THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2020, 124, 20−27

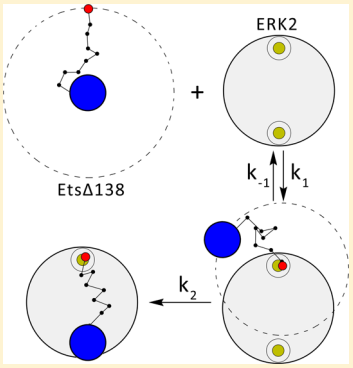
pubs.acs.org/JPCB

Role of Intrinsically Disordered Regions in Acceleration of Protein−Protein Association

Mikita M. Misiura[†] and Anatoly B. Kolomeisky*[,†,‡,¶]

[†]Department of Chemistry and Center for Theoretical Biological Physics, [‡]Department of Chemical and Biomolecular Engineering, and [¶]Department of Physics and Astronomy, Rice University, Houston, Texas 77005, United States

ABSTRACT: Although intrinsically disordered proteins and intrinsically disordered regions (IDRs) in folded proteins are not able to form stable structures, it is known that they play critically important roles in various biological processes. However, despite multiple studies, the molecular mechanisms of their functions remain not fully understood. In this work, we theoretically investigate the role of IDRs in acceleration of protein−protein association processes. Our hypothesis is that, in protein pairs with several independent binding sites, the association process goes faster if some of these binding sites are located on IDRs or connected by IDRs. To test this idea, we employed analytical modeling, numerical calculations, and Brownian dynamics computer simulations to calculate protein−protein association reaction rates for the ERK2−EtsΔ138 system, belonging to the RAS−RAF−MEK−ERK signaling pathway in living cells. It is found that putting a binding site on IDR accelerates the association process by a factor of 3 to 4. Possible molecular explanations for these observations are presented, and other systems that might use this mechanism are also mentioned.



INTRODUCTION

For a long time, it was generally accepted that, in order to function, proteins must have particular well-defined 3D structures.[1] This is known as structure−function paradigm in biology. Any disorder in the protein structures was considered to be harmful for their functionality. However, starting from 1950s, many unstructured proteins, now called intrinsically disordered proteins (IDPs), and proteins with some unstructured regions, called intrinsically disordered regions (IDRs), have been discovered.[2,3] It is now clear that some proteins can be perfectly functional while having no particular structure or while being only partially folded.[4,5] Such proteins (or regions of proteins) are characterized by relatively low sequence complexities, frequent repeats, large fractions of hydrophilic, polar, and charged amino acids, and low fractions of bulky hydrophobic amino acids.[2−7]

In recent years, IDPs and IDRs have been intensively investigated using a variety of experimental and computational methods.[2,3,8] It is now well established that IDPs and IDRs are frequently found in nature, and the number of such proteins is increasing with increasing complexity of living organisms.[3] Several advantages of having disordered domains in proteins in comparison with rigidly structured proteins have been suggested for various biological processes.[3,7,9,10] For example, it was argued that folded proteins function in a relatively narrow range of conditions in which their structures are stable. At the same time, the absence of such well-defined structures in the case of IDPs and IDRs means that there is a broader range of environments in which they can perform their duties, and for this reason, they are remarkably multifunctional. In addition, the lack of single folded structures allows disordered proteins or disordered domains to bind to multiple targets with different structures and to adopt different conformations upon binding. Furthermore, IDPs and IDRs are more flexible in post-translation modifications, and they can also form membraneless organelles, which were shown to be a critically important part of cellular machinery.[3,11] However, despite significant progress in our understanding of IDPs and IDRs, many questions on molecular mechanisms of underlying processes remain unanswered.[3,12]

It was suggested that the presence of disorder might be especially advantageous for association of proteins with other biological molecules. A so-called "fly-casting" mechanism has been proposed to explain the speeding of molecular recognition by proteins with disordered domains.[13,14] Theoretical calculations estimated that partially unfolded proteins might associate with DNA up to 1.6 time faster than fully folded proteins.[13] Partially disordered proteins have somewhat larger sizes and are able to weakly bind to DNA from a longer distance, allowing the protein molecule to start organizing its structure. This brings the protein molecule closer to DNA due to increasing strength of intermolecular interactions, and binding and folding can take place cooperatively. For partially unfolded proteins, the interactions are long-range but weak, while for fully folded proteins, interactions are strong but short-range. This eventually leads to faster binding kinetics to DNA for partially unfolded proteins in comparison with fully folded proteins.[13] A related "dock-and-coalesce" mechanism has been proposed for association of disordered proteins with structured targets when binding is not accompanied by the

Received: September 16, 2019
Revised: December 4, 2019
Published: December 5, 2019

DOI: 10.1021/acs.jpcb.9b08793
J. Phys. Chem. B 2020, 124, 20−27

folding.[8] In this scenario, a segment of IDP first docks to its specific site on the target molecule, simplifying the search of their corresponding sites for other domains, which bind shortly after in a sequential manner. Eventually, all domains coalesce around the original binding site by associating with their corresponding binding sites.

In this paper, we investigate theoretically the kinetics of protein−protein association when one of the proteins has one binding site located on IDR. Our goal is to quantitatively estimate the effects of disordered regions on the dynamics of protein−protein association reactions. We postulate that, if such a process involves more than one binding site, then putting one of the sites on IDR will accelerate the association. To test this hypothesis, we apply our method for a system of two proteins, ERK2 and EtsΔ138, which were shown to bind each other using at least two binding sites (see Figure 1).[15−20]

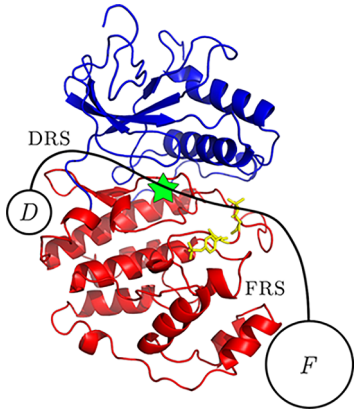


Figure 1. Simplified structural representation of the activated ERK2 protein (2P−ERK2, PDB entry 2ERK) bound to the substrate molecule EtsΔ138. The N-terminal domain of ERK2 is shown in blue, and the C-terminal domain is shown in red. A green star shows an approximate position of the active site. EtsΔ138 is shown as a cartoon (as there are no experimentally determined structures of the complex). The circle with F in the middle marks the bound PNT domain of the EtsΔ138 (residues 40−138) on which the F site is located. The disordered region of EtsΔ138 (residues 1−39) is shown as a solid black line that connects the F circle with the circle representing the D site on IDR.

