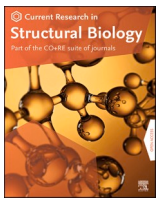
Current Research in Structural Biology 8 (2024) 100158

Contents lists available at ScienceDirect

Current Research in Structural Biology

journal homepage: www.journals.elsevier.com/current-research-in-structural-biology



Research Article

Conformational analysis of the IQSEC2 protein by statistical thermodynamics

Michael Shokhen [a,**], Amnon Albeck [a], Veronika Borisov [b], Yonat Israel [b], Nina S. Levy [b], Andrew P. Levy [b,*]

[a] Department of Chemistry, Bar Ilan University, Ramat Gan, Israel
[b] Technion Faculty of Medicine, Technion Israel Institute of Technology, Haifa, Israel

ARTICLE INFO

Handling editor: A Wlodawer

Keywords:
Accelerated molecular dynamics
Statistical thermodynamics
Ensemble method of Gibbs
Protein folding
Protein dimerization

ABSTRACT

Mutations in the IQSEC2 gene result in severe intellectual disability, epilepsy and autism. The primary function of IQSEC2 is to serve as a guanine exchange factor (GEF) controlling the activation of ARF6 which in turn mediates membrane trafficking and synaptic connections between neurons. As IQSEC2 is a large intrinsically disordered protein little is known of the structure of the protein and how this influences its function. Understanding this structure and function relationship is critical for the development of novel therapies to treat IQSEC2 disease. We therefore sought to identify IQSEC2 conformers in unfolded and folded states and analyze how conformers differ when binding to ARF6 and thereby influence GEF catalysis. We simulated the folding process of IQSEC2 by accelerated molecular dynamics (aMD). Following the ensemble method of Gibbs, we proposed that the number of microstates in the ensemble replicating a protein macroscopic system is the total number of MD snapshots sampled on the production MD trajectory. We divided the entire range of reaction coordinate into a series of consecutive, non-overlapping bins. Thermal fluctuations of biomolecules in local equilibrium states are Gaussian in form. To predict the free energy and entropy of different conformational states using statistical thermodynamics, the density of states was estimated taking into account how many MD snapshots constitute each conformational state. IQSEC2 dimers derived from the most stable folded and unfolded conformers of IQSEC2 were generated by protein-protein docking and then used to construct IQSEC2-ARF6 encounter complexes. We suggest that IQSEC2 folding and dimerization are two competing processes that may be used by nature to regulate the process of GDP exchange on ARF6 catalyzed by IQSEC2.

1. Introduction

IQSEC2 is a guanine nucleotide exchange factor (GEF) promoting the exchange of GDP for GTP on ARF6 leading to its activation (Levy et al., 2023). ARF6 plays a major role in regulating the trafficking of glutamate receptors on neurons and in the growth of dendritic spines playing a key role in the consolidation of memories. Mutations in IQSEC2 result in severe intellectual disability, drug resistant seizures and autism. Our laboratory has been focusing on treating IQSEC2 disease using gene therapy with the adeno-associated viral (AAV) platform as this platform is currently the only FDA approved means for treating single gene disorders (Ling et al., 2023). However, due to the maximum size constraints of the genetic cargo which can be carried by AAV (4.2 kb for the open reading frame for the gene of interest) it is necessary to reduce the size of the 1488 amino acid sequence of IQSEC2 by at least 150 amino acids (Wu et al., 2010). In order to engineer the IQSEC2 protein to create a functional mini IQSEC2 protein it is necessary to understand the folding of IQSEC2 and how this folding affects its function and whether specific deletions in IQSEC2 allow retention of the structure and function of IQSEC2. IQSEC2 is an intrinsically disordered protein (IDP) (Shokhen et al., 2023). The folding and structure of IQSEC2 is poorly understood. There is no X ray crystallographic structure for the IQSEC2 molecule. The most popular and widely used method to model protein structure AlphaFold (Jumper et al., 2021) failed to generate a 3D structure of IQSEC2 consistent with the biology in experimental analysis (Shokhen et al., 2023) in line with the previously described failure of AlphaFold to predict the 3D structure of IDPs (Azzaz et al., 2022; Ruff and Pappu, 2021). IDPs exist as ensembles of conformers without a well-defined

* Corresponding author.
** Corresponding author.
E-mail addresses: michael.shokhen@biu.ac.il (M. Shokhen), alevy@technion.ac.il (A.P. Levy).

https://doi.org/10.1016/j.crstbi.2024.100158
Received 1 August 2024; Received in revised form 9 September 2024; Accepted 1 October 2024
Available online 3 October 2024

equilibrium structure and play diverse roles in cell signaling (Bondos et al., 2022; Tesei et al., 2024). The folding energy landscapes of ordered globular proteins is considered a rugged funnel-shaped pathway leading to one global potential minimum in the funnel bottom (Socci et al., 1998). On the other hand, the folding energy landscape of IDPs is shallower than in the case of ordered globular proteins with multiple minima corresponding to different low-energy conformational states (Uversky, 2017; Strodel, 2021). The dominating paradigm in the literature proposes that folding of IDPs occurs upon binding to another protein. Hence, contacts with this IDP binding partner are the major driving force for folding (Strodel, 2021; Habchi et al., 2014; Toto et al., 2020; Malagrinò et al., 2022). Two variants of how IDP folding occurs have been identified by NMR relaxation dispersion experiments: (1) "induced fit" – folding after association with the target protein, and (2) "conformational selection" - binding with the target protein in a pre-folded state (Araia et al., 2015).

Conformational heterogeneity of IDPs or intrinsically disordered regions (IDR) of proteins makes it virtually impossible to identify their full 3D structures in X-ray crystallography (Piovesan et al., 2022). Combining AMBER20 conventional molecular dynamics (Case et al., 2022), the deep convolutional residual neural networks (ResNet) method for predicting protein 3D structure implemented in the RaptorX server (Xu et al., 2021), and molecular modeling by YASARA Structure software (Kreiger et al., 2015; Kreiger et al., 2014), we have recently reported on the structure of IQSEC2 in an extended conformation. This study allowed us to analyze and explain how mutations in the IQ region of IQSEC2 mediate ARF6 dysregulation (Shokhen et al., 2023). However, a major limitation of deep learning algorithms such as RaptorX is that they only provide a static structure prediction in contradiction to the nature of IDPs which exist as an ensemble of conformers. In order to overcome this limitation a variety of enhanced sampling all-atom molecular dynamics algorithms (Bernardi et al., 2015; Spiwok et al., 2015; Chipot, 2023; Ray and Parrinello, 2023) have been proposed for the energy landscape analysis (Chipot, 2023), conformational search and protein folding simulations of IDPs (Kasahara et al., 2019; Saikia and Baruah, 2024; Shrestha et al., 2019; Willea et al., 2019; Shrestha et al., 2021; Chávez-García et al., 2022; Gurumoorthy et al., 2023; Zhu et al., 2023).

In the present study we have simulated the folding process of IQSEC2 by accelerated molecular dynamics (aMD) allowing us to identify multiple conformers of IQSEC2 and to determine how these conformers differ in their ability to interact with ARF6 and thereby influence GEF catalysis on ARF6. The aMD algorithm adds a non-negative ΔV boost potential to the system potential V(r) when it is below the reference energy E and thus reduces the time of simulating biomolecular conformational transitions through reduced energy barriers by orders of magnitude compared to classical MD (Hamelberg et al., 2004, 2007; Pierce et al., 2012; Miao et al., 2014, 2015). aMD and its algorithm enhancements are widely used for biomolecule conformational sampling (Mandal et al., 2023; Glaser et al., 2021), protein folding and unfolding (Zhou et al., 2019; Tyagi et al., 2019; Ermakova et al., 2022) protein-ligand and protein-protein binding (Smith and Carlson, 2021; Zhao et al., 2023; Bhattarai et al., 2021; Martí et al., 2022), analysis and calculation of the free energy landscapes, and potentials of mean force (PMF) of all these processes (Gedeon et al., 2015; Wang et al., 2021; Pawnikar et al., 2022). aMD does not require any predefined reaction coordinate (Miao et al., 2015). In the post-processing analysis we used the ensemble method of Gibbs (Hill, 1986) in combination with Gaussian peaks deconvolution in Origin(Pro) software (Origin (Pro); https:www.originlab.com) in order to identify snapshot population values of every conformational state (macrostate) of IQSEC2 in the folding process without the need of prior knowledge of V(r) for every snapshot. We then used statistical thermodynamics to calculate the free energy landscape and potential of mean force of the folding process allowing us to identify the different conformers of IQSEC2 and how these conformers differ in their interaction with ARF6 and thereby influence GEF catalysis. Finally, we show in agreement with prior experimental data (Myers et al., 2012) a key role for IQSEC2 dimerization as a driving force for its GEF catalysis for ARF6.

