Special Issue on "Theoretical and Computational Chemistry"

J. Indian Chem. Soc.,
Vol. 96, July 2019, pp. 981-990

Transition state ensemble in salt induced protein folding follows Hammond's postulate

Hiranmay Maity and Govardhan Reddy*

Solid State and Structural Chemistry Unit, Indian Institute of Science, Bengaluru-560 012, India

E-mail: greddy@iisc.ac.in

Manuscript received online 02 May 2019, revised and accepted 28 May 2019

The transition state ensemble (TSE) plays a key role in determining the folding time scale of small single domain proteins, which exhibit two-state folding kinetics. Salts can change the stability of protein conformations and strongly influence the protein structure in the TSE. In this paper we studied the effect of protective salts, which stabilise the protein folded states, on the TSE of DNA binding domain of lac repressor protein (*lac*-DBD) and N-terminal domain of ribosomal protein L9 (NTL9) using a coarse-grained protein model and computer simulations. The effect of salts on the protein conformations is taken into account using the molecular transfer model (MTM). We show that the TSE of *lac*-DBD is homogeneous, while the TSE of NTL9 is heterogeneous. The protective salts used in the study shift the position of the TSE for both the proteins towards the unfolded state in agreement with Hammond's postulate.

Keywords: Transition state, P_{fold} analysis, salt effect, Brownian dynamics.

1. Introduction

Salts are extensively used to regulate the activity of proteins in various biological processes. They are regularly used in protein folding studies to understand how they modulate the stability of protein conformations and shift the equilibrium between folded and unfolded states. Most of the small single domain proteins follow two-state folding kinetics i.e. the folded and unfolded states are separated by a transition state (TS) barrier without any intermediate states[1,2]. Salts can strongly influence protein folding kinetics by affecting the TS structures. Unlike the TS in the chemical reactions of small organic molecules, the TS in protein folding is not unique but an ensemble of structures due to the possibility of formation and breaking of many physical contacts between the residues present in the protein. Knowledge of protein structure in the transition state ensemble (TSE) can help us understand the wide variation in the folding time scale of small single domain proteins, which ranges from microseconds to seconds. The short lifetime of the protein structures in the TSE makes it extremely difficult to infer information about these structures directly from experiments. In this paper we studied the effect of protective salts, which stabilize the folded states of proteins, on the TSE of two small single domain proteins using a coarse-grained protein model and computer simulations. We chose DNA binding domain of lac repressor (*lac*-DBD) and N-terminal domain of ribosomal protein L9 (NTL9) as model systems as these proteins are well characterized experimentally.

Experiments typically use denaturants such as urea and guanidinium chloride to perturb the folded states of protein. The effect of denaturants on the relative position of the TS along a folding reaction coordinate is inferred using a technique called linear free energy relationship, which involves measuring the rates of protein folding and unfolding in the presence of denaturants[3,4]. Information such as the compactness of structures in the TSE are inferred from the average change in the solvent accessible surface area (SASA) between the TSE and the unfolded (folded) state, which is related to the slope of the line (*m*-value) obtained by plotting the logarithm of the folding (unfolding) rate as a function of denaturant concentration, [*C*]. Experiments suggest that the position of TS moves toward the native state in the presence of denaturants supporting Hammond's postulate[5]. However, from these experiments, it is difficult to extract structural information such as the topology of protein structures in the TSE.

Protein engineering techniques such as Φ and ψ-value analysis can provide structural information about the TSE[6–9].

In these techniques, mutations are introduced into the protein, and changes in the free energy of the folded state and TS relative to the unfolded state are measured. From these measurements, the extent of structure around the mutated residues in the TS is inferred. Both Φ and ψ-value analysis show that in the TSE, the protein partially attains its native state structure.

However, these techniques disagree on the extent of structure in the TSE for some proteins like protein L and ubiquitin. ψ-value analysis, for example, depicts an extensive native-like topology for ubiquitin[10] and protein L[11], whereas, Φ-value analysis suggests much smaller and polarized structure[12,13]. Complementing the experiments, computer simulations can play an important role in elucidating the structure of the proteins in the TSE. Simulations show that TSE of protein L and ubiquitin are extensive with native-like topology agreeing with the ψ-value analysis[14,15]. Simulations also have the advantage of providing detailed structural information, which can be further tested by experiments.

