting drug binding.

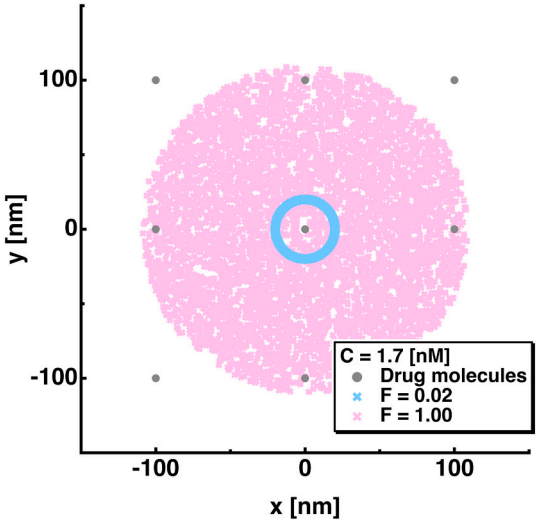

Fig. 4. The positions of the 2nd domain at the time when the 1st domain has bound to a drug molecule, the coordinate of which is translated to the origin (x, y) = (0, 0). The positions of the 2nd domain were extracted from all the runs at two conditions of F = 0.02 and 1.00, which are shown in cyan and magenta markers, respectively. Drug molecules are shown in grey and their concentration is 1.7 [nM] (C = 1.7). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

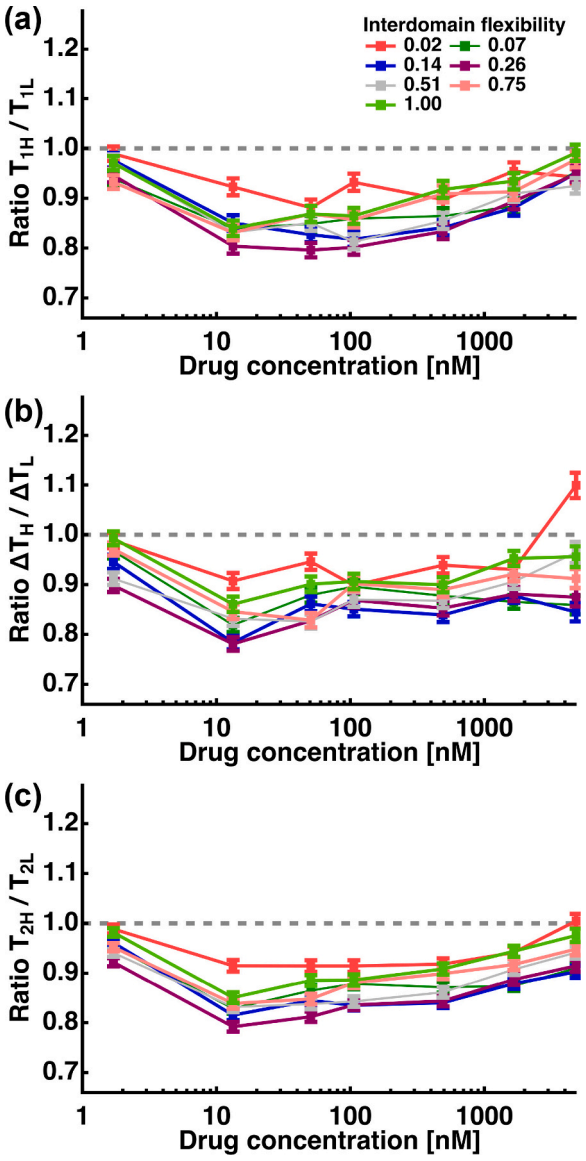

Fig. 5. Separation of the contributions of interdomain flexibility and interdomain distance to the parameters obtained for simulations at different interdomain flexibility. T_{1H}, ΔT_H, and T_{2H} denote those obtained for the simulation shown in Fig. 3 while T_{1L}, ΔT_L, and T_{2L} denote those obtained for the simulation for "rigid" molecules the interdomain distance of which is allowed to move only between $0.95R_{max} < R < R_{max}$ at each F value in Table 1. Error bars show the standard error of the mean.