2. Methods

2.1. Folding of IQSEC2 by aMD simulations

We used accelerated molecular dynamics (aMD) implemented in AMBER20 software (Case et al., 2022) to simulate folding of an IQSEC2 fragment containing amino acid residues 1–1094 of an extended conformation previously generated by molecular modeling (Shokhen et al., 2023). The C-end Glu1094 was capped with an NMe (NHCH3 amide C-terminal protein capping) group which is customarily done in MD stimulations of proteins. The simulations were conducted using the ff14SB forcefield (Maier et al., 2015) in a rectangular periodic simulation box (140.537 Å, 161.795 Å, 166.377 Å) with explicit 116110 water molecules in a TIP3P model. The system was neutralized by addition of 316 Na+ and 309 Cl-ions in physiological concentration. The total number of atoms in the simulation cell was 365846. After preparing the input topology and structural files, the following steps were done before performing the production molecular dynamic simulations: (1) minimizing only the water, restraining the protein (20000 cycles); (2) short simulation to let water move (NPT, 310K), restraining the protein; (3) total minimization of water and protein (20000 cycles); (4) molecular dynamics of 1.4 ns to heat the system, restraining the protein (NVT, from 0 to 310K); (5) relaxing the system, restraining the protein heavy atoms (NPT, 310K, 1 ns); and (6) relaxing the system (NPT, 310K, 5 ns). In the equilibration stage, a temperature of 310 K and stable density were reached. For the actual simulations evaluating the complex interactions at the final (7) production stage, molecular dynamics simulations were performed at 310 K with a Langevin thermostat and NPT ensemble, using a pressure of 1 atm and a SHAKE constraint of 2 fs. An 8 Å cutoff radius was used for range-limited interactions, with Particle Mesh Ewald electrostatics for long-range interactions (Darden et al., 1993). The minimum distance from the protein atoms to the wall of the solvent box was determined to be 12 Å. The production molecular dynamics simulation was a set of sequential 100 ns steps. Every step was restarted from the previous one, applying a random seeds generator. The total production time was 4 μs. The system reached convergence in the last 800 ns, as estimated by the RMSD of the Cα atoms, Fig. S1. The snapshots were saved every 10 ps. An average dihedral energy and total potential energy were accepted from 5 ns of cMD equilibration stage (step 6) and used as a boost reference for the subsequent aMD production simulation. The aMD dual-boosting modification of the potential (Hamelberg et al., 2007) was defined by the following equations:

$$V(r)^* = V(r) + \Delta V(r) \tag{1}$$

$$\Delta V(r) = \left(E_p - V(r)\right)^2 / \left(\alpha P + E_p - V(r)\right) + \left(E_d - V_d(r)\right)^2 / \left(\alpha D + E_d - V_d(r)\right) \tag{2}$$

where $V(r)$ is the normal potential and $V_d(r)$ is the normal torsional potential. E_p and E_d are average potential and dihedral energies that serve as a reference energy. The terms αP and αD are factors that determine inversely the strength with which the boost is applied. The algorithm incorporating E_p, E_d, αP and αD was acquired from Pierce et al. (2012). A feature of aMD is that the shape of the modified potential conserves the underlying shape of the $V(r)$ real one, such that minima are maintained as minima and barriers are preserved as barriers (Hamelberg et al., 2004). Thus, adding the aMD potential simply modifies the relationship between energy differences, so the distribution of sampling of different structures is still related to the original potential distribution.

2.2. Reaction coordinate definition

The folding of big proteins is the result of large-scale motions of protein tertiary-structure domains. Such conformational changes take place on the nanosecond-microsecond time scale (Haran and Mazal, 2020). At the microscopic level, folding of a large protein is a rare event that occurs in a process dominated by the slowest degrees of freedom (Chipot, 2023; Chen and Chipot, 2022). For this reason the orbit in configuration space of the process of protein folding is projected on a one dimensional reaction coordinate corresponding to the slowest degree of freedom which is the largest conformational movement of protein domains in the process.

We study in this work the folding process of IDP protein IQSEC2 by aMD simulations. The structure generated by molecular modeling of amino acid residues 1–1094 of IQSEC2 has an extended conformation (Shokhen et al., 2023). aMD does not require any predefined reaction coordinate unlike other biased free energy calculation methods (Miao et al., 2015). Thus, aMD can be used for exploring the conformational space of biomolecules without a prior knowledge or restraints for calculations of the free energy landscape in post-simulation analysis (Miao et al., 2015). . The intrinsically disordered regions (IDR) limit the rate of the folding process, and of course the largest Ca-Ca distance identified between the initial extended and final folded states on the production MD trajectory is the optimal variant for conformational space sampling in post processing analysis. Specifically, the reaction coordinate for IQSEC2 folding was defined as the maximum distance in all pairs of Cα atoms between two conformations – the initial extended and final folded state:

$$R_c = \max\left[D_{ext}(C\alpha_k - C\alpha_l) - D_{fold}(C\alpha_k - C\alpha_l)\right] \quad (3)$$

As a result, the R_c was identified as a distance between the Cα atoms of Gly167 and Pro410 of IQSEC2.

2.3. The ensemble method of Gibbs

We present here an approach based on the statistical thermodynamics ensemble method of Gibbs. Ensemble refers to the collection of all possible microstates induced by thermal conformational fluctuations of the target macrosystem (Hill, 1986; Cooper, 1984). The microstates of the ensemble are distributed uniformly (Hill, 1986). It is well established that any protein system in thermal equilibrium with the environment fluctuates between conformational states (Cooper, 1984; Wei et al., 2016). The overall set of such states is characterized as the conformational landscape of the protein system. Based on the ensemble paradigm of statistical thermodynamics, the conformational landscape may be considered as an ensemble.

2.4. Histogram analysis of sampled MD snapshots

In aMD simulations of biomolecular systems the probability distributions of each snapshot (microstate) along a selected reaction coordinate R_c is calculated by Boltzmann's law $p^*(r) = e^{-\beta(V(r)+\Delta V(r))}$. $V(r)$ is the original potential, $\Delta V(r)$ is the boost potential, and r denotes the atomic coordinates. In post-processing analysis $p^*(A)$ can be reweighted to recover the canonical ensemble distribution, $p(A)$ (Miao et al., 2015). In this study we do not use a reweighting procedure. To calculate $p(r)$ we used a feature of aMD for which the shape of the modified potential in aMD conserves the underlying shape of the actual $V(r)$, such that minima are maintained as minima and barriers are preserved as barriers (Hamelberg et al., 2004). Thus, the probability distribution of snapshots is still related to the original potential distribution (Hamelberg et al., 2004; Miao et al., 2015). We accepted an approximation that the number of microstates in the ensemble replicating a protein macroscopic system is the total number of MD snapshots sampled on the production MD trajectory. In the language of probability theory (Feller, 2018), the sampled MD simulated snapshots may be considered as points of discrete sample space of size Ω of mutually exclusive independent random events. As follows from the second postulate of statistical thermodynamics, all Ω snapshots (microstates) have equal distribution probability (Hill, 1986):

$$q = 1/\Omega \quad (4)$$

Suppose that a thermodynamically equilibrated molecular system exists in dynamic equilibrium between K stationary conformational macrostates i. The probability p_i of every i-th conformational state occurrence may be estimated as a summarized probability of L_i snapshots composing this i-th state, where q is defined by eq (4):

$$p_i = qL_i \quad (5)$$

Obviously:

$$\sum_i^K L_i = \Omega \quad (6)$$

$$\sum_i^K p_i = 1 \quad (7)$$

We have applied a histogram technique to divide the entire range of the R_c variable into a series of consecutive, non-overlapping bins of 0.1 Å length for the collected snapshots. The R_c decreases from 124 Å in the stretched state to 26 Å in the folded state. The total number of bins is 981, which corresponds to $M = 400000$ collected snapshots. The method counts how many snapshots fall into each bin at the corresponding position on the R_c reaction coordinate. The MD snapshot distribution values, normalized by M number, give the values of the probability distribution p_k of every k-th bin, as follows from eqs (5)–(7). A graph of the snapshot distribution function $F(R_c)$ is presented in Fig. 1A.