This system plays a major role in extracellular biological signaling.[15] ERK2 is a globular (fully structured) protein with two binding sites,[21] while the EtsΔ138 protein has a structured PNT domain with one binding site and a disordered peptide chain on which the second binding site is located.[22−25] The binding sites of ERK2 are called DRS and FRS (D recruiting site and F recruiting site), and they bind the D and F sites of EtsΔ138, respectively.[16−18,26−29] The F site of the EtsΔ138 protein is in the structured PNT domain, while the D site is located on IDR. The molecular mechanisms of activation of ERK2 signaling proteins are still not well explained.[15,30] For example, it is not clear why two spatially separated binding sites are required for its action. Our theoretical approach allows us to clarify some aspects of the complex mechanisms of protein−protein association for ERK2.

THEORY

To estimate the effect of IDRs in protein−protein association, we consider a minimal theoretical model as presented in Figure 2. Two association scenarios are considered. First, we analyze the binding of two globular proteins with D_{ERK2} and R_{ERK2} describing the diffusion coefficient and the radius of ERK2, respectively, with sizes of binding sites of ERK2 being R_{FRS} and R_{DRS}, and with D_{PNT} and R_{PNT} describing the diffusion coefficient and the radius of the structured part of EtsΔ138, respectively (Figure 2a). The unstructured part of EtsΔ138 is omitted in this case, and thus, this process describes the association of two proteins without involvement of disordered domains. It is assumed that the reaction is diffusion-limited, and the association rate constant k (see Figure 2a) can be estimated using a Hill−Berg−Purcell formula,[31,32] which describes the irreversible binding of a spherical particle (the PNT domain of EtsΔ138) by a perfectly absorbing circular disk (the FRS binding site) located on an otherwise reflecting sphere (the ERK2 molecule),

$$k = 4(D_{PNT} + D_{ERK2})R_{FRS} \quad (1)$$

where R_{FRS} is a radius of the F recruiting site of ERK2.

Figure 2b shows another scenario of protein−protein association, which involves IDR. In this case, both proteins have two binding sites. The distance between two binding sites on ERK2 (denoted as d) is fixed since ERK2 is fully structured. The distance between two binding sites on EtsΔ138 (labeled r in Figure 2b) is a dynamic quantity since one of them is located on a disordered segment. In our model, the IDR domain is viewed as a polymer chain attached to the first binding site on the structured domain and with the second binding site at another end of the polymer chain (see Figure 2b). The polymer chain is constantly moving, exploring the space around, which leads to varying distances between binding sites of EtsΔ138. Assuming that the polymer segment follows the statistics of a Gaussian chain and can explore fully the surrounding space, we can estimate the probability density function for distances between binding sites[33]

$$p_{eq}(r) \propto 4\pi r^2 \exp\left[-\frac{3r^2}{2 < r^2 >}\right] \quad (2)$$

where $<r^2>$ is an average squared distance between binding sites (averaged over all possible polymer chain conformations). It was already shown that this expression describes quite well the experimental data on end-to-end distance distributions of unfolded peptide chains.[34]

The overall effective kinetic rate constant for the two-step association mechanism in Figure 2b is given by

$$k' = \frac{k_1 \times k_2}{k_1 + k_{-1} + k_2} \simeq \frac{k_2}{1 + K_D + k_2/k_1} \quad (3)$$

where $K_D = k_{-1}/k_1$ is a dissociation equilibrium constant of the intermediate state with only the D site bound to DRS (Figure 2b). In addition, the rate constant $k_1 = k_{on}/V$ has also a dimension of inverse time because we assumed a unit volume in our theoretical analysis. The intermediate state is unstable and short-lived because the created bond D−DRS is quite weak and the system can easily dissociate back into the unbound state or proceed forward with the association at the second binding site F (to the FRS site).

To calculate the effective rate constant k' for the two-step association process, one needs to evaluate the parameters K_D, k_1, and k_2. The dissociation equilibrium constant can be obtained from the experimental measurements. To estimate k_1, the binding rate constant for the first step, we again employ the Hill−Berg−Purcell formula but with some modifications to

DOI: 10.1021/acs.jpcb.9b08793
J. Phys. Chem. B 2020, 124, 20−27

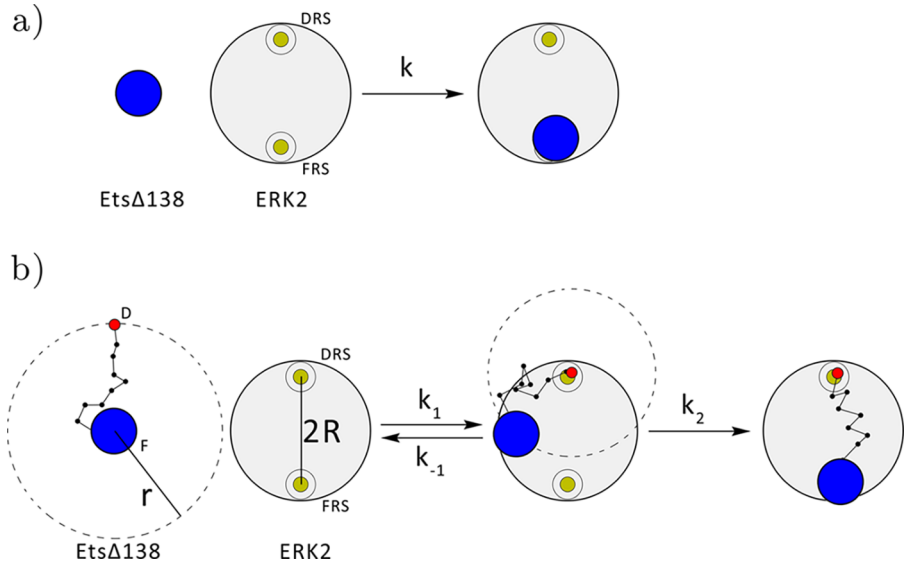


Figure 2. Minimal theoretical model to investigate the role of disordered domains in protein−protein association. (a) One-step binding of two globular proteins without IDR; (b) two-step binding of a globular protein and a protein with the binding site on IDR. Details are in the text.