Simulations using coarse-grained protein models are playing an important role in elucidating the effects of denaturants[16,17], and salts[18] on the folding thermodynamics of proteins. In this paper using the coarse-grained self-organized polymer model with side chains (SOP-SC)[17,19] and molecular dynamics simulations, we studied the effect of protective salts on the TSE of *lac*-DBD and NTL9. The salt effects on the proteins are taken into account using the molecular transfer model (MTM)[18], which utilizes the experimentally measured transfer free energies of various model compounds in different salt solutions[20,21]. The TSE of *lac*-DBD is computed in the presence of ammonium sulphate ($(NH_4)_2SO_4$), potassium chloride (KCl) and ammonium chloride (NH_4Cl), whereas the TSE of NTL9 is constructed in the presence of ammonium sulphate ($(NH_4)_2SO_4$), potassium fluoride (KF) and ammonium chloride (NH_4Cl). Simulations show that in the absence of salt and at the melting temperature, both *lac*-DBD and NTL9 fold in two state manner following nucleation-collapse mechanism[22]. Simulations show that the TSE of

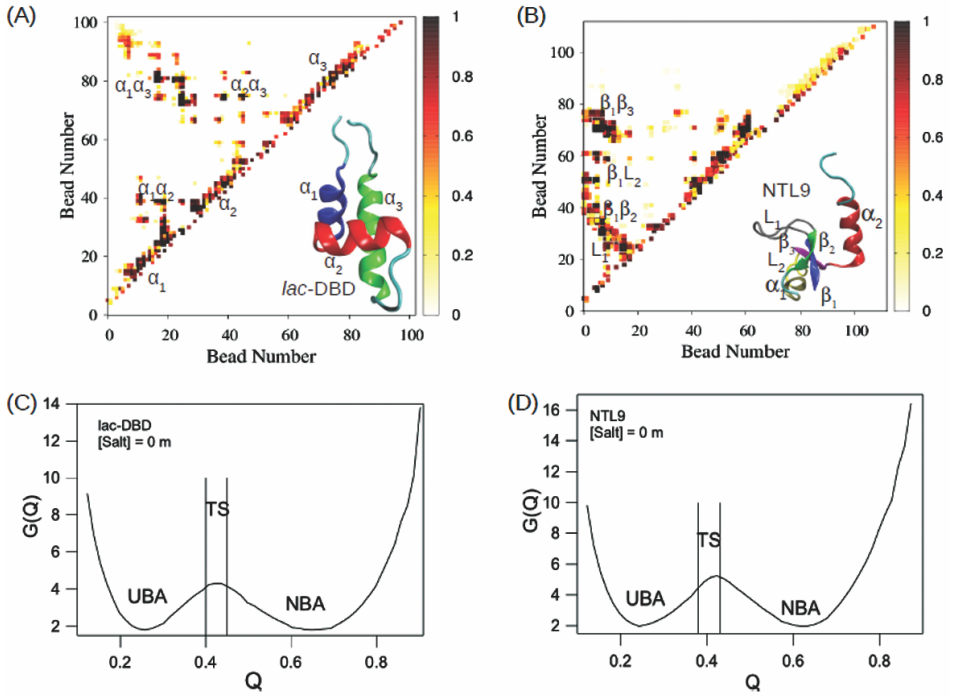

Fig. 1. (A) Contact map and cartoon representation of folded structure of *lac*-DBD (PDB ID: 1OSL). Three α-helices - α_1, α_2 and α_3 are shown in blue, red and green, respectively. (B) Contact map and cartoon representation of NTL9 in the folded state (PDB ID: 1CQU). Two α-helices - α_1 and α_2 are in tan and red, respectively, and the three strands - β_1, β_2 and β_3 are in blue, green and magenta respectively. The loop, L_1, in grey connects β_1 and β_2. The figures are rendered using visual molecular dynamics (VMD)[29]. (C) Free energy surface of *lac*-DBD projected onto the fraction of native contacts Q at T_m and [salt] = 0 molal (m). Putative structures are selected from the TS region using the criteria $0.41 < Q < 0.45$. (D) Free energy of NTL9 is projected onto Q at T_m and [salt] = 0 m. Putative structures for TS analysis are selected using the criteria $0.38 < Q < 0.43$.

lac-DBD is homogeneous whereas TSE of NTL9 is heterogeneous[18]. All the salts used in the present study are protective salts, and in the presence of these salts, fraction of native-like contacts present in the TSE of both *lac*-DBD and NTL9 decreased indicating that the TS moved towards the unfolded state in agreement with Hammond's postulate.