2.5. Conformational analysis

Biomolecules in thermal contact with the environment exist, at the molecular level, as fluctuating dynamic systems (Cooper, 1984; Wei et al., 2016). The system fluctuates around each local minimum with amplitude values inverse to the depth of the potential well. In other words, the deeper the local minimum, the more time the system spends in this state and, therefore, the closer it is to local thermodynamic equilibrium in this macroscopic (conformational) state. All physical quantities describing the system in equilibrium fluctuate about their mean values (Hill, 1986; Cooper et al., 1984; Landau and Lifshitz, 1980). These fluctuations are small and are Gaussian in form (Hill, 1986; Landau and Lifshitz, 1980). Therefore, fluctuations of a quantity x characterizing the system (R_c in our case) of the i-th conformational state in thermodynamic equilibrium may be estimated by the standard deviation (STD) σ from its mean value μ: (Hill, 1986; Landau and Lifshitz, 1980).

$$\sigma = \sqrt{\frac{\sum (\boldsymbol{x} - \mu)^2}{L_i - 1}} \quad (8)$$

Where L_i is the number of snapshots composing the i-th conformational state. Indeed, a snapshots distribution plot in Fig. 1A demonstrates that the peaks corresponding to the most stable conformational states have a well-defined Gaussian shape. Most peaks in Fig. 1A are overlapping, so we performed a Gaussian peaks deconvolution using the method implemented in the Origin (Pro) software (Origin (Pro) https:www.originlab.com). Starting from $R_c > 91$ Å, the snapshot population was too small, so a joint peaks deconvolution analysis in the total area of R_c variation was not possible on the scale used. Thus, we divided the region of R_c variation into two intervals [26 Å $\leq R_c \leq$ 91 Å] and [91 Å $\leq R_c \leq$ 124 Å]. The results of the peak deconvolution analysis, which identified 7 folded state (FS) peaks (FS1, …,FS7) in the first R_c interval and 4

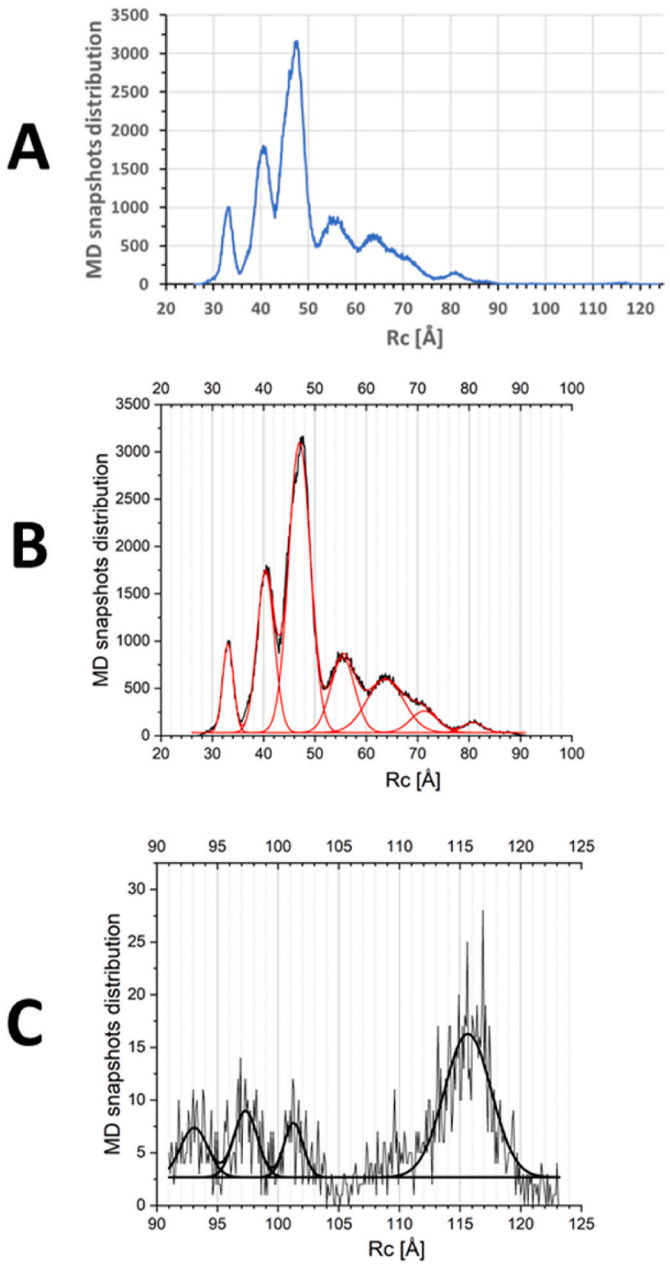


Fig. 1. MD snapshot distribution as a function of the IQSEC2 folding along the R_c reaction coordinate. A. Graph of histogram analysis of aMD simulations of the folding of IQSEC2. **B, C.** Gaussian peaks deconvolution.

unfolded state (US) peaks (US1, …,US4) in the second R_c interval, are plotted in Fig. 1B and C. The corresponding Origin (Pro) statistical analysis is presented in Tables S1 and S2. The correlation coefficients R^2 when approximating snapshot distributions with Gaussian peaks are equal to 0.993 and 0.657 at R_c intervals with high and low snapshot populations, respectively. The latter option is well illustrated by the plot in Fig. 1C with a high amplitude of fluctuations in the distribution of snapshots of sparsely populated shallow minima. We adopted a Gaussian approximation for the analysis of the distribution of MD snapshots at each identified peak over the entire range of R_c variations. The peak boundaries on R_c were estimated by the standard deviations (STD) σ from its mean value μ, eq (8), as presented in Tables S3 and S4. Gaussian peaks occupied 11 fragments on R_c, determined by the corresponding pairs of parameters (μ, σ). There were also 12 interval states (IS1,..,IS12) - 10 between the Gaussian peaks on R_c and 2 areas on the left and right boundaries. First, we determined a pair of parameters (μ, σ) for every 12 intervals based on their snapshot populations and boundaries in R_c, and then we carried out their statistical analysis as presented in Table S5.

2.6. Conformational free energies from first principles

To predict the free energy and entropy of different conformational states by statistical thermodynamics, it is necessary to estimate the density of states and count how many MD snapshots (ensemble microstates) L_i composes each *i-th* conformational state (macrostate). We considered that the system was in an equilibrated stationary conformational state for every local minimum. The probability p_i of independent mutually exclusive events *i* is defined by eq (9) (Feller, 2018):

$$p_i = L_i / \sum_i L_i \tag{9}$$

It should be noted that values for L_i only take into account MD snapshots occupying the interval (μ-σ, μ+σ) on R_c for each macrostate (conformational state) as determined in the statistical analysis in the previous section, Tables S3–S5. Accordingly, with population values L_i and the corresponding probability occurrences of conformational states in hand, it was possible to predict their relative energetic stability by means of statistical thermodynamics. Since kinetic experiments in enzymology are usually performed at normal atmospheric pressure, we conducted the MD simulations in frames of (*N, P, T*) ensemble at constant pressure $P = 1$ bar $= 100$ kJ/M^3, T = 310K, and *N* total number of atoms in the system. Hence, we used a Gibbs free energy G_i for estimation of a *i*-th conformational state stability:

$$G_i = E_i + PV_i - TS_i \tag{10}$$

Where E_i is the internal energy of a thermodynamic system (the sum of kinetic and potential energies), P, V_i, T, and S_i are pressure, volume, temperature, and entropy of the system, respectively. P and T are constant parameters in the (*N,P,T*) ensemble. We used basic formulas of statistical thermodynamics for the definitions of S and E adopted for this study. The probabilities p_i of macrostates defined in eq (9) may be equivalently expressed as the Boltzmann distribution law:

$$p_i = e^{-E_i/RT}/Q \tag{11}$$

$$Q = \sum_i e^{-E_i/RT} \tag{12}$$

Where T is the absolute temperature, and Q is the conformational partition function. The entropy S_i of an *i*-th conformational state was defined by eq (13):

$$S_i = -R\, p_i\, Ln\, p_i \tag{13}$$

where R is the Gas constant. Note that our calculations of thermodynamic functions in this work were conducted in molar units, so in eqs (11)–(13) the Boltzmann constant k was substituted for the Gas constant R. The internal energy E_i of a conformational state *i* according to eq (11) is defined by eq (14):

$$E_i = -RT\, Ln\, p_i - RT\, LnQ \tag{14}$$

Experiment can give only the difference of ΔE_{ij} or ΔS_{ij} values between two conformational states *i* and *j*, but not absolute values (Hill, 1986), so we estimated the stability of a stationary conformational state *i* as the ΔG_{i0} difference calculated relative to the most stable state *0*:

$$\Delta G_{i0} = \Delta E_{i0} + P\Delta V_{i0} - T\Delta S_{i0} \tag{15}$$

where P, T are constants of the (*N, P, T*) ensemble. The $RT\ LnQ$ is eliminated in the calculation of ΔE_{i0} in eq (15) since Q is constant according to eqs (9) and (11).