3.6. Implications of the current simulation results

N^{pro} has emerged as a potential target for new drug developments and thus been particularly paid much attention. However, there has been no study on what effects the elusive interdomain flexibility might have on its behavior in terms of drug encounter, the understanding of which would be required for efficient in silico drug development against N^{pro} of various pathogens in the future. It is thus important to use the system where the protein size is similar to that of N^{pro} and drug concentration is within the realistic range in order to obtain fundamental information on the effects of interdomain flexibility. The present study suggests that both the interdomain flexibility and the drug concentration affect the encounter rate between the drug molecules and the N^{pro} molecules, which puts an emphasis on the importance of experimental characterization of physicochemical properties of their flexible linker. This would then affect the approaches and procedures employed in the future in silico studies aimed for searching the potential drugs against N^{pro} and estimating their efficacy at a molecular level. The current study thus provides a foundation for further research to improve the methodology in new drug developments and estimation of molecular interactions between N^{pro} and drugs/ligands.

It has been shown that more than 70% of eukaryotic proteome are proteins where domains are connected by linkers with variable lengths and sequences [35]. Some of them are proteins containing two domains and both domains have a ligand binding site such as an RNA-binding protein Maleless (MLE) [36] and a S-type lectin Galectin-4 involved in tumor progression [37]. Although the behavior of the parameters analyzed in this study depends on the structural features of the targeted proteins such as the domain size and the linker length given a realistic range of drug concentration, the findings obtained in the present study could also be useful to design further experimental and theoretical approaches to elucidate the molecular mechanism of biochemical reactions that such proteins are involved in because of the simplification of the system employed here.

4. Conclusion

Two-dimensional random-walk simulation was carried out in order to study the effects of the flexibility of the interdomain linker of a N^{pro}-type protein on the encounter rate of its domains with drug molecules. The simulation has shown that the increase in the interdomain flexibility facilitates the binding of the 1st NTD with a drug molecule. Furthermore, it has been found that the time interval between the 1st and the 2nd drug bindings decreases as the interdomain flexibility increases, suggesting that the 2nd drug binding is accelerated once the 1st drug binding has occurred. Moreover, the magnitude of the acceleration of the drug binding is more significant at lower drug concentrations, showing that the interdomain linker plays a pivotal role in these conditions. The simulation at the lowest concentration even implies that the interdomain flexibility might slow down the binding events. The present study thus suggests the importance of interdomain flexibility in terms of drug binding rate and thus would give a future research direction to improve the approaches for more efficient drug developments. Simultaneously, this study suggests that it is imperative to study the dynamical nature of interdomain flexible linkers in addition to the binding kinetics between a drug molecule and an active site in the domain. The powerful experimental techniques to study such a dynamical aspect include nuclear magnetic resonance [38] and incoherent neutron scattering [39]. Moreover, since the dynamical aspect of the molecule is closely related to its structure, the structural characterization of flexible linkers by, for instance, X-ray scattering [40] is also promising. Future studies combining these methods would provide detailed structural and dynamical information on the interdomain linker in a wide range of spatial and temporal scales, which would advance our understanding of the molecular mechanism of functions of flexible proteins not limited to N^{pro}, and hence contribute to the further improvement in the accuracy of molecular dynamics simulation, leading to more efficient new drug developments targeting these proteins.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

I am grateful to Prof. Judith Peters for her encouragement. This work was partially supported by QST President's Strategic Grant (Exploratory Research).

Appendix A. Supplementary materials

The gallery of the dependence of the unnormalized T_1, ΔT, and T_2 values on the interdomain flexibility at each drug concentration is shown in Fig. S1 in the supplementary material.