2. Materials and methods

2.1. Self organized polymer with side chain (SOP-SC) model for proteins

We have used the SOP-SC model[17,19] to describe the protein conformations. In this model, an amino acid is represented by two interacting beads - one for the backbone atoms and another for the side chain atoms. The backbone bead is placed at C_α position and the side chain bead is placed at the center of mass of the side chain atoms. We used the solution NMR structure of *lac*-DBD[23] and NTL9[24] with protein data bank (PDB) IDs: 1OSL and 1CQU, respectively to construct the initial coarse-grained structures of the proteins. Hydrogen atoms are added to the PDB structure before calculating the center of mass of the side chain. The energy of a protein conformation in the SOP-SC model is a sum of bonded (E_B), non-bonded (E_{NB}) and electrostatic interactions (E_{el}). The non-bonded interactions are a sum of native (E_{NB}^N) and non-native (E_{NB}^{NN}) interactions. In NTL9, approximately 1/3 of the total residues (18 out of 56) are charged and electrostatic interactions can play a crucial role in the folding of this protein. Electrostatic interactions are included in the energy function of NTL9 but neglected for *lac*-DBD. The energy function of a protein conformation represented by coordinates {**r**} in the SOP-SC model in the absence of salt, [salt] = 0, is given by

$$E_{CG}(\{\mathbf{r}\}, 0) = E_B + E_{NB}^N + E_{NB}^{NN} + \lambda E_{el} \quad (1)$$

In the above equation, the values of λ are 0 and 1 for *lac*-DBD and NTL9, respectively. Detailed description of the SOP-SC model and its energy function is described in the Supporting information of Ref.[18]

2.2. Molecular transfer model (MTM)

The effect of salt on a protein conformation is taken into account using the MTM[18]. In the presence of a salt with concentration [salt] the total energy is given by

$$E_{CG}(\{\mathbf{r}\}, [\text{salt}]) = E_{CG}(\{\mathbf{r}\}, 0) + \Delta G_{tr}(\{\mathbf{r}\}, [\text{salt}]) \quad (2)$$

where $E_{CG}(\{\mathbf{r}\}, 0)$ is given by eq. (1) and $\Delta G_{tr}(\{\mathbf{r}\}, [\text{salt}])$ is the transfer free energy associated with a protein conformation when it is transferred from water to a salt solution with concentration, [salt]. $\Delta G_{tr}(\{\mathbf{r}\}, [\text{salt}])$ of a protein conformation is given by

$$\Delta G_{tr}(\{\mathbf{r}\}, [\text{salt}]) = \sum_{k=1}^{N} \delta g_{tr,k}([\text{salt}])\alpha_k(\{\mathbf{r}\})/\alpha_{Gly\text{-}k\text{-}Gly} \quad (3)$$

where N is the number of beads in the SOP-SC model, $\delta g_{tr,k}([\text{salt}])$ is the transfer free energy of bead k from water to a solution with salt concentration [salt], $\alpha_k(\{\mathbf{r}\})$ is the solvent accessible surface area (SASA) of the bead k in a protein conformation described by positions {**r**}, $\alpha_{Gly\text{-}k\text{-}Gly}$ is the SASA of the bead k in the tripeptide *Gly-k-Gly*. Values of van der Wall radii to calculate $\alpha_k(\{\mathbf{r}\})$ are given in Table S3, $\alpha_{Gly\text{-}k\text{-}Gly}$ are given in Table S4 and $\delta g_{tr,k}([\text{salt}])$ can be calculated using the data in Table S5 in Ref.[18].

2.3. Simulations and data analysis:

We have performed P_{fold} analysis to identify the TSE of proteins[25]. In P_{fold} analysis initially we select putative TS conformations from the barrier region of the free energy surface projected onto the fraction of native-like contacts, Q, at the melting temperature, T_m. Multiple short trajectories are spawned using Brownian dynamics simulations starting from each putative TS conformation and commitment probability P_{fold} is calculated by computing the fraction of trajectories landed into the folded state without visiting the unfolded state. Brownian dynamics simulations are performed using Ermak-McCammon algorithm[26], $\vec{r}_i(t+h) = \vec{r}_i(t) + \frac{h}{\zeta}\vec{F}_c + \vec{\Gamma}$, where $\vec{\Gamma}$ is a random displacement with a Gaussian distribution with mean zero and variance $\left\langle \Gamma(h)^2 \right\rangle = \frac{2k_B Th}{\zeta}$. The friction coefficient $\zeta \approx 5.0\ m/\tau_H$ and, the value of h varies from $0.01\tau_H$ to $0.05\tau_H$ depending on the salt concentration. $\vec{F}_c$ is the deterministic force computed using eq. (1) and eq. (2) in absence and presence of salt, respectively. In the simulations, the characteristic unit of length a = 1 Å, energy $\in$ = 1 kcal/mol, and mass $m = 1.8\times10^{-22}$ g (typical mass of the bead).