We calculated the values of the protein system volume enclosed in the solvent accessible surface (SAS) for all 11 Gaussian and 12 interval states in their mean 3D structures identified in previous statistical

analyses. All p_i-s of macrostates were calculated by eq (9). The most stable local equilibrium state is in the conformer corresponding to the FS3 Gaussian peak with $\mu = 47.0$ Å on R_c, so all relative free energy values were calculated using FS3 as the zero level, where $\Delta G_{i0} = G_i - G_0$. The detailed calculations of the thermodynamic functions by eq (9),13–15 are presented in the *Thermodynamics.xlsx* file in SI.

2.7. Computational assessment of how IQSEC2 forms dimers and how IQSEC2 interacts with ARF6

IQSEC2 dimers derived from the most stable folded (FS3) and unfolded (US4) conformers of IQSEC2 were generated by the docking server ClusPro, which implements rigid body methods and fast Fourier transform (FFT) algorithms allowing for the sampling of billions of complex conformations (Kozakov et al., 2017). The N-terminal amino acid residues 1–67 of IQSEC2 were used as attraction parameters in the protein-protein docking for the formation of these dimers as these residues have a canonical coiled-coil domain sequence which has previously been shown to mediate the dimerization of many proteins (Burkhard et al., 2001) and have been proposed to mediate the dimerization of IQSEC2 (Myers et al., 2012). We previously reported on the molecular modeling of the ARF6-Sec7-PH complex and equilibrated it in aMD simulations (Shokhen et al., 2023). In this study using YASARA Structure software, we aligned Sec7-PH domains of the ARF6-Sec7-PH complex with IQSEC2 (residues 1–1094 residues) for the folded FS3 and unfolded US4 IQSEC2 conformers. After deleting Sec7-PH fragment (amino acids 749–1094) of the ARF6-Sec7-PH complex the geometry of the IQSEC2 dimers and IQSEC2 ARF6 complexes was optimized by YASARA Structure software using the ff14SB forcefield in a periodic simulation cell filled with explicit water molecules and Na+ and Cl-ions in physiological concentrations. The ΔG free energies of binding at 37 °C of the IQSEC2 dimers and IQSEC2-ARF6 encounter complexes were estimated by the PRODIGY web server an online tool for the prediction of binding affinity in protein–protein complexes (Xue et al., 2016; Vangone and Bonvin, 2015).

3. Results

3.1. Use of potential of mean force analysis to describe the folding of IQSEC2

As described in methods we first sought to determine how IQSEC2 folds. The calculated ΔG_{i0} values for the folding of IQSEC2 are presented in Fig. 2 as a graph of $\Delta G_{ij}(R_c)$ Potential of Mean Force (PMF). Table 1 summarizes the details characterizing the PMF graph with a detailed thermodynamics analysis presented in the *Thermodynamics.xlsx* file in SI. There are 7 minima in the folded state (FS) and 2 in the unfolded state (US). The FS3 conformation represents the most stable of the conformational states of the folded protein and the US4 conformation with $\mu =$ 115.6 Å on R_c and $\Delta G_{i0} =$ 15.5 kJ/mol is the most stable of the two conformational states of the unfolded protein. The upper potential barrier with a height of $\Delta G_{j0}^{\neq} =$ 22.6 kJ/mol and $\mu =$ 99.4 Å on R_c corresponds to the IS9 region and is the rate limiting transition state (TS) for the folding-unfolding equilibrium, as presented in Fig. 3. The folding-unfolding rate limiting constants k_f and k_u can be calculated by the classical kinetic equation (Behar et al., 2017):

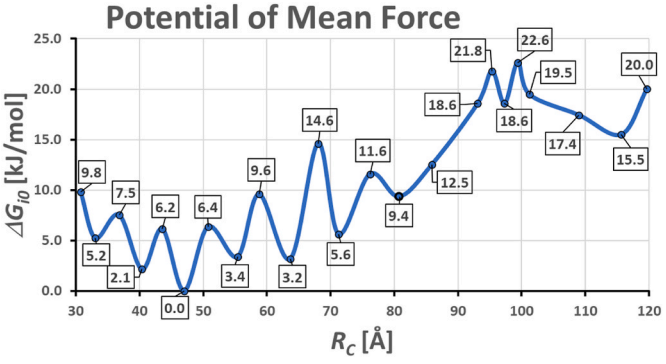


Fig. 2. Potential of mean force graph of the IQSEC2 folding along the R_c reaction coordinate. Labels present the ΔG_{i0} values calculated relative to the most stable FS3 folded conformer, for details see Table 1 and *Thermodynamics.xlsx* file in SI.

Table 1
IQSEC2 conformational macrostates on the R_c reaction coordinate. The $\mu(R_c)$ are mean values in $(\mu - \sigma, \mu + \sigma)$ intervals on R_c corresponding to the number of snapshots of each macroscopic state of IQSEC2 during the folding process. Potential minimum states are shown in **bold**.

Conformational states	Snapshots count L_i	$\mu(R_c)$ Å	ΔG_{i0} kJ/mol
IS1	3706	30.77	9.8
FS1	**19771**	**33.06**	**5.2**
IS2	8575	36.80	7.5
FS2	**58247**	**40.38**	**2.1**
IS3	15304	43.57	6.2
FS3	**126557**	**46.99**	**0.0**
IS4	13675	50.76	6.4
FS4	**41230**	**55.39**	**3.4**
IS5	4333	58.78	9.6
FS5	**44383**	**63.68**	**3.2**
IS6	691	68.05	14.6
FS6	**18439**	**71.25**	**5.6**
IS7	2225	76.19	11.6
FS7	**5569**	**80.76**	**9.4**
IS8	1750	85.90	12.5
US1	172	93.08	18.6
IS9	50	95.35	21.8
US2	**179**	**97.31**	**18.6**
IS10	38	99.37	22.6
US3	127	101.26	19.5
IS11	281	108.99	17.4
US4	**617**	**115.65**	**15.5**
IS12	108	119.67	20.0
Total sum	**366027**		

$$k_i = k_{0f}\, e^{-\Delta G_i^{\neq}/RT} \tag{16}$$

Where $i = f$ or u, and $k_{0f} = (1\mu s)^{-1}$ is the upper limit of the diffusion-controlled rate of protein folding (Hagen et al., 1996). As a result, for the rate limiting steps $k_f = 6.36 \times 10^4$ sec^{-1} and $k_u = 1.56 \times 10^2$ sec^{-1}, the corresponding transmission path times are $t_f = 1.57 \times 10^{-5}$ s and $t_u = 6.43 \times 10^{-3}$ s. The predicted folding of IQSEC2 in ~16 μs correlates with the previously described rate of closing of a long loop in a chemically unfolded cytochrome which occurred in 35–40 μs (Hagen et al., 1996).