reflect the fact that the binding is reversible. This time EtsΔ138 represents an absorbing circular disk, which binds to the ERK2 molecule

$$k_1 = \frac{k_{on}}{V} = \frac{4\pi(D_{PNT} + D_{ERK2})\langle r\rangle P_{bind}}{V} \quad (4)$$

where $\langle r\rangle$ is the average distance between binding sites of EtsΔ138 and P_{bind} is the probability of binding during each encounter. In this expression, we view r as an effective radius of EtsΔ138, which is larger than that of the rigid part alone (and can also be significantly larger than ERK2). The idea behind this approximation is the following. The binding site D is located on IDR, and it diffuses significantly faster than the structured PNT domain of the EtsΔ138 protein. Thus, the whole protein can be viewed as an effectively larger particle with a radius $\langle r\rangle$ (shown in Figure 2b as a dashed circle). Because the structured part of the EtsΔ138 protein (the PNT domain) is much larger and heavier than the segment with IDR, the overall diffusion constant of the center of mass of this protein can still be approximated by D_{PNT}. The phenomenological parameter P_{bind} reflects the reversibility of this process, that is, that not every encounter of two proteins will end up with them binding together. In this work, we use $P_{bind} = 0.1$ because the produced intermediate complex is assumed to be quite weak. The exact value of this parameter does not influence the physics of this process as long as it is quite small, which is realistic in our case.

The estimation of the rate constant k_2 is a more difficult task. To calculate it, we simplify the problem and assume that the D site of EtsΔ138 is bound to the center of ERK2, which allows us to reduce the problem to be one-dimensional (1D). Here, we utilize the theoretical approach developed by Szabo, Schulten, and Schulten (SSS), who calculated the rates of intramolecular reactions between two reacting groups connected by a polymer chain.[35] Below we present the detailed explanations of how we evaluate the rate constant k_2.

Figure 3 illustrates the method of calculation for the rate k_2. We have the D binding site on EtsΔ138 already bound to the DRS site of ERK2. Although the F site actively explores 3D space around the DRS, the association with the FRS site can be

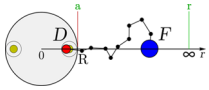


Figure 3. Theoretical estimation of the k_2 rate constant. The red vertical line at $r = R$ (radius of ERK2) denotes the absorbing boundary. To calculate k_2, we simplify the problem and assume that the D site of EtsΔ138 is bound to the center of ERK2, which allows us to reduce the problem to be one-dimensional. The finite surface binding rate is introduced to account for the presence of the ERK2 because not every trajectory will end up with the F site bound to the FRS of ERK2 (see text for details).

viewed as an effective 1D motion along the axis that connects the DRS and FRS sites (see Figure 3). Here, we assume that the rotational diffusion of the F site is so fast allowing for the problem to be reduced to the effective one-dimensional search for the site FRS. Hence, in order to calculate k_2, we have to estimate the mean first-passage time for the F binding site on EtsΔ138 to reach the position R starting from a specific configuration at time zero. When $R = 0$, the problem is equivalent to calculating the rate constants of intramolecular polymer chain reactions that was analyzed before.[35−38]

Employing the theoretical framework developed by SSS, we can estimate the mean time for the reaction as

$$\tau = \frac{1}{k_2} = \frac{1}{D_F}\int_R^\infty \frac{dr}{p_{eq}(r)}\left[\int_r^\infty p_{eq}(x)dx\right]^2 + \frac{1}{\kappa_F p_{eq}(R)} \quad (5)$$

where κ_F is the surface binding rate. This equation means that the reaction time is a sum of two contributions. One of them is the time to reach the surface (first term), and another one is the time to react at the surface (second term). The surface binding rate can be estimated by combining Collins−Kimball and Hill−Berg−Purcell formulas,[31,32,39] leading to

$$k = 4D_F R_{FRS} = \frac{4\pi D_F R_{ERK2} 4\pi R_{ERK2}^2 \kappa_F}{4\pi D_F R_{ERK2} + 4\pi R_{ERK2}^2 \kappa_F} \quad (6)$$

From this equation, it can be easily found that

DOI: 10.1021/acs.jpcb.9b08793
J. Phys. Chem. B 2020, 124, 20−27

$$\kappa_F = \frac{D_F R_{FRS}}{\pi {R_{ERK2}}^2 (1 - R_{FRS}/\pi R_{ERK2})} \quad (7)$$

It should be noted that, in eqs 5−7, we assume that $D_F = D_{ERK2} + D_{PNT}$. Finally, the explicit values for the rate constant k' can be obtained by solving numerically eqs 2−5 and 7.

To quantify the role of the disordered region, we consider the ratio of rate constants, $k_{eff}/k = (k + k')/k$, for the reaction of association with the IDR segment and without the IDR segment. One should note that the effective rate constant $k_{eff} = k + k'$ reflects the fact that association in the case of two binding sites might still proceed via the one-step mechanism, explaining the summation of both terms.