References

[1] P. Zhou, X. Lou Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C.L. Huang, H.D. Chen, J. Chen, Y. Luo, H. Guo, R. Di Jiang, M.Q. Liu, Y. Chen, X. R. Shen, X. Wang, X.S. Zheng, K. Zhao, Q.J. Chen, F. Deng, L.L. Liu, B. Yan, F. X. Zhan, Y.Y. Wang, G.F. Xiao, Z.L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature. 579 (2020) 270–273, https://doi.org/10.1038/s41586-020-2012-7.

[2] S.A. Kopecky-Bromberg, L. Martínez-Sobrido, M. Frieman, R.A. Baric, P. Palese, Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and Nucleocapsid proteins function as interferon antagonists, J. Virol. 81 (2007) 548–557, https://doi.org/10.1128/jvi.01782-06.

[3] N.K. Dutta, K. Mazumdar, J.T. Gordy, The Nucleocapsid protein of SARS–CoV-2: a target for vaccine development, J. Virol. 94 (2020) 1–2, https://doi.org/10.1128/jvi.00647-20.

[4] C. ke Chang, C.M.M. Chen, M. hui Chiang, Y. lan Hsu, T. huang Huang, Transient Oligomerization of the SARS-CoV N Protein - Implication for Virus Ribonucleoprotein Packaging, PLoS One 8 (2013), https://doi.org/10.1371/journal.pone.0065045.

[5] D.E. Koshland, Application of a theory of enzyme specificity to protein synthesis, Proc. Natl. Acad. Sci. 44 (1958) 98–104, https://doi.org/10.1073/pnas.44.2.98.

[6] R. putri Indahningrum, Investigation of N terminal domain of SARS CoV 2 nucleocapsid protein with antiviral compounds based on molecular modeling approach, Sci. Prepr. (2020), https://doi.org/10.14293/S2199-1006.1.SOR-.PPPT99I.v1.

[7] P. Sarma, N. Shekhar, M. Prajapat, P. Avti, H. Kaur, S. Kumar, S. Singh, H. Kumar, A. Prakash, D.P. Dhibar, B. Medhi, In-silico homology assisted identification of inhibitor of RNA binding against 2019-nCoV N-protein (N terminal domain), J. Biomol. Struct. Dyn. (2020) 1–9, https://doi.org/10.1080/07391102.2020.1753580.

[8] R. Rolta, R. Yadav, D. Salaria, S. Trivedi, M. Imran, A. Sourirajan, D.J. Baumler, K. Dev, In silico screening of hundred phytocompounds of ten medicinal plants as potential inhibitors of nucleocapsid phosphoprotein of COVID-19: an approach to prevent virus assembly, J. Biomol. Struct. Dyn. (2020) 1–18, https://doi.org/10.1080/07391102.2020.1804457.

[9] H. Kaur, N. Shekhar, S. Sharma, P. Sarma, A. Prakash, B. Medhi, Ivermectin as a potential drug for treatment of COVID-19: an in-sync review with clinical and computational attributes, Pharmacol. Rep. (2021), https://doi.org/10.1007/s43440-020-00195-y.

[10] W. Zeng, G. Liu, H. Ma, D. Zhao, Y. Yang, M. Liu, A. Mohammed, C. Zhao, Y. Yang, J. Xie, C. Ding, X. Ma, J. Weng, Y. Gao, H. He, T. Jin, Biochemical characterization of SARS-CoV-2 nucleocapsid protein, Biochem. Biophys. Res. Commun. (2020), https://doi.org/10.1016/j.bbrc.2020.04.136 (S0006-291X(20)30876–7).

[11] D.-H. Kim, K.-H. Han, Transient secondary structures as general target-binding motifs in intrinsically disordered proteins, Int. J. Mol. Sci. 19 (2018) 3614, https://doi.org/10.3390/ijms19113614.