3.2. Determination of which IQSEC2 conformers can form an encounter complex with ARF6

We wished to determine which of the IQSEC2 conformers would be predicted to be catalytically active. We examined all of the folded and unfolded IQSEC2 conformers for their ability to form encounter complexes with ARF6. Both the IQSEC2 US2 and US4 unfolded states can form complexes with ARF6. Among the folded conformers, only FS1, FS4, and FS7 can bind ARF6 because there are unresolvable conformational clashes between the ARF6 loop (amino acids 37–50 of ARF6) and the PH domain loop (IQSEC2 amino acids 953–972) of the FS2, FS3, FS5 and FS6 IQSEC2 conformers. Moreover, FS3 and FS6 have unresolvable conformational clashes with the α-helical fragment 878–895 of the IQSEC2 catalytic Sec7 domain. Fig. 4 shows the computational 3D structures of such unrealized complexes of ARF6 with conformers FS2 and FS3. The relative concentrations of active (IQSEC2 capable of forming a complex with ARF6) and inactive folded conformers of

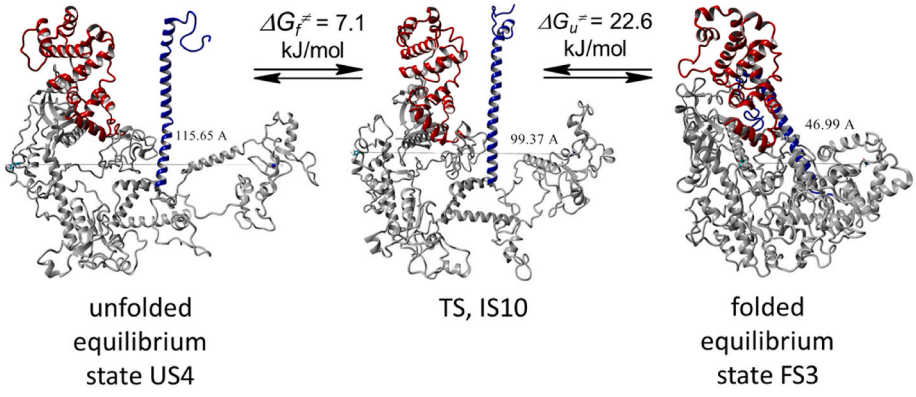


Fig. 3. 3D structures of the most stable folded (FS3) and unfolded (US4) IQSEC2 conformers, and the IS10 macrostate which is the rate limiting transition state (TS) of the folding-unfolding equilibrium. The numerical values are distances in Å between the Cα atoms of Gly167 and Pro410 of IQSEC2. The latter was accepted as the R_c reaction coordinate of folding. Colors: N-terminus (1–67)-blue; Sec7 domain (746–939)-red; the remaining IQSEC2 fragments – grey.

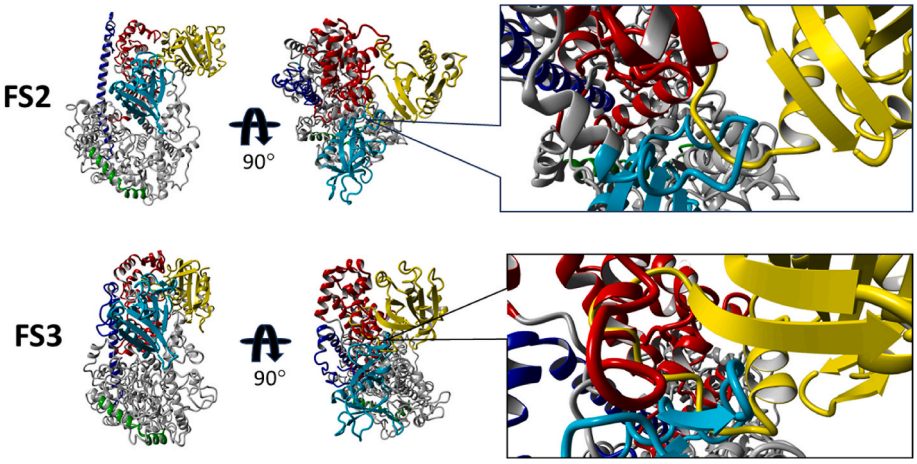


Fig. 4. The majority of folded conformers of IQSEC2 cannot bind ARF6. The3D structures of ARF6-IQSEC2 encounter complexes which are predicted not to occur due to conformational clashes between linked loops of ARF6 and the PH domain in the FS2 and FS3 IQSEC2 conformers. FS3 also has an unresolvable conformational clash with an alpha-helix within Sec7 domain. Colors: N-terminus (1–67)-blue; IQ domain (347–376) - green; Sec7 (746–939) - red; PH (951–1085) - cyan; the remaining IQSEC2 fragments – grey; ARF6-yellow.

IQSEC2 can be calculated from the corresponding snapshots population in Table 1. There are 20% active and 80% inactive folded conformers. Among all folded conformers of IQSEC2, the most stable FS3 dominates with a population of 40.3%. Free energies of ARF6 binding with the most stable unfolded conformer US4, the US4-US4 dimer, and Sec7-PH domain (749–1094) were calculated by the PRODIGY server. The contribution to the intermolecular bond energy of electrostatic interactions between any contacting pair of oppositely charged amino acid side chains and especially ionic hydrogen bonds at the interface was an order of magnitude greater than the contribution of any pair of van der Waals interactions. Table 2 presents N – O interatomic distances between functional groups of oppositely charged amino acid side chains in the contact surface of ARF6 encounter complexes with either monomeric US4, the US4-US4 dimer or SEC7-PH domain, as well as the free energies of binding of these complexes predicted by PRODIGY. These data predict that the US4 conformer should be able to interact well with ARF6, and to be catalytically active, whether it is in a monomeric or dimeric form.

Table 2
Encounter complexes of ARF6 with IQSEC2. The encounter complex of ARF6 with monomeric US4 IQSEC2 and with the US4-US4 IQSEC2 dimer is compared to the complex of ARF6 with Sec7-PH (749–1094) IQSEC2 fragment. Ionic hydrogen bond distances are shown in bold. Free energies of binding ΔG are predicted by the PRODIGY server.

Interactions of oppositely charged ionic side chains of amino acid residues		Distances [Å] between N and O atoms of functional groups		
IQSEC2	ARF6	SEC7-PH-ARF6	US4-ARF6	US4:US4 dimer-ARF6
Asp 441	Lys 32		4.88	7.33
Glu 849	Lys 26	5.38	6.04	7.50
Lys 852	Asp 22	4.53	6.40	6.76
G 854	Lys 69	10.19	8.24	8.44
Asp 879	Lys 69	**2.71**	**2.74**	**2.70**
Asp 894	Arg 15	4.44	5.27	6.20
Lys 913	Glu 111	6.38	3.96	4.77
Asp 919	Lys 69	10.15	7.92	10.76
ΔG [kJ/mol]		−63.1	−70.2	−64.0

3.3. *Computational prediction of the structure of IQSEC2 dimers and their ability to interact with ARF6*

Meyers et al. demonstrated experimentally that IQSEC2 predominately exists as a dimer in its active form and proposed that dimerization \was mediated via its N-terminal coiled-coil domain (IQSEC2 amino acid residues 1–67 (Myers et al., 2012). The 3D structures of the IQSEC2 monomer and dimer have not been described. Based on the 3D structures of the folded FS3 and unfolded US4 conformers of IQSEC2 we

generated their dimers FS3-FS3 and US4-US4 by means of the ClusPro protein-protein docking method as described in Methods (Kozakov et al., 2017). The folded FS3 dimer has a weaker stability as predicted by the PRODIGY server (Xue et al., 2016; Vangone and Bonvin, 2015) and has a free energy of −50.2 kJ/mol as compared to −58.5 kJ/mol for the unfolded US4 dimer, which is due to the much smaller contact surface of 673.4 $Å^2$ between the N-termini in the FS3 dimer compared to the N-termini in the US4 dimer with a contact surface value of 1289.2. $Å^2$ (Fig. 5). The N-termini are in close contact along their axes in the US4-US4 dimer in contrast to the FS3-FS3 dimer. The interactions of amino acid pairs mediating the contact of both FS3-FS3 and US4-US4 dimers are presented in the files FS3-FS3 interactions.xlsx and US4-US4 interactions.xlsx in SI. The free energy of binding of the US4 dimer is contributed by 7 pairs of interactions of oppositely charged ionic amino acid residues, in contrast to the FS3 dimer, where there is not a single pair of ionic interactions.

4. Discussion

We present here an original algorithm for protein conformational analysis using principles of statistical thermodynamics based on the ensemble method of Gibbs. Since it is a general approach, it could be applied as a post-processing method with molecular dynamics simulations for any biomolecule. Here we have applied the method to an intrinsically disordered protein IQSEC2 and identify multiple potential conformers which are predicted to have different biological activities. Specifically, we have shown that IQSEC2 can form a complex with ARF6 only in an unfolded state and that it can also dimerize. We suggest that competition between the folding and dimerization of IQSEC2 is used by nature as a balance to quantitatively regulate the GDP exchange on ARF6 catalyzed by IQSEC2. The computational results presented here are consistent with experimentally determined results concerning how the GEF activity of IQSEC2 is regulated physiologically by Ca+2 in the neuron (Myers et al., 2012; Bai et al., 2023; Rogers et al., 2019).