SIMULATIONS

Our theoretical model of the role of IDRs in protein−protein association involves several approximations, so we also performed Brownian dynamics (BD) computer simulations to test our hypothesis. Here, we employed the method of computational calculation for protein−protein binding rate constants developed by Northrup et al.[40,41] The method works as follows. One of the protein molecules is fixed at the center of coordinates, and the other one is placed at some distance b from the first one. The mutual orientation of two molecules is randomly chosen, and b is selected in such a way that the interaction between two proteins can be viewed as isotropic at this distance. Then BD simulation is started and is continued until two proteins are either bound or until the second protein molecule moves too far from the first molecule. For the latter scenario, a cutoff distance q is specified (usually set to be equal to $3b$ or larger). The simulation is repeated multiple times, and the probability β for two proteins to bind each other is obtained as a fraction of successful binding events to the number of simulations. The desired rate constant can then be calculated using the following expressions, which can be viewed as a generalization of a Debye−Smoluchowskii equation[40]

$$k = \frac{4\pi D b \beta}{1 - (1 - \beta)\frac{k_D(b)}{k_D(q)}} \quad (8)$$

$$k_D(x) = \frac{4\pi D}{\int_x^\infty \exp\left[\frac{U(r)}{k_B T}\right]\frac{dr}{r^2}} \quad (9)$$

where $U(r)$ is the intermolecular interaction potential between two protein molecules.

The Gromacs 4.5.4 MD package with custom interaction potentials was utilized to perform BD computer simulations.[42−44] The geometry of BD simulations is shown schematically in Figure 4. ERK2 was fixed in the center of coordinates and represented by two beads corresponding to its two binding sites (the FRS and the DRS sites). EtsΔ138 was represented as a string of beads connected by stiff springs. The first bead in the chain represented the F binding site, and it was assumed to be larger than all the other beads (the radius of the F site is 1.5 nm; for other beads, the radius is 0.15 nm), and the last bead represented the D binding site. In all simulations, the F site has a smaller diffusion coefficient than all other beads (which all have identical diffusion coefficients). The number of beads was varied as explained below. Both binding sites of EtsΔ138 were set to interact with their recruiting sites of ERK2 using Gaussian-like potentials shown in Figure 4B. The barrier

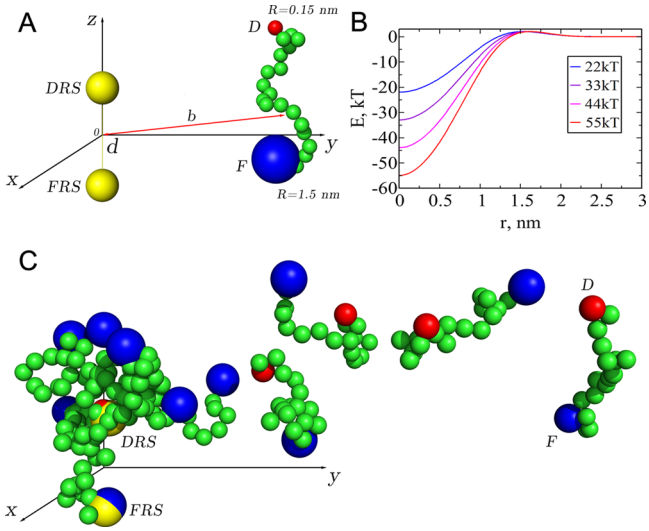


Figure 4. (A) Geometry of the system for computer simulations (sphere sizes and distances are not to scale). ERK2 is represented by two connected yellow beads, corresponding to its two binding sites. EtsΔ138 is represented using a large bead for the PNT domain (shown in blue) connected by a chain of identical beads representing IDR (shown in green) to the binding site D (shown in red). (B) Examples of potentials used to describe the interactions between the DRS and D sites and between the FRS and F sites. The height of the barrier is constant, and it is kept at $2k_BT$. Details are in the text. (C) Example trajectory of EtsΔ138 binding to ERK2 (sphere sizes are not to scale). ERK2 is fixed on the left, while EtsΔ138 starts on the right and slowly diffuses toward ERK2. First, the D site is bound to DRS, and then, after a while, the F site is bound to FRS.

for the binding was always assumed to be $2k_BT$, while the binding energy was varied in the range of 11 to $55k_BT$.

In the beginning of each simulation run, the EtsΔ138 protein configuration was randomly generated at a random location on a sphere of radius b = 12 nm with the center on ERK2. The simulation was then continued until the F site of EtsΔ138 was bound or until EtsΔ138 diffused further than q = 40 nm from ERK2. For each data point, at least 1000 simulations were performed to collect good statistics on the probability of binding β. First 500 runs were performed with the DRS−D site interaction turned on (the effect of IDR is taken into account), and then 500 runs were performed with identical parameters; however, when the DRS−D site interactions were turned off, the effect of IDR is neglected. Then the association constants were calculated using eqs 8 and 9. The resulting ratio of rate constants is denoted as $(k + k')/k$ to match the notation we used in analytical calculations. Error bars (standard deviations) are estimated using a standard bootstrapping approach.

RESULTS AND DISCUSSION

Currently, there is no clear understanding on why ERK2 has two binding sites, which are located at some distance from its catalytic site.[15,16,45] In this work, we propose a hypothesis that one of the binding sites is needed to speed up the overall binding process, while the other side orients the substrate molecule into a position inside the catalytic site to be phosphorylated more efficiently. This idea is supported by the fact that EtsΔ138 and ERK2 were shown to associate very fast.[46] The experimentally measured association rate constant for these proteins is $k_{on} = 5 \times 10^6$ M^{-1} s^{-1}, which is about 5

DOI: 10.1021/acs.jpcb.9b08793
J. Phys. Chem. B 2020, 124, 20−27

times faster than the diffusion limit for globular proteins in the absence of long-range interactions.[8,47] This clearly indicates that there is a special mechanism of accelerated association between these proteins, and our theoretical method presents a possible scenario for this to happen.