We computed Q to monitor the protein folding kinetics, which is defined as $Q = \frac{1}{N_N}\sum_{i=1}^{N_N}\Theta\left(\delta - |r_i - r_i^0|\right)$, where N_N (= $N_N^{bb} + N_N^{bs} + N_N^{ss}$) is the total number of native contacts

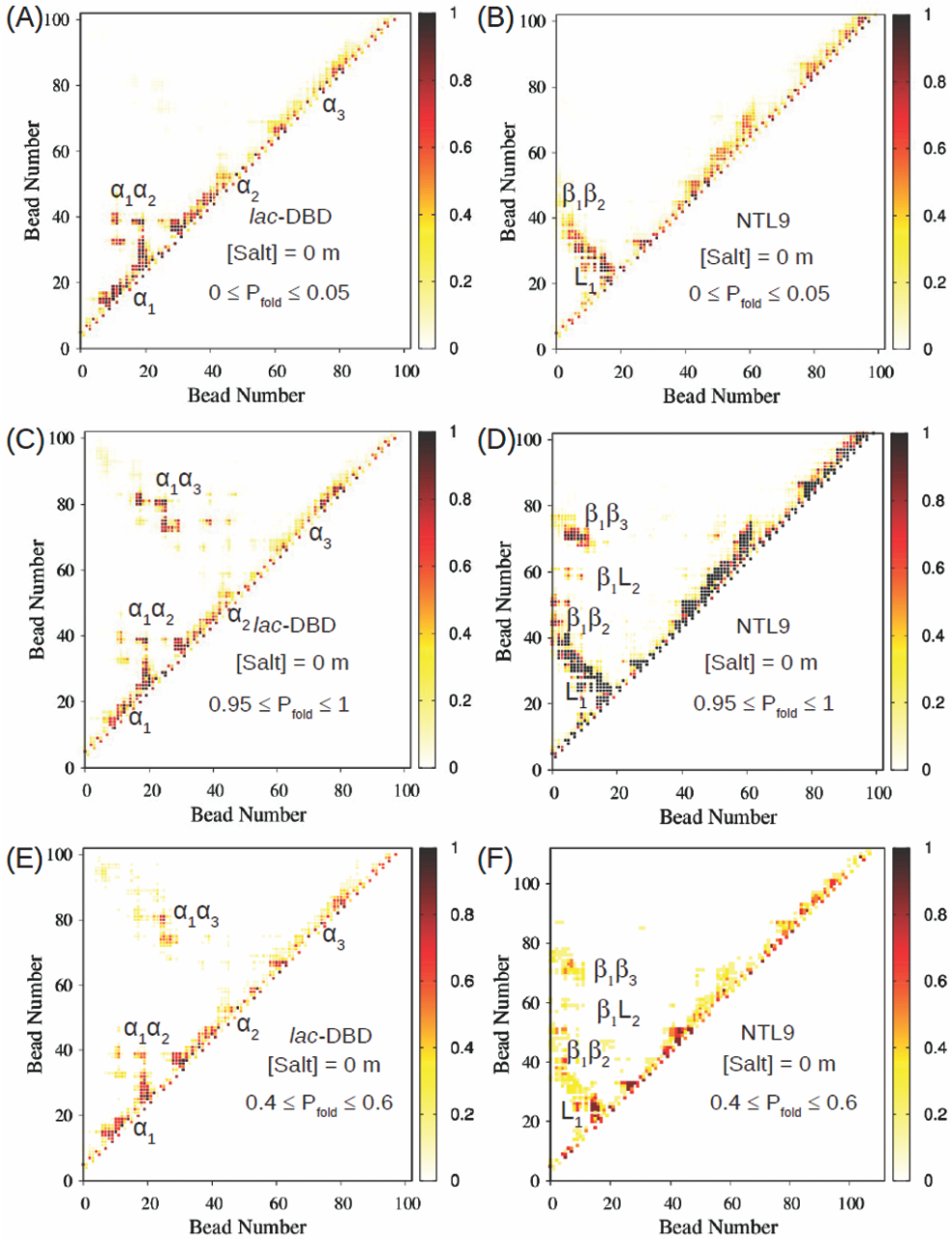

Fig. 2. (A) Contact map of *lac*-DBD putative TS structures, which satisfy $0 \le P_{fold} \le 0.05$ shows that long-ranged contacts between α_1 and α_3 are absent in these structures. (B) Contact map of NTL9 putative TS structures, which satisfy $0 \le P_{fold} \le 0.05$ shows that protein conformations are devoid of prominent long range contacts between β_1 and L_2 or β_1 and β_3. (C) Contact map of *lac*-DBD putative TS structures, which satisfy $0.95 \le P_{fold} \le 1$ shows that strong contacts are present between α_1 and α_3 SSEs. (D) Contact map of NTL9 putative TS structures, which satisfy $0.95 \le P_{fold} \le 1$ shows that long-ranged contacts between the residues of β_1 and L_2 or β_3 are present. (E) Contact map of the TSE of *lac*-DBD, which satisfy $0.4 \le P_{fold} \le 0.6$ shows partial contacts formed between α_1 and α_3 are crucial in the formation of critical nucleus. (F) Contact map of the TSE of NTL9, which satisfy $0.4 \le P_{fold} \le 0.6$ shows that long-ranged contacts between β_1 and L_2 or β_3 are essential in the formation of critical nucleus.

present between pairs of beads in the folded state of SOP-SC model of protein. N_N^{bb}, N_N^{bs} and N_N^{ss} are the number of native contacts present between backbone-backbone, backbone-side chain and side chain-side chain beads, respec-

tively (Table S2 in Ref.[18]). r_i is the distance between the *i*-th pair of beads, and r_i^0 being the corresponding distance in the folded state, Θ is Heaviside step function, and δ = 2 Å.

3. Results and discussion

3.1. TSE of lac-DBD and NTL9 in the absence of salts

lac-DBD and NTL9 are experimentally[27,28] well characterized. The *lac*-DBD structure with the PDB ID: 1OSL has 62 residues with three α-helices (α_1 to α_3) and an unstructured C-terminal part. In the experiments[27] only the structured part of the protein is used to study the folding thermodynamics. In order to compare the simulation data with the experiments, we also studied the folding of only the structured