Our modeling suggests that IQSEC2 can exist in multiple conformers both in a folded state and in an unfolded state and we have provided the predicted frequency distribution of these conformers. We have shown that about 80% of the folded conformers are predicted to be catalytically inactive while the unfolded conformers are all active as identified by the ability of these respective conformers to form encounter complexes with ARF6.

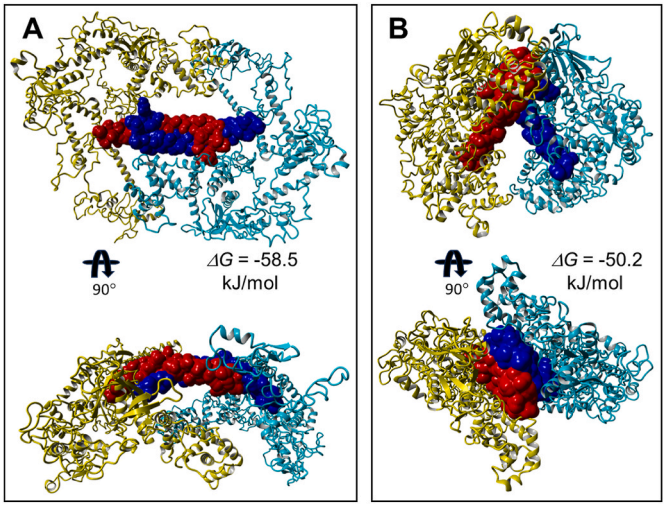


Fig. 5. Three-dimensional structures in two projections of (A) US4-US4 and (B) FS3-FS3 dimers of IQSEC2 predicted by the ClusPro protein-protein docking server. The binding free energies were estimated by the PRODIGY server. The R and L component monomers of the dimer are represented as a ribbon, while their N-terminal fragments are shown as vdW surfaces. Colors: molecule R - yellow, its N-terminal helical domain (1–67) - red; the corresponding colors of the L molecule are cyan and blue, respectively.

What regulates the transition from folded inactive conformers to unfolded active conformers of IQSEC2? Experimental studies from our lab and others have shown that in the setting of low intracellular Ca+2 in the neuron (<100 nm) apocalmodulin (apoCM) is bound to the IQ region of IQSEC2 and IQSEC2 is catalytically inactive (Bai et al., 2023; Rogers et al., 2019). A rise in intracellular Ca+2 due to the binding of the neurotransmitter glutamate to a receptor on the surface of the neuron results in the dissociation of apoCM from IQSEC2 and the activation of the GEF activity of IQSEC2 (Brown et al., 2016). We propose that the Ca2+ dependent binding of apoCM controls the activity of IQSEC2 by regulating the folding of IQSEC2. When apoCM is bound this promotes transition of IQSEC2 into a folded state to which ARF6 cannot bind and IQSEC2 is thus inactive catalytically. When Ca+2 levels in the neuron rise and apoCM dissociates from the folded IQSEC2, this promotes transition into an unfolded state thereby providing access of the SEC7 region of IQSEC2 to ARF6. Our modeling has also shown that IQSEC2 can form dimers predominately in the unfolded state. Myers has shown that IQSEC2 dimer formation, as mediated by its N terminal region, is required for the physiological Ca+2 induced increase in IQSEC2 activity (Myers et al., 2012). We have shown that the electrostatic attraction between oppositely charged ionic amino acid residues in the IQSEC2 N-terminal fragment dominates the stability of the US4-US4 dimer.

Study limitations

Although our results are well supported by experimental data and are consistent with the physiological regulation of IQSEC2 catalysis, further molecular modeling and experimental validation will be necessary in order to study the competing kinetics of the folding and dimerization of IQSEC2.

Medical significance and potential importance of this study

Understanding the regulation of IQSEC2 is critical for the development of new treatments for children with mutations in IQSEC2. Specifically the model described here is being incorporated into an algorithm to design and test IQSEC2 minigenes that may be used for gene therapy for children with IQSEC2 disease. We are currently attempting to select regions of IQSEC2 which may be deleted and yet preserve folding and dimerization. The studies described here will allow computational testing of residues which can be deleted and yet preserve folding, formation of ARF6 encounter complexes and dimerization. Once we have computationally designed an active minigene we can test its biological activity in neurons derived from patient induced pluripotent stem cells (Brant et al., 2021), as well as in mouse models of human IQSEC2 disease (Rogers et al., 2019; Brant et al., 2021; Jackson et al., 2019; Jabarin et al., 2021; Kane et al., 2022).

Author contributions

MS: Conceptualization, Methodololgy, Writing – original draft, Writing – review & editing. AA: Conceptualization, Resources, Writing – review & editing. YB: Conceptualization, Validation, Writing – review & editing. YI: Conceptualization, Validation, Writing – review & editing. NSL: Conceptualization, Validation, Writing – review & editing. APL: Conceptualization, Methodololgy, Validation, Resources, Writing – original draft, Writing – review & editing

Funding

This work was funded in part by a grant from the Israel Science Foundation (grant # 591/23) and the United States-Israel Binational Science Foundation (grant #2023012) both to APL.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Andrew P Levy reports financial support was provided by Israel Science Foundation. If there are other authors, they 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crstbi.2024.100158.

References

Araia, M., Sugasea, K., Dysona, H.J., Wright, P.E., 2015. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. Proc. Natl. Acad. Sci. U.S.A. 112 (31), 9614–9619. https://doi.org/10.1073/pnas.1512799112.

Azzaz, F., Yahi, N., Chahinian, H., Fantini, J., 2022. The epigenetic dimension of protein structure is an intrinsic weakness of the alphafold program. Biomolecules 12 (10), 527. https://doi.org/10.3390/biom12101527.

Bai, G., Li, Hao, Qin, P., Guo, Y., Yang, W., Lian, Y., Ye, F., Chen, J., Wu, M., Huang, R., Li, J., Lu, Y., Zhang, M., 2023. Ca+2 induced release of IQSEC2/BRAG1 autoinhibition under physiological and pathological conditions. J. Cell Biol. 222 (12), e202307117. https://doi.org/10.1083/jcb.202307177.

Bondos, S.E., Dunker, A.K., Uversky, V.N., 2022. Intrinsically disordered proteins play diverse roles in cell signaling. Cell Commun. Signal. 20, 20.

Behar, I., Jernigan, R.L., Dill, K.A., 2017. The principles of protein folding kinetics, Chapter 6, Protein Actions, pg. Garland Science 129–160. https://dasher.wustl.edu/bio5357/readings/behar-chapter6.pdf.

Bernardi, R.C., Melo, M.C.R., Schulten, K., 2015. Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochim. Biophys. Acta 1850 (5), 872–877. https://doi.org/10.1016/j.bbagen.2014.10.019.

Bhattarai, A., Pawnikar, S., Miao, Y., 2021. Mechanism of ligand recognition by human ACE2 receptor. J. Phys. Chem. Lett. 12 (20), 4814–4822. https://doi.org/10.1021/acs.jpclett.1c01064.

Brown, J.C., Petersen, A., Zhong, L., Himelright, M.L., Murphy, J.A., Walikonis, R.S., Gerges, N.Z., 2016. Bidirectional regulation of synaptic transmission by BRAG1/IQSEC2 and its requirement in long-term depression. Nat. Commun. 7, 11080. https://doi.org/10.1038/ncomms11080.

Brant, B., Tchelet Stern, T., Shekhidem, H., Mizrahi, L., Rosh, I., Stern, Y., Ofer, P., Asleh, A., Umanah, G., Jada, R., Levy, N.S., Levy, A.P., Shani Stern, S., 2021. IQSEC2 mutation associated with epilepsy, intellectual disability and autism results in hyperexcitability of patient derived neurons and deficient synaptic transmission. Mol. Psychiatr. 26 (12), 7498–7508. https://doi.org/10.1038/s41380-021-01281-0.

Burkhard, P., Stetefeld, J., Strelkov, S.V., 2001. Coiled coils: a highly versatile protein folding motif. Trends Cell Biol. 11 (2), 82–88. https://doi.org/10.1016/s0962-8924(00)01898-5.

Case, D.A., Aktulga, H.M., Belfon, K., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., et al., 2022. Amber 2020. University of California, San Francisco.

Chávez-García, C., Hénin, J., Karttunen, M., 2022. Multiscale computational study of the conformation of the full-length intrinsically disordered protein MeCP2. J. Chem. Inf. Model. 62 (4), 958–970. https://doi.org/10.1021/acs.jcim.1c01354.