Results from our numerical calculations are presented in Figure 5. The radius R of ERK2 was set to 2.5 nm (the blue vertical line in Figure 5), as estimated from the structure of ERK2. The radius of the F recruiting site of ERK2 was set to be equal to 0.5 nm. The average distance between D and F binding sites of EtsΔ138 ($\sqrt{<r^2>}$ to be exact) was varied in the range of 1 to 250 nm. As one can see, the proposed two-step binding mechanism indeed leads to faster association in comparison with the one-step binding scenario. Our theoretical results suggest that this acceleration is well within 1 order of magnitude, and it is highly dependent on the average distance between the binding sites of EtsΔ138. For every value of the equilibrium dissociation constant K_D for the intermediate state, there is an optimum distance at which the maximal acceleration is achieved. The smaller the K_D (the intermediate state is more stable), the larger the optimal distance (see Figure 5). The experimentally estimated value of K_D for EtsΔ138 is 30 μM, giving the maximum possible acceleration to be close to 3.[17] For shorter than optimal distances, the formation of the intermediate state is slow because the IDR-located binding site is too close to the slowly diffusing F binding site, effectively searching a smaller volume around the F site and leading to smaller k_1 rates (see Figure 2b). For larger than optimal distances, the formation of the intermediate state is a fast step (k_1 is large), and the rate k_2 becomes rate-limiting. This is because, after the D site is found, the F site is on average too far from the D site and hence from the FRS site, and it has to search a larger volume before associating with the FRS (Figure 2b). It should be also noted that our theory works better when the average distance between binding sites of EtsΔ138 is significantly larger than the size of ERK2 and gets less accurate when r approaches R.

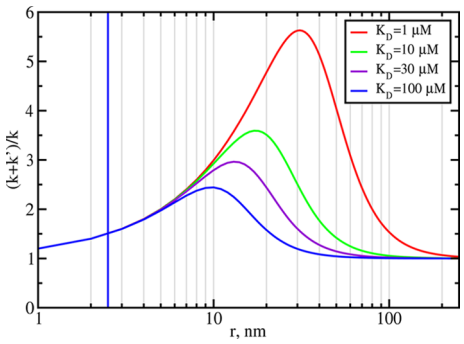


Figure 5. Numerical estimates of the acceleration in the protein−protein association rate in the presence of the IDR-located region as a function of the distance between binding sites of EtsΔ138. The blue vertical line marks the radius of ERK2. Parameters used: R_{ERK2} = 2.5 nm, R_F = 0.5 nm, P_{bind} = 0.1.

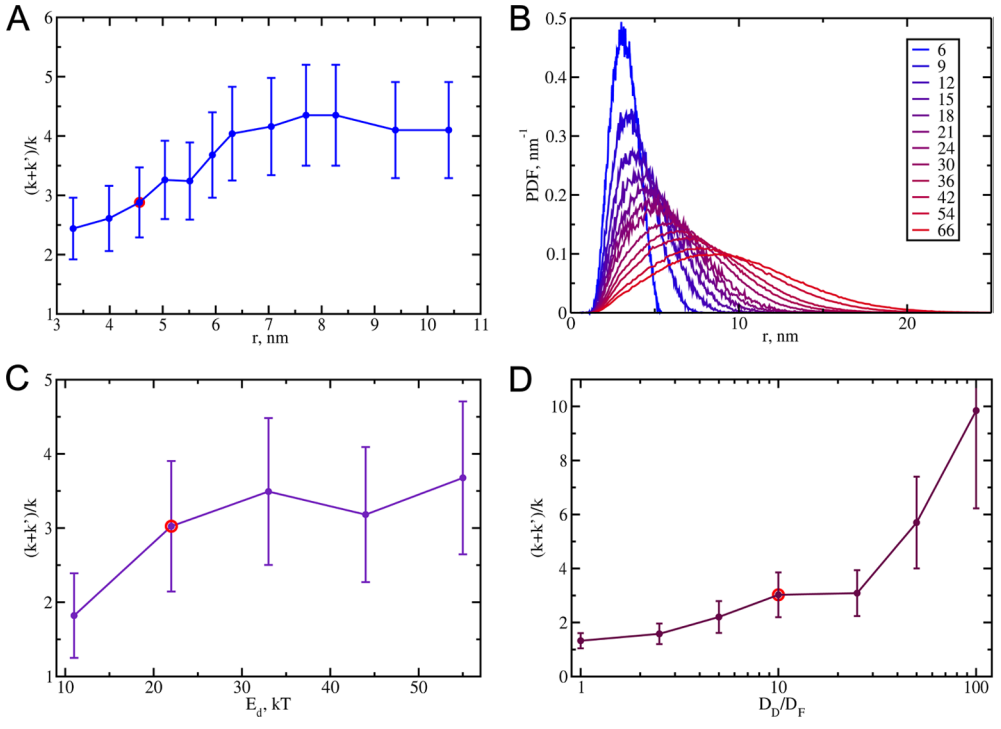


Figure 6. Brownian dynamics computer simulation results for protein−protein association. Error bars show standard deviations. (A) Acceleration of the protein−protein association rate as a function of average distance between binding sites of EtsΔ138. In these simulations, D_D/D_F = 10 is used, and both binding energies are equal to $22k_BT$. The red circle marks the position of what we consider the most realistic value for our ERK2−EtsΔ138 system. (B) Distributions of distances between D and F binding sites of EtsΔ138 from the simulations in panel (A). The legend shows then the numbers of beads utilized to simulate IDR. The larger is the number of beads, the larger is the average distance between the binding sites. (C) Acceleration of the protein−protein association rate as a function of the binding energy between the DRS site on ERK2 and the D site on EtsΔ138. For simulations, D_D/D_F = 10 is used, and the binding energy of the F site is equal to $22k_BT$. (D) Acceleration of the protein−protein association rate as a function of the D_D/D_F ratio. Binding energies for both D and F sites are equal to $22k_BT$.

DOI: 10.1021/acs.jpcb.9b08793
J. Phys. Chem. B 2020, 124, 20−27

Our numerical calculations are based on the approximate one-dimensional description of the association process with the additional assumption that the D site diffuses significantly faster than the F site. Thus, to verify our theoretical results, we performed BD computer simulations with a more realistic description of the system under study. The results from simulations are presented in Figure 6. One can see that, in agreement with numerical calculations (Figure 5), increasing the distance between the binding sites on EtsΔ138 leads to increased acceleration of protein−protein association. Overall, the BD computer simulations predict the acceleration of protein−protein association to be of the order of 3 to 4, which is consistent with our theoretical results. One difference between our theoretical predictions and theoretical results is that, in BD simulations, we could not reach the regime where increasing the average distance between binding sites of EtsΔ138 leads to a decreased association rate, although we do see some decline for the largest average distances in Figure 6A. The actual range of parameters we need to see a significant decrease is unreachable in our simulations because, to reach it, we have to significantly increase the number of beads (see Figure 6B), which makes the computational cost of such simulation prohibitive. Red circles in Figure 6 mark, in our opinion, sets of the most realistic parameters to describe the ERK2−EtsΔ138 system.