part of the protein (residues Met1 to Arg51) (Fig. 1A) and ignored the disordered part (residues Cys52 to Leu62). NTL9 is an α/β protein with 56 residues. It has three β-strands (β_1 to β_3) and two α-helices (α_1 and α_2). β_1 and β_2 are connected via loop, L_1 (Fig. 1B). Native contacts between the coarse-grained protein beads in the folded state are shown in the contact map (Fig. 1). The local contacts within a secondary structural element (SSE) appear along the diagonal of the contact map whereas tertiary contacts between two SSEs appear as off-diagonal contacts.

At the melting temperature, T_m (= 323 K for *lac*-DBD and 370 K for NTL9), the proteins show cooperative folding[18], and they hop between folded and unfolded state separated by a critical Q value, Q_c (= 0.41 for NTL9 and 0.43 for *lac*-DBD). Q_c is the position of the TS when the folding free energy is projected onto Q (Fig. 1C, D). The free energy minima at $Q > Q_c$ and $Q < Q_c$ corresponds to the native basin of attraction (NBA) and unfolded basin of attraction (UBA), respectively. To identify TSE using P_{fold}, we initially selected putative structures from the TS region. The TS region considered for *lac*-DBD is $0.41 \leq Q \leq 0.45$ and for NTL9 is $0.38 \leq Q \leq 0.43$. For each putative structure, we generated five hundred Brownian dynamics trajectories for short time period of 15000 τ_{BD} so that the trajectories end up either in the NBA or UBA state. We computed P_{fold} by calculating the fraction of trajectories landed in the NBA state without visiting the UBA state. The P_{fold} values of putative TS conformations are between 0 and 1. If P_{fold} = 0, then none of the trajectories landed in the NBA and if it is 1 then all the trajectories landed in the NBA. We used the criteria that a putative conformation belongs to TSE if the P_{fold} value is between 0.4 and 0.6. The TSE computed using the P_{fold} shows that *lac*-DBD is homogeneous whereas TSE of NTL9 is heterogeneous and diffusive[18].

lac-DBD and NTL9 at T_m fold through nucleation-condensation mechanism where the tertiary contacts between different SSEs almost form simultaneously to stabilize the protein folding nucleus. The structures in the TSE show native-like tertiary contacts. To identify the critical contacts