Chen, H., Chipot, C., 2022. Enhancing sampling with free-energy calculations. Curr. Opin. Struct. Biol. 77, 102497. https://doi.org/10.1016/j.sbi.2022.102497.

Chipot, C., 2023. Free energy methods for the description of molecular processes. Annu. Rev. Biophys. 52, 113–138. https://doi.org/10.1146/annurev-biophys-062722-093258.

Cooper, A., 1984. Protein fluctuations and the thermodynamic uncertainty principle. Prog. Biophys. Mol. Biol. 44 (3), 181–214. https://doi.org/10.1016/0079-6107(84)90008-7.

Darden, T., York, D., Pedersen, L., 1993. Particle Mesh Ewald: an nlog(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092. https://doi.org/10.1063/1.464397.

Ermakova, E., Makshakova, O., Zuev, Y., Sedov, I., 2022. Beta-rich intermediates in denaturation of lysozyme: accelerated molecular dynamics simulations. J. Biomol. Struct. Dyn. 40 (24), 13953–13964. https://doi.org/10.1080/07391102.2021.1997823.

Feller, W., 2018. An Introduction of Probability Theory and its Applications, I. Wiley, India Pvt. Ltd.

Gedeon, P.C., Thomas, J.R., Madura, J.D., 2015. Accelerated molecular dynamics and protein conformational change: a theoretical and practical guide using a membrane embedded model neurotransmitter transporter. Methods Mol. Biol. 1215, 253–287. https://doi.org/10.1007/978-1-4939-1465-4_12.

Glaser, M., Bruce, N.J., Han, S.B., Wade, R.C., 2021. Simulation of the positive inotropic peptide S100A1ct in aqueous environment by Gaussian accelerated molecular dynamics. J. Phys. Chem. B 125 (18), 4654–4666. https://doi.org/10.1021/acs.jpcb.1c00902.

Gurumoorthy, V., Shrestha, U.R., Zhang, Q., Pingali, S.V., Boder, E.T., Urban, V.S., Smith, J.C., Petridis, L., O'Neill, H., 2023. Disordered domain shifts the conformational ensemble of the folded regulatory domain of the multidomain oncoprotein c-Src. Biomacromolecules 24 (2), 714–723. https://doi.org/10.1021/acs.biomac.2c01158.

Habchi, J., Tompa, P., Longhi, S., Uversky, V.N., 2014. Introducing protein intrinsic disorder. Chem. Rev. 114 (13), 6561–6588. https://doi.org/10.1021/cr400514h.

Hagen, S.J., Hofrichter, J., Szabo, A., Eaton, W.A., 1996. Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. Proc. Natl. Acad. Sci. USA 93 (21), 11615–11617. https://doi.org/10.1073/pnas.93.21.11615.

Hamelberg, D., Mongan, J., McCammon, J.A., 2004. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J. Chem. Phys. 120 (24), 11919–11929. https://doi.org/10.1063/1.1755656.

Hamelberg, D., de Oliveira, C.A.F., McCammon, J.A., 2007. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. J. Chem. Phys. 127 (15), 155102. https://doi.org/10.1063/1.2789432.

Haran, G., Mazal, H., 2020. How fast are the motions of tertiary-structure elements in proteins? J. Chem. Phys. 153 (1), 130902. https://doi.org/10.1063/5.0024972.

Hill, T.L., 1986. An Introduction to Statistical Thermodynamics. Dover Publications; Inc., N.Y.

Jabarin, R., Levy, N., Abergel, Y., Berman, J.H., Zag, A., Netser, S., Levy, A.P., Wagner, S., 2021. Pharmacological modulation of AMPA receptors rescues specific impairments in social behavior associated with the A350V IQSEC2 mutation. Transl. Psychiatry 11 (1), 234. https://doi.org/10.1038/s41398-021-01347-1.

Jackson, M.R., Loring, K.E., Homan, C.C., Thai, M.H.N., Maattanen, L., Arvio, M., Jarvela, I., Shaw, M., Gardner, A., Gecz, J., Shoubridge, C., 2019. Heterozygous loss of function of *IQSEC2/Iqsec2* leads to increased activated Arf6 and severe neurocognitive seizure phenotype in females. Life Sci. Alliance 2 (4), e201900386. https://doi.org/10.26508/lsa.201900386.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zıdek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinksi, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596 (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Kane, O., McCoy, A., Jada, R., Borisov, V., Zag, L., Zag, A., Rozales, K., Shalgi, R., Levy, N.S., Levy, A.P., Marsh, E., 2022. Characterization of spontaneous seizures and EEG abnormalities in a mouse model of the human A350V IQSEC2 mutation and identification of a possible target for precision medicine based therapy. Epilepsy Res. 182, 106907. https://doi.org/10.1016/j.eplepsyres.2022.106907.

Kasahara, K., Terazawa, H., Takahashi, T., Higo, J., 2019. Studies on molecular dynamics of intrinsically disordered proteins and their fuzzy complexes: a mini-review. Comput. Struct. Biotechnol. J. 17, 712–720. https://doi.org/10.1016/j.csbj.2019.06.009.

Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D., Vajda, S., 2017. The ClusPro web server for protein-protein docking. Nat. Protoc. 12 (2), 255–278. https://doi.org/10.1038/nprot.2016.169.

Krieger, E., Vriend, G., 2015. New ways to boost molecular dynamics simulations. J. Comput. Chem. 36 (13), 996–1007. https://doi.org/10.1002/jcc.23899.

Krieger, E., Vriend, G., 2014. YASARA view - molecular graphics for all devices - from smartphones to workstations. Bioinformatics 30 (20), 2981–2982. https://doi.org/10.1093/bioinformatics/btu426.

Landau, L.D., Lifshitz, E.M., 1980. Course of Theoretical Physics. Volume 5. Statistical Physics, Part 1. Third revised and enlarged edition. Pergamon Press Ltd., N.Y.

Levy, N.S., Borisov, V., Lache, O., Levy, A.P., 2023. Molecular insights into IQSEC2 disease. Int. J. Mol. Sci. 24 (5), 4984. https://doi.org/10.3390/ijms24054984.

Ling, Q., Herstine, J.A., Bradbury, A., Gray, S.J., 2023. AAV-based in vivo gene therapy for neurological disorders. Nat. Rev. Drug Discov. 22 (10), 789–806. https://doi.org/10.1038/s41573-023-00766-7.

Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., Simmerling, C., 2015. ff14SB: improving the accuracy of protein sidechain and backbone parameters from ff99SB. J. Chem. Theor. Comput. 11 (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

Malagrinò, F., Diop, A., Pagano, L., Nardella, C., Toto, A., Gianni, S., 2022. Unveiling induced folding of intrinsically disordered proteins – protein engineering, frustration and emerging themes. Curr. Opin. Struct. Biol. 72, 153–160. https://doi.org/10.1016/j.sbi.2021.11.004.

Mandal, P., Rani, P., Chandra, G., Singh, D.V., 2023. Flap sub-domain dynamics of serine-threonine phosphatase (Stp1) of Staphylococcus aureus: an accelerated molecular dynamics simulation study. J. Biomol. Struct. Dyn. 41 (13), 6413–6421. https://doi.org/10.1080/07391102.2022.2107575.

Martí, D., Alemán, C., Ainsley, J., Ahumada, O., Torras, J., 2022. IgG1-b12−HIV-gp120 interface in solution: a computational study. J. Chem. Inf. Model. 62 (2), 359–371. https://doi.org/10.1021/acs.jcim.1c01143.

Miao, Y., Sinko, W., Pierce, L., Bucher, D., Walker, R.C., McCammon, J.A., 2014. Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. J. Chem. Theor. Comput. 10 (7), 2677–2689. https://doi.org/10.1021/ct500090q.

Miao, Y., Feixas, F., Eun, C., McCammon, J.A., 2015. Accelerated molecular dynamics simulations of protein folding. J. Comput. Chem. 36 (20), 1536–1549. https://doi.org/10.1002/jcc.23964.

Myers, K.R., Wang, G., Sheng, Y., Conger, K.K., Casanova, J.E., Zhu, J.J., 2012. Arf6-GEF BRAG1 Regulates JNK-mediated synaptic removal of GluA1-containing AMPA receptors: a new mechanism for nonsyndromic X-linked mental disorder. J. Neurosci. 32 (34), 11716–11726. https://doi.org/10.1523/JNEUROSCI.1942-12.2012.