We estimated the binding energy between the D site on EtsΔ138 and the DRS site on ERK2 to be $22k_BT$. However, this quantity has never been measured, and we decided to test the effect of varying this quantity on the acceleration of the protein−protein association rate in the BD computer simulations, as illustrated in Figure 6C. The results show that the acceleration increases for stronger interaction energies. However, the effect is quite weak: increasing the binding energy from 11 to $55k_BT$ only changes the acceleration from ∼2 to ∼3.5. Thus, the acceleration weakly depends on the strength of the binding interaction, and our choice of the binding energy seems to be appropriate.

Another important factor in the protein−protein association of EtsΔ138 and ERK2 is the relative values of the diffusion coefficients for the D and F sites. Since the F site is located on the structured PNT domain of EtsΔ138, it is reasonable to assume that it diffuses slower than the D site. However, the exact ratio D_D/D_F is unknown and is not an easy parameter to estimate. In previous simulations and theoretical calculations, we set it equal to 10 as we think that this value seems to be the most plausible (based, for example, on measurements of diffusion coefficients of peptides and amino acids[48] and estimations using the Stokes−Einstein formula), but testing the different values of diffusion coefficients is necessary. Figure 6D shows the results of the BD simulations for different D_D/D_F ratios. When the ratio is not large (up to $D_D/D_F = 10$), the acceleration is almost constant and close to 2. This can be understood using the following arguments. The two-step mechanism opens a new channel for achieving a protein−protein association. When the diffusion constants for the D and F sites are comparable, it is similar to having two parallel reactions with almost the same reaction rates, and this explains the factor of 2. As D_D/D_F reaches 100, the acceleration increases up to about 10, although we think that such a ratio is unrealistic for this particular system. However, if we take into account that EtsΔ138 is just a part (first 138 residues) of a much larger transcription factor protein Ets1, the D_D/D_F ratio for the full substrate working in vivo might be larger due to the presence of the ETS domain connected to the PNT domain via another long disordered protein region.[20] This structured domain might serve as additional weight (or "anchor") and might slow down the PNT domain even further while not affecting the diffusion of the D site too much. In our theoretical approach, this situation would correspond to higher values of the parameter P_{bind}. These arguments suggest that, in live cells, the effect of the acceleration due to the two-step mechanism might be even stronger.

Overall, both theory and simulations predict modest acceleration of the protein−protein binding in the presence of IDR-located binding sites. For a wide range of parameters, we estimate the acceleration to be in the range of approximately 3 to 4. This is consistent with experimental observations on association rates for EtsΔ138 and ERK2 being about 5 times faster than the diffusion limit for globular proteins of the same size.[46] It is important to note that many other substrates of ERK2 were also shown to bind it at multiple sites,[15,20,49−51] and many of them utilize the DRS site on ERK2. Interestingly, kinases MEK1 and MEK2 that activate ERK2 in vivo also have intrinsically disordered regions that they use to bind ERK2.[24,52−54] This suggests that they might potentially employ the same binding acceleration mechanism that we describe in this work for the ERK2−EtsΔ138 system. Given that both the activation of ERK2 and the phosphorylation of substrates by active ERK2 belong to the RAS−RAF−MEK−ERK signaling cascade, IDR-accelerated interactions can serve to significantly accelerate the propagation of biological signals. For example, if we assume the moderate value of acceleration (3 times) for both the ERK2 activation and the phosphorylation of Ets1 by active ERK2 molecules, then the whole process accelerates by a factor of $3 \times 3 = 9$, which can be crucial for on-time delivery of extracellular signals.

It is important to note that the protein−protein binding acceleration mechanism proposed here is fully consistent with the general dock-and-coalesce picture of the binding of disordered proteins with structured targets.[8] In our model, the D site on the flexible IDR segment first attaches to the corresponding site on the rigid ERK2 molecule, which allows for the F sites to find its corresponding site faster.

SUMMARY AND CONCLUSIONS

We developed a new theoretical approach to investigate the role of IDR in protein−protein association processes. A hypothesis that, for systems with several binding sites, putting at least one of them in a disordered region would accelerate binding is proposed and tested using analytical arguments, numerical calculations, and BD computer simulations. For theoretical analysis, the binding of the ERK2 enzyme to the EtsΔ138 substrate is studied. A modest acceleration by a factor of 3 to 4 in the protein−protein association rates is observed for a large range of parameters, which is consistent with experimental observations showing that the binding rate for this system is larger than the diffusion limit for globular proteins. Thus, our theoretical picture provides a molecular explanation for the existence of several binding sites in many enzymatic systems. The importance of this mechanism for faster delivery of biological signals is also discussed.

Although the theoretical model presented in this work allows us to study complex processes that take place during the protein−protein association, it is important to emphasize the limitations of our theoretical method, which neglects many

DOI: 10.1021/acs.jpcb.9b08793
J. Phys. Chem. B 2020, 124, 20−27

important features of these processes. These limitations include the consideration of protein molecules as purely globular, given that the known structures is not entirely correct; the assumption of equilibrium for the motion of the disordered region, that is, that it is able to visit all regions in the space before the reaction is taking place; and the assumption of fast rotational motion of protein domains in comparison with other transitions in the system. However, despite the use of multiple simplifications and approximations, our theoretical approach is at the very least able to clarify some molecular details of the protein−protein association. It will be important to test our theoretical predictions with experimental methods and with more advanced theoretical computations.

AUTHOR INFORMATION

Corresponding Author

*E-mail: tolya@rice.edu.

ORCID

Anatoly B. Kolomeisky: 0000-0001-5677-6690

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was supported by the Welch Foundation (Grant C-1559), by the NSF (Grant CHE-1664218), and by the Center for Theoretical Biological Physics sponsored by the NSF (Grant PHY-1427654).