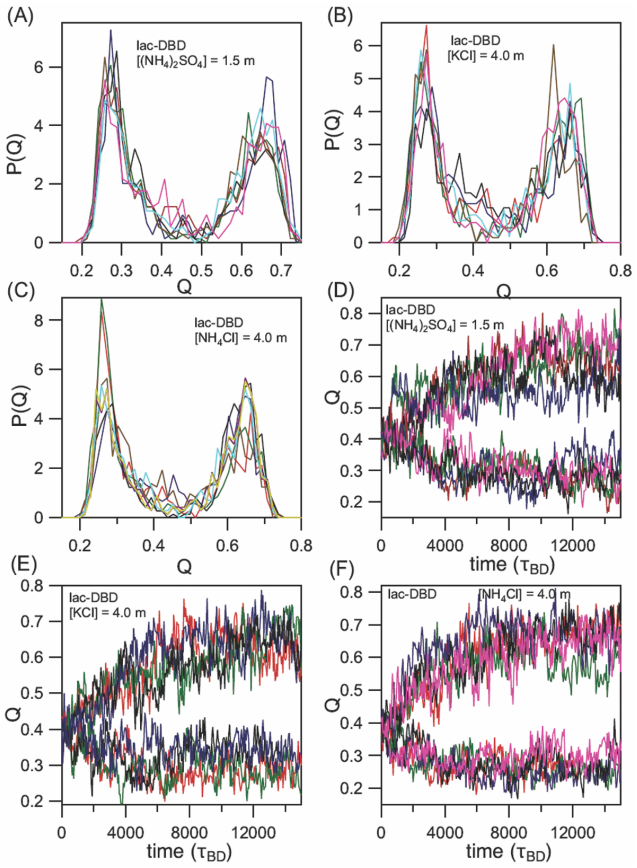

Fig. 3. P_{fold} analysis of *lac*-DBD at T = 323 K and (A) [$(NH_4)_2SO_4$] = 1.5 m, (B) [KCl] = 4.0 m, and (C) [NH_4Cl] = 4.0 m. Probability distribution of final Q values, $P(Q)$, computed from five hundred short Brownian dynamics trajectories spawned using a protein conformation from the TSE of *lac*-DBD as the starting initial conformation. $P(Q)$ for different TS conformations are shown in different colors. Representative Brownian dynamics simulation trajectories spawned using a protein structure from the TSE obtained at T = 323 K and (D) [$(NH_4)_2SO_4$] = 1.5 m, (E) [KCl] = 4.0 m, and (F) [NH_4Cl] = 4.0 m as the initial starting conformation show that the trajectories land with almost equal probability in both the NBA and UBA states.