Origin(Pro), Version 23. OriginLab Corporation, Northampton, MA, USA. https://www.originlab.com/.

Pawnikar, S., Bhattarai, A., Wang, J., Miao, Y., 2022. Binding analysis using accelerated molecular dynamics simulations and future perspectives. Adv. Appl. Bioinform. Chem. 15, 1–19. https://doi.org/10.2147/AABC.S247950.

Pierce, L.C.T., Salomon-Ferrer, R., de Oliveira, C.A.F., McCammon, J.A., Walker, R.C., 2012. Routine access to millisecond time scale events with accelerated molecular dynamics. J. Chem. Theor. Comput. 8 (9), 2997–3002. https://doi.org/10.1021/ct300284c.

Piovesan, D., Monzon, A.M., Quagliaa, F., Tosatto, S.C.E., 2022. Databases for intrinsically disordered proteins. Acta Crystallogr. D78, 144–151. https://doi.org/10.1107/S2059798321012109.

Ray, D., Parrinello, M., 2023. Kinetics from metadynamics: principles, applications, and outlook. J. Chem. Theor. Comput. 19 (17), 5649–5670. https://doi.org/10.1021/acs.jctc.3c00660.

Rogers, E.J., Jada, R., Schragenheim-Rozales, K., Sah, M., Cortes, M., Florence, M., Levy, N.S., Moss, R., Walikonis, R.S., Palty, R., Shalgi, R., Lichtman, D., Kavushansky, A., Gerges, N.Z., Kahn, I., Umanah, G.K.E., Levy, A.P., 2019. An IQSEC2 mutation associated with intellectual disability and autism results in decreased surface AMPA Receptors. Front. Mol. Neurosci. 12, 43. https://doi.org/10.3389/fnmol.2019.00043.

Ruff, K.M., Pappu, R.V., 2021. AlphaFold and implications for intrinsically disordered proteins. J. Mol. Biol. 433 (20). https://doi.org/10.1016/j.jmb.2021.167208, 167208–167208.

Saikia, B., Baruah, A., 2024. Recent advances in de novo computational design and redesign of intrinsically disordered proteins and intrinsically disordered protein regions. Arch. Biochem. Biophys. 752, 109857. https://doi.org/10.1016/j.abb.2023.109857.

Shokhen, M., Walikonis, R., Uversky, V.N., Allbeck, A., Zezelic, C., Feldman, D., Levy, N.S., Levy, A.P., 2023. Molecular modeling of ARF6 dysregulation caused by mutations in IQSEC2. J. Biomol. Struct. Dyn. 42 (3), 1268–1279. https://doi.org/10.1080/07391102.2023.2199085.

Shrestha, U.R., Juneja, P., Zhang, Q., Gurumoorthy, V., Borreguero, J.M., Urban, V., Cheng, X., Pingali, S.V., Smith, J.C., O'Neill, H.M., Petridis, L., 2019. Generation of the configurational ensemble of an intrinsically disordered protein from unbiased molecular dynamics simulation. Proc. Natl. Acad. Sci. U.S.A. 116 (41), 20446–20452. https://doi.org/10.1073/pnas.1907251116.

Shrestha, U.R., Smith, J.C., Petridis, L., 2021. Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations. Commun. Biol. 4 (1), 243. https://doi.org/10.1038/s42003-021-01759-1.

Smith, R.D., Carlson, H.A., 2021. Identification of cryptic binding sites using MixMD with standard and accelerated molecular dynamics. J. Chem. Inf. Model. 61 (3), 1287–1299. https://doi.org/10.1021/acs.jcim.0c01002.

Socci, N.D., Onuchic, J.N., Wolynes, P.G., 1998. Protein folding mechanisms and the multidimensional folding funnel. Proteins 32, 136–158. https://doi.org/10.1002/(SICI)1097-0134(19980801)32:2<136::AID-PROT2>3.0.CO;2-J.

Spiwok, V., Sucur, Z., Hosek, P., 2015. Enhanced sampling techniques in biomolecular simulations. Biotechnol. Adv. 33 (6 Pt2), 1130–1140. https://doi.org/10.1016/j.biotechadv.2014.11.011.

Strodel, B., 2021. Energy landscapes of protein aggregation and conformation switching in intrinsically disordered proteins. J. Mol. Biol. 433 (20), 167182. https://doi.org/10.1016/j.jmb.2021.167182.

Tesei, G., Trolle, A.I., Jonsson, N., Betz, J., Knudsen, F.E., Pesce, F., Johansson, K.E., Lindorff-Larsen, K., 2024. Conformational ensembles of the human intrinsically disordered proteome. Nature 626, 897–904. https://doi.org/10.1038/s41586-023-07004-5.

Toto, A., Malagrinò, F., Visconti, L., Troilo, F., Pagano, L., Brunori, M., Jemth, P., Gianni, S., 2020. Templated folding of intrinsically disordered proteins. J. Biol. Chem. 295 (19), 6586–6593. https://doi.org/10.1074/jbc.REV120.012413.

Tyagi, Chetna, Marik, Tamás, Vágvölgyi, C., Kredics, L., Ötvös, F., 2019. Accelerated molecular dynamics applied to the peptaibol folding problem. Int. J. Mol. Sci. 20 (17), 4268. https://doi.org/10.3390/ijms20174268.

Uversky, V.N., 2017. Flexibility of the "rigid" classics or rugged bottom of the folding funnels of myoglobin, lysozyme, RNase A, chymotrypsin, cytochrome c, and carboxypeptidase A1. Intrinsically Disord. Proteins 5 (1), e1355205. https://doi.org/10.1080/21690707.2017.1355205.

Vangone, A., Bonvin, A.M., 2015. Contacts-based prediction of binding affinity in protein–protein complexes. Elife 4, e07454. https://doi.org/10.7554/eLife.07454.

Wang, J., Arantes, P.R., Bhattarai, A., Hsu, R.V., Pawnikar, S., Huang, Y.-M.M., Palermo, G., Miao, Y., 2021. Gaussian accelerated molecular dynamics (GaMD): principles and applications. Wiley Interdiscip. Rev. Comput. Mol. Sci. 11 (5), e1521. https://doi.org/10.1002/wcms.1521.

Wei, G., Xi, W., Nussinov, R., Ma, B., 2016. Protein ensembles: how does nature harness thermodynamic fluctuations for life? the diverse functional roles of conformational ensembles in the cell. Chem. Rev. 116 (11), 6516–6551. https://doi.org/10.1021/acs.chemrev.5b00562.

Willea, H., Doroshd, L., Amidiana, S., Schmitt-Ulmse, G., Stepanova, M., 2019. Combining molecular dynamics simulations and experimental analyses in protein misfolding. Adv. Protein Chem. Struct. Biol. 118, 33–110. https://doi.org/10.1016/bs.apcsb.2019.10.001.

Wu, Z., Yang, H., Colosi, P., 2010. Effect of genome size on AAV vector packaging. Mol. Ther. 18 (1), 80–86. https://doi.org/10.1038/mt.2009.255.

Xue, Li C., Rodrigues, J.P., Kastritis, P.L., Bonvin, A.M., Vangone, A., 2016. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. Bioinformatics 32, 3676–3678.

Xu, J., McPartlon, M., Li, J., 2021. Improved protein structure prediction by deep learning irrespective of co-evolution information. Nat. Mach. Intell. 3, 601–609.

Zhao, Y., Zhang, J., Zhang, H., Gu, S., Deng, Y., Tu, Y., Hou, T., Kang, Y., 2023. Sigmoid accelerated molecular dynamics: an efficient enhanced sampling method for biosystems. J. Phys. Chem. Lett. 14 (4), 1103–1112. https://doi.org/10.1021/acs.jpclett.2c03688.

Zhou, S., Shi, D., Liu, X., Yao, X., Da, L.-T., Liu, H., 2019. pH-induced misfolding mechanism of prion protein: insights from microsecond-accelerated molecular dynamics simulations. ACS Chem. Neurosci. 10 (6), 2718–2729. https://doi.org/10.1021/acschemneuro.8b00582.

Zhu, J., Li, Z., Tong, H., Lu, Z., Zhang, N., Wei, T., Chen, H.-F., 2023. Phanto IDP: compact model for precise intrinsically disordered protein backbone generation and enhanced sampling. Briefings Bioinf. 25 (1), bbad429. https://doi.org/10.1093/bib/bbad429.