[12] S.-H. Lee, D.-H. Kim, J.J. Han, E.-J. Cha, J.-E. Lim, Y.-J. Cho, C. Lee, K.-H. Han, Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins, Curr. Protein Pept. Sci. 13 (2012) 34–54, https://doi.org/10.2174/138920312799277974.

[13] Y.-C. Chang, T.G. Oas, Osmolyte-induced folding of an intrinsically disordered protein: folding mechanism in the absence of ligand, Biochemistry. 49 (2010) 5086–5096, https://doi.org/10.1021/bi100222h.

[14] C.J. Smoyer, S.L. Jaspersen, Breaking down the wall: the nuclear envelope during mitosis, Curr. Opin. Cell Biol. 26 (2014) 1–9, https://doi.org/10.1016/j.ceb.2013.08.002.
[15] M.P. Stewart, J. Helenius, Y. Toyoda, S.P. Ramanathan, D.J. Muller, A.A. Hyman, Hydrostatic pressure and the actomyosin cortex drive mitotic cell rounding, Nature. 469 (2011) 226–230, https://doi.org/10.1038/nature09642.
[16] R. Vancraenenbroeck, Y.S. Harel, W. Zheng, H. Hofmann, Polymer effects modulate binding affinities in disordered proteins, Proc. Natl. Acad. Sci. 116 (2019) 19506–19512, https://doi.org/10.1073/pnas.1904997116.
[17] C.M. Davis, M. Gruebele, S. Sukenik, How does solvation in the cell affect protein folding and binding? Curr. Opin. Struct. Biol. 48 (2018) 23–29, https://doi.org/10.1016/j.sbi.2017.09.003.
[18] S. Sukenik, M. Salam, Y. Wang, M. Gruebele, In-cell titration of small solutes controls protein stability and aggregation, J. Am. Chem. Soc. 140 (2018) 10497–10503, https://doi.org/10.1021/jacs.8b04809.
[19] D. Moses, F. Yu, G.M. Ginell, N.M. Shamoon, P.S. Koenig, A.S. Holehouse, S. Sukenik, Revealing the hidden sensitivity of intrinsically disordered proteins to their chemical environment, J. Phys. Chem. Lett. 11 (2020) 10131–10136, https://doi.org/10.1021/acs.jpclett.0c02822.
[20] F. Spitzer, Principles of Random Walk, 2nd edn, Springer Science & Business Media, 1976, https://doi.org/10.1007/978-1-4757-4229-9.
[21] J. Helenius, G. Brouhard, Y. Kalaidzidis, S. Diez, J. Howard, The depolymerizing kinesin MCAK uses lattice diffusion to rapidly target microtubule ends, Nature. 441 (2006) 115–119, https://doi.org/10.1038/nature04736.
[22] B.M. Regner, D. Vučinić, C. Domnisoru, T.M. Bartol, M.W. Hetzer, D. M. Tartakovsky, T.J. Sejnowski, Anomalous diffusion of single particles in cytoplasm, Biophys. J. 104 (2013) 1652–1660, https://doi.org/10.1016/j.bpj.2013.01.049.
[23] J. Han, J. Herzfeld, Macromolecular diffusion in crowded solutions, Biophys. J. 65 (1993) 1155–1161, https://doi.org/10.1016/S0006-3495(93)81145-7.
[24] S. Kang, M. Yang, Z. Hong, L. Zhang, Z. Huang, X. Chen, S. He, Z. Zhou, Z. Zhou, Q. Chen, Y. Yan, C. Zhang, H. Shan, S. Chen, Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites, Acta Pharm. Sin. B (2020), https://doi.org/10.1016/j.apsb.2020.04.009.
[25] R. Zhou, R. Zeng, A. von Brunn, J. Lei, Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein, Mol. Biomed. 1 (2020) 1–11, https://doi.org/10.1186/s43556-020-00001-4.
[26] P. Bernado, E. Mylonas, M.V. Petoukhov, M. Blackledge, D.I. Svergun, Structural characterization of flexible proteins using small-angle X-ray scattering, J. Am. Chem. Soc. 129 (2007) 5656–5664, https://doi.org/10.1021/ja069124n.