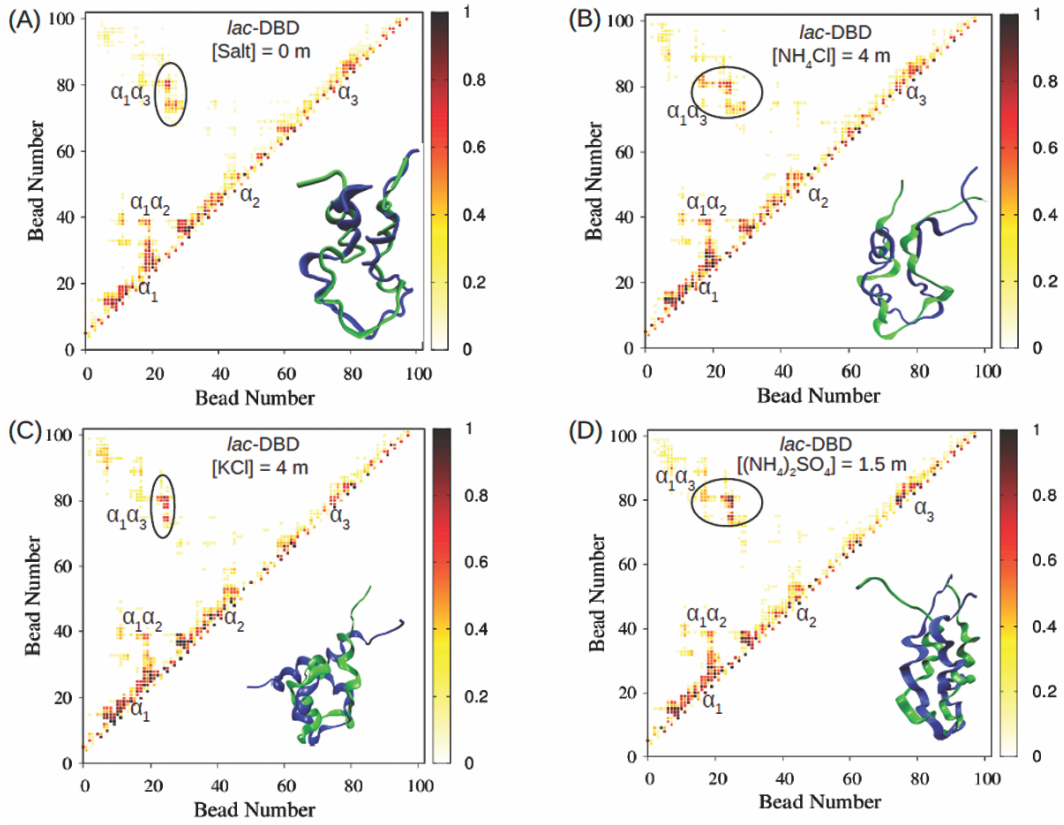

Fig. 4. Contact map of *lac*-DBD TSE and two superimposed representative conformations in blue and green from the TSE obtained at T = 323 K and (A) [salt] = 0 m, (B) [NH_4Cl] = 4.0 m, (C) [KCl] = 4.0 m, and (D) [$(NH_4)_2SO_4$] = 1.5 m. Long-ranged tertiary contacts present between the residues are highlighted.

present between the residues in the protein conformations, which form the TSE, we constructed the contact maps of putative transition structure, which have P_{fold} values in the range $0 \leq P_{fold} \leq 0.05$ and $0.95 \leq P_{fold} \leq 1$, and, compared them with the contact map of the structures in the TSE obtained using the condition $0.4 \leq P_{fold} \leq 0.6$ (Fig. 2). For *lac*-DBD, the contact map generated using the putative TS structures, which satisfy the condition $0 \leq P_{fold} \leq 0.05$ shows formation of prominent contacts in $\alpha_1\alpha_2$ and very few weak $\alpha_1\alpha_3$ contacts (Fig. 2A) indicating just $\alpha_1\alpha_2$ contacts are not sufficient for the formation of critical folding nucleus. On the other hand, the contact map of the putative structures, which satisfy $0.95 \leq P_{fold} \leq 1$ have strong native-like contacts in both the SSEs $\alpha_1\alpha_2$ and $\alpha_1\alpha_3$ (Fig. 2C). The contact map of *lac*-DBD structures in the TSE shows presence of partially formed contacts in both $\alpha_1\alpha_2$ and $\alpha_1\alpha_3$. The contacts in $\alpha_1\alpha_3$ however are important tertiary contacts, which control the formation of TS structure (Fig. 2E). The TSE of *lac*-DBD is homogeneous as all the protein conformations have similar tertiary contacts between the SSEs.

Similarly, the contact map of NTL9 putative TS structures, which satisfy the condition $0 \leq P_{fold} \leq 0.05$ shows formation of local native-like contacts in the SSEs $\beta_1\beta_2$ and loop region L_1 (Fig. 2B). However, these contacts are not sufficient enough to form a stable folding nucleus. The contact map of putative TS structures, which satisfy the condition $0.95 \leq P_{fold} \leq 1$ shows formation of strong native-like tertiary contacts in the SSEs β_1L_2 and $\beta_1\beta_3$ region (Fig. 2D), which are again partially formed in the TSE of NTL9 (Fig. 2F). Structural characterization of the TSE of NTL9 showed that it is heterogeneous in nature[18] as protein conformations can have contacts between different SSEs. The protein conformations in the TSE have tertiary contacts in the SSEs β_1L_2 or $\beta_1\beta_3$ or both. These contacts are missing in the putative TS conformations, which land in the UBA in P_{fold} analysis.