REFERENCES

(1) Lodish, H.; Berk, A.; Kaiser, C. A.; Krieger, M.; Scott, M. P.; Bretscher, A.; Ploegh, H.; Matsudaira, P. *Molecular Cell Biology*; Macmillan, 2008.

(2) Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **2014**, *83*, 553−584.

(3) Uversky, V. N. Intrinsically Disordered Proteins and Their 'Mysterious'(Meta) Physics. *Front. Phys.* **2019**, *7*, 10.

(4) Wright, P. E.; Dyson, H. J. Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure-Function Paradigm. *J. Mol. Biol.* **1999**, *293*, 321−331.

(5) Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D.; et al. Extreme Disorder in an Ultrahigh-Affinity Protein Complex. *Nature* **2018**, *555*, 61.

(6) Uversky, V. N. A Decade and a Half of Protein Intrinsic Disorder: Biology Still Waits for Physics. *Protein Sci.* **2013**, *22*, 693−724.

(7) Wright, P. E.; Dyson, H. J. Intrinsically Disordered Proteins in Cellular Signalling and Regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18.

(8) Zhou, H.-X.; Bates, P. A. Modeling Protein Association Mechanisms and Kinetics. *Curr. Opin. Struct. Biol.* **2013**, *23*, 887−893.

(9) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradović, Z.; Dunker, A. K. Intrinsic Disorder in Cell-Signaling and Cancer-Associated Proteins. *J. Mol. Biol.* **2002**, *323*, 573−584.

(10) Fuxreiter, M. Fuzziness: Linking Regulation to Protein Dynamics. *Mol. BioSyst.* **2012**, *8*, 168−177.

(11) Uversky, V. N. Intrinsically Disordered Proteins in Overcrowded Milieu: Membrane- Less Organelles, Phase Separation, and Intrinsic Disorder. *Curr. Opin. Struct. Biol.* **2017**, *44*, 18−30.

(12) Drake, J. A.; Pettitt, B. M. Force Field-Dependent Solution Properties of Glycine Oligomers. *J. Comput. Chem.* **2015**, *36*, 1275−1285.

(13) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. Speeding Molecular Recognition by Using the Folding Funnel: the Fly-Casting Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8868−8873.

(14) Huang, Y.; Liu, Z. Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding-Binding Process: a Critical Assessment of the Fly-Casting Mechanism. *J. Mol. Biol.* **2009**, *393*, 1143−1159.

(15) Futran, A. S.; Link, A. J.; Seger, R.; Shvartsman, S. Y. ERK as a Model for Systems Biology of Enzyme Kinetics in Cells. *Curr. Biol.* **2013**, *23*, R972−R979.

(16) Lee, S.; Warthaka, M.; Yan, C.; Kaoud, T. S.; Ren, P.; Dalby, K. N. Examining Docking Interactions on ERK2 With Modular Peptide Substrates. *Biochemistry* **2011**, *50*, 9500−9510.

(17) Callaway, K. A.; Rainey, M. A.; Riggs, A. F.; Abramczyk, O.; Dalby, K. N. Properties and Regulation of a Transiently Assembled ERK2 Ets-1 Signaling Complex. *Biochemistry* **2006**, *45*, 13719−13733.

(18) Waas, W. F.; Rainey, M. A.; Szafranska, A. E.; Dalby, K. N. Two Rate-Limiting Steps in the Kinetic Mechanism of the Serine/Threonine Specific Protein Kinase ERK2: a Case of Fast Phosphorylation Followed by Fast Product Release. *Biochemistry* **2003**, *42*, 12273−12286.

(19) Lee, T.; Hoofnagle, A. N.; Kabuyama, Y.; Stroud, J.; Min, X.; Goldsmith, E. J.; Chen, L.; Resing, K. A.; Ahn, N. G. Docking Motif Interactions in MAP Kinases Revealed by Hydrogen Exchange Mass Spectrometry. *Mol. Cell* **2004**, *14*, 43−55.

(20) Sharrocks, A. D. The ETS-Domain Transcription Factor Family. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 827.

(21) Zhang, F.; Strand, A.; Robbins, D.; Cobb, M. H.; Goldsmith, E. J. Atomic Structure of the MAP Kinase ERK2 at 2.3 Å Resolution. *Nature* **1994**, *367*, 704.

(22) Slupsky, C. M.; Gentile, L. N.; Donaldson, L. W.; Mackereth, C. D.; Seidel, J. J.; Graves, B. J.; McIntosh, L. P. Structure of the Ets-1 Pointed Domain and Mitogen- Activated Protein Kinase Phosphorylation Site. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 12129−12134.

(23) Piserchio, A.; Warthaka, M.; Kaoud, T. S.; Callaway, K.; Dalby, K. N.; Ghose, R. Local Destabilization, Rigid Body, and Fuzzy Docking Facilitate the Phosphorylation of the Transcription Factor Ets-1 by the Mitogen-Activated Protein Kinase ERK2. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E6287−E6296.

(24) Jacobs, D.; Glossip, D.; Xing, H.; Muslin, A. J.; Kornfeld, K. Multiple Docking Sites on Substrate Proteins Form a Modular System That Mediates Recognition by ERK MAP Kinase. *Genes Dev.* **1999**, *13*, 163−175.

(25) Mackereth, C. D.; Schärpf, M.; Gentile, L. N.; MacIntosh, S. E.; Slupsky, C. M.; McIntosh, L. P. Diversity in Structure and Function of the Ets Family PNT Domains. *J. Mol. Biol.* **2004**, *342*, 1249−1264.

(26) Callaway, K.; Waas, W. F.; Rainey, M. A.; Ren, P.; Dalby, K. N. Phosphorylation of the Transcription Factor Ets-1 by ERK2: Rapid Dissociation of ADP and Phospho-Ets-1. *Biochemistry* **2010**, *49*, 3619−3630.

(27) Abramczyk, O.; Rainey, M. A.; Barnes, R.; Martin, L.; Dalby, K. N. Expanding the Repertoire of an ERK2 Recruitment Site: Cysteine Footprinting Identifies the D-recruitment Site as a Mediator of Ets-1 Binding. *Biochemistry* **2007**, *46*, 9174−9186.