[27] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF chimera–a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612, https://doi.org/10.1002/jcc.20084.
[28] A. Ortega, D. Amoros, J. Garcia de la Torre, Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models, Biophys. J. 101 (2011) 892–898, https://doi.org/10.1016/j.bpj.2011.06.046.
[29] H.J. Kuh, S.H. Jang, M.G. Wientjes, J.L.S. Au, Computational model of intracellular pharmacokinetics of paclitaxel, J. Pharmacol. Exp. Ther. 293 (2000) 761–770.
[30] S.Y. Lin, C.L. Liu, Y.M. Chang, J. Zhao, S. Perlman, M.H. Hou, Structural basis for the identification of the N-terminal domain of coronavirus nucleocapsid protein as an antiviral target, J. Med. Chem. 57 (2014) 2247–2257, https://doi.org/10.1021/jm500089r.
[31] D.P. Valencia, F.J. González, Understanding the linear correlation between diffusion coefficient and molecular weight. A model to estimate diffusion coefficients in acetonitrile solutions, Electrochem. Commun. 13 (2011) 129–132, https://doi.org/10.1016/j.elecom.2010.11.032.
[32] R.G.D. Steel, A rank sum test for comparing all pairs of treatments, Technometrics. 2 (1960) 197–207, https://doi.org/10.1080/00401706.1960.10489894.
[33] M. Dwass, Some k-sample rank-order tests, in: S.S.G.I. Olkin, W. Hoeffding, W. G. Madow, H.B. Mann (Eds.), Contrib. to Probab. Stat., Stanford University Press, 1960, pp. 198–202.
[34] F. Wilcoxon, Individual comparisons by ranking methods, Biom. Bull. 1 (1945) 80–83, https://doi.org/10.2307/3001968.
[35] M.H. Nguyen, M. Martin, H. Kim, F. Gabel, O. Walker, M. Hologne, Molecular recognition of ubiquitin and Lys63-linked diubiquitin by STAM2 UIM-SH3 dual domain: the effect of its linker length and flexibility, Sci. Rep. 9 (2019) 1–12, https://doi.org/10.1038/s41598-019-51182-0.
[36] P.K. Ankush Jagtap, M. Müller, P. Masiewicz, S. von Bülow, N.M. Hollmann, P.-C. Chen, B. Simon, A.W. Thomae, P.B. Becker, J. Hennig, Structure, dynamics and roX2-lncRNA binding of tandem double-stranded RNA binding domains dsRBD1,2 of Drosophila helicase Maleless, Nucleic Acids Res. 47 (2019) 4319–4333, https://doi.org/10.1093/nar/gkz125.
[37] J. Kopitz, S. Ballikaya, S. André, H.-J. Gabius, Ganglioside GM1/galectin-dependent growth regulation in human neuroblastoma cells: special properties of bivalent Galectin-4 and significance of linker length for ligand selection, Neurochem. Res. 37 (2012) 1267–1276, https://doi.org/10.1007/s11064-011-0693-x.
[38] M. Kovermann, P. Rogne, M. Wolf-Watz, Protein dynamics and function from solution state NMR spectroscopy, Q. Rev. Biophys. 49 (2016) 1–43, https://doi.org/10.1017/S0033583516000019.
[39] F. Gabel, D. Bicout, U. Lehnert, M. Tehei, M. Weik, G. Zaccai, Protein dynamics studied by neutron scattering, Q. Rev. Biophys. 35 (2002) 327–367, https://doi.org/10.1017/s0033583502003840.
[40] T. Matsuo, Usefulness of medium-angle X-ray scattering for structural characterization of flexible proteins studied by computer simulations, Biochem. Biophys. Res. Commun. 525 (2020) 830–835, https://doi.org/10.1016/j.bbrc.2020.02.150.