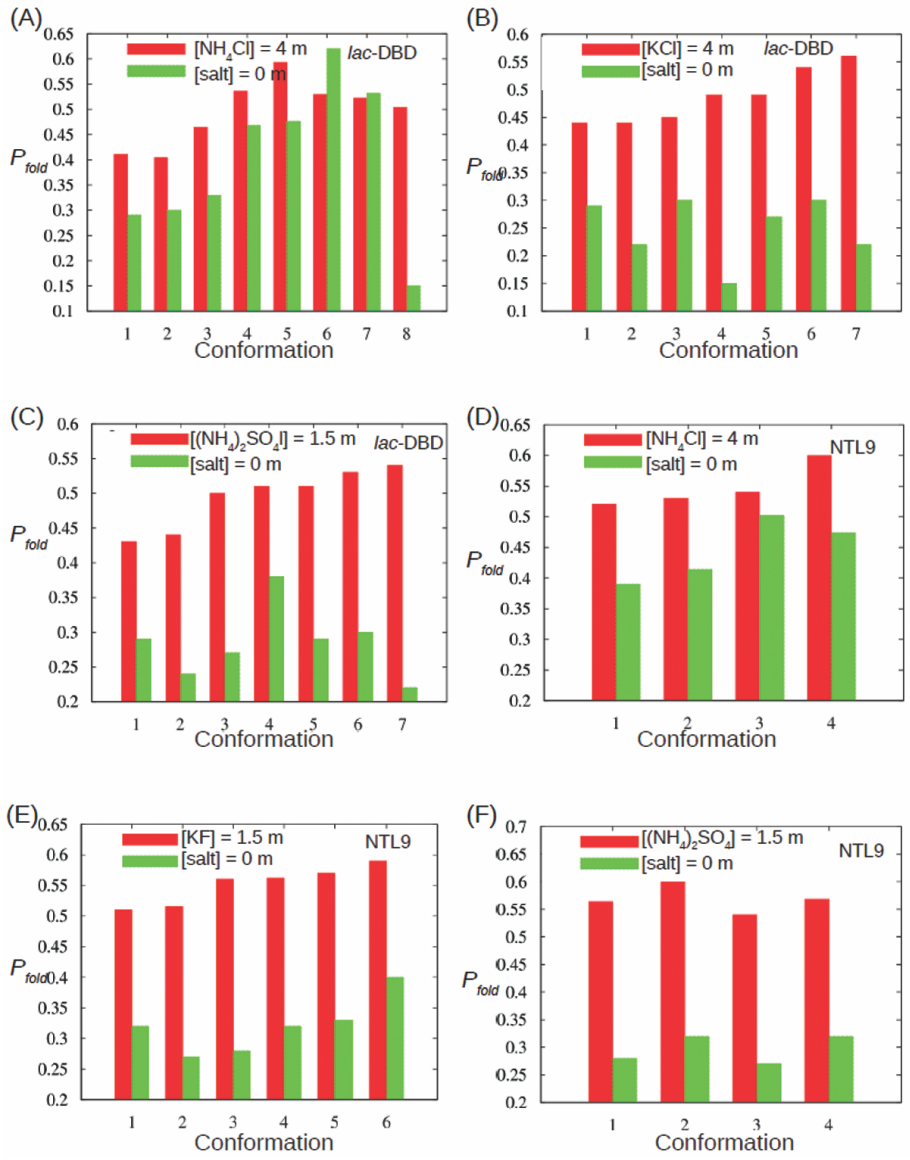

Fig. 5. P_{fold} values for different TS structures from the *lac*-DBD TSE obtained at T = 323 K and (A) [NH_4Cl] = 4 m, (B) [KCl] = 4 m and (C) [$(NH_4)_2SO_4$] = 1.5 m are in red and their corresponding P_{fold} values at [salt] = 0 m are in green. Similarly P_{fold} values for different TS structures from the NTL9 TSE obtained at T = 370 K and (D) [NH_4Cl] = 4 m, (E) [KF] = 1.5 m and (F) [$(NH_4)_2SO_4$] = 1.5 m are in red and their corresponding P_{fold} values at [salt] = 0 m are in green.

3.2. Salt effects on the TSE of lac-DBD

We have studied the effects of $(NH_4)_2SO_4$, KCl and NH_4Cl on the TSE of *lac*-DBD. Among the three salts, $(NH_4)_2SO_4$ strongly stabilizes the NBA state, whereas, KCl and NH_4Cl moderately stabilize the NBA state. We have extracted the TSE of *lac*-DBD in the presence of salts using P_{fold} analysis. Protective salts can decrease the height of the free energy barrier between UBA and NBA states, and as a result in high salt concentrations an unfolded protein conformation can easily transit to the NBA state on short simulation time scales making the P_{fold} calculations difficult. To illustrate the effect of salts on the *lac*-DBD TSE, the optimum salt concentrations we chose for $(NH_4)_2SO_4$, KCl and NH_4Cl are 1.5 molal (m), 4 m and 4 m, respectively. The short simulation trajecto-