(28) Rainey, M. A.; Callaway, K.; Barnes, R.; Wilson, B.; Dalby, K. N. Proximity-Induced Catalysis by the Protein Kinase ERK2. *J. Am. Chem. Soc.* **2005**, *127*, 10494−10495.

(29) Seidel, J. J.; Graves, B. J. An ERK2 Docking Site in the Pointed Domain Distinguishes a Subset of ETS Transcription Factors. *Genes Dev.* **2002**, *16*, 127−137.

(30) Misiura, M.; Kolomeisky, A. B. Kinetic Network Model to Explain Gain-of-Function Mutations in ERK2 Enzyme. *J. Chem. Phys.* **2019**, *150*, 155101.

(31) Hill, T. L. Effect of Rotation on the Diffusion-Controlled Rate of Ligand-Protein Association. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 4918−4922.

(32) Berg, H. C.; Purcell, E. M. Physics of Chemoreception. *Biophys. J.* **1977**, *20*, 193−219.

(33) Rubinstein, M.; Colby, R. H. *Polymer Physics*; Oxford University Press: New York, 2003; Vol. 23.

DOI: 10.1021/acs.jpcb.9b08793
J. Phys. Chem. B 2020, 124, 20−27

(34) Möglich, A.; Joder, K.; Kiefhaber, T. End-to-End Distance Distributions and Intrachain Diffusion Constants in Unfolded Polypeptide Chains Indicate Intramolecular Hydrogen Bond Formation. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 12394−12399.
(35) Szabo, A.; Schulten, K.; Schulten, Z. First Passage Time Approach to Diffusion Controlled Reactions. *J. Chem. Phys.* **1980**, *72*, 4350−4357.
(36) Deutch, J. M. A Simple Method for Determining the Mean Passage Time for Diffusion Controlled Processes. *J. Chem. Phys.* **1980**, *73*, 4700−4701.
(37) Kramers, H. A. Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions. *Physica* **1940**, *7*, 284−304.
(38) Hänggi, P.; Talkner, P.; Borkovec, M. Reaction-Rate Theory: Fifty Years after Kramers. *Rev. Mod. Phys.* **1990**, *62*, 251.
(39) Collins, F. C.; Kimball, G. E. Diffusion-Controlled Reaction Rates. *J. Colloid Sci.* **1949**, *4*, 425−437.
(40) Northrup, S. H.; Allison, S. A.; McCammon, J. A. Brownian Dynamics Simulation of Diffusion-Influenced Bimolecular Reactions. *J. Chem. Phys.* **1984**, *80*, 1517−1524.
(41) Northrup, S. H.; Erickson, H. P. Kinetics of Protein-Protein Association Explained by Brownian Dynamics Computer Simulation. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 3338−3342.
(42) Bekker, H.; Berendsen, H. J. C.; Dijkstra, E. J.; Achterop, S.; Van Drunen, R.; Van der Spoel, D.; Sijbers, A.; Keegstra, H.; Reitsma, B.; Renardus, M. K. R. Gromacs: A Parallel Computer for Molecular Dynamics Simulations. *Physics Computing '92*; World Scientific Publishing: Singapore, 1993; pp 252−256.
(43) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: a Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(44) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1-2*, 19−25.
(45) Waas, W. F.; Dalby, K. N. Transient Protein-Protein Interactions and a Random-Ordered Kinetic Mechanism for the Phosphorylation of a Transcription Factor by Extracellular-Regulated Protein Kinase 2. *J. Biol. Chem.* **2002**, *277*, 12532−12540.
(46) Lee, S.; Warthaka, M.; Yan, C.; Kaoud, T. S.; Piserchio, A.; Ghose, R.; Ren, P.; Dalby, K. N. A Model of a MAPK•Substrate Complex in an Active Conformation: A Computational and Experimental Approach. *PLoS One* **2011**, *6*, No. e18594.
(47) Schreiber, G.; Haran, G.; Zhou, H.-X. Fundamental Aspects of Protein-Protein Association Kinetics. *Chem. Rev.* **2009**, *109*, 839−860.
(48) Longsworth, L. G. Diffusion Measurements, at 25, of Aqueous Solutions of Amino Acids, Peptides and Sugars. *J. Am. Chem. Soc.* **1953**, *75*, 5705−5709.
(49) Zhou, B.; Wu, L.; Shen, K.; Zhang, J.; Lawrence, D. S.; Zhang, Z.-Y. Multiple Regions of MAP Kinase Phosphatase 3 Are Involved in Its Recognition and Activation by ERK2. *J. Biol. Chem.* **2001**, *276*, 6506−6515.
(50) Zhang, J.; Zhou, B.; Zheng, C.-F.; Zhang, Z.-Y. A Bipartite Mechanism for ERK2 Recognition by Its Cognate Regulators and Substrates. *J. Biol. Chem.* **2003**, *278*, 29901−29912.
(51) Jacobs, D.; Beitel, G. J.; Clark, S. G.; Horvitz, H. R.; Kornfeld, K. Gain-of-Function Mutations in the Caenorhabditis Elegans lin-1 ETS Gene Identify a C-Terminal Regulatory Domain Phosphorylated by ERK MAP Kinase. *Genetics* **1998**, *149*, 1809−1822.
(52) Roskoski, R., Jr. MEK1/2 Dual-Specificity Protein Kinases: Structure and Regulation. *Biochem. Biophys. Res. Commun.* **2012**, *417*, 5−10.
(53) Tanoue, T.; Adachi, M.; Moriguchi, T.; Nishida, E. A Conserved Docking Motif in MAP Kinases Common to Substrates, Activators and Regulators. *Nat. Cell Biol.* **2000**, *2*, 110.
(54) Sheridan, D. L.; Kong, Y.; Parker, S. A.; Dalby, K. N.; Turk, B. E. Substrate Discrimination Among Mitogen-Activated Protein Kinases Through Distinct Docking Sequence Motifs. *J. Biol. Chem.* **2008**, *283*, 19511−19520.

DOI: 10.1021/acs.jpcb.9b08793
J. Phys. Chem. B 2020, 124, 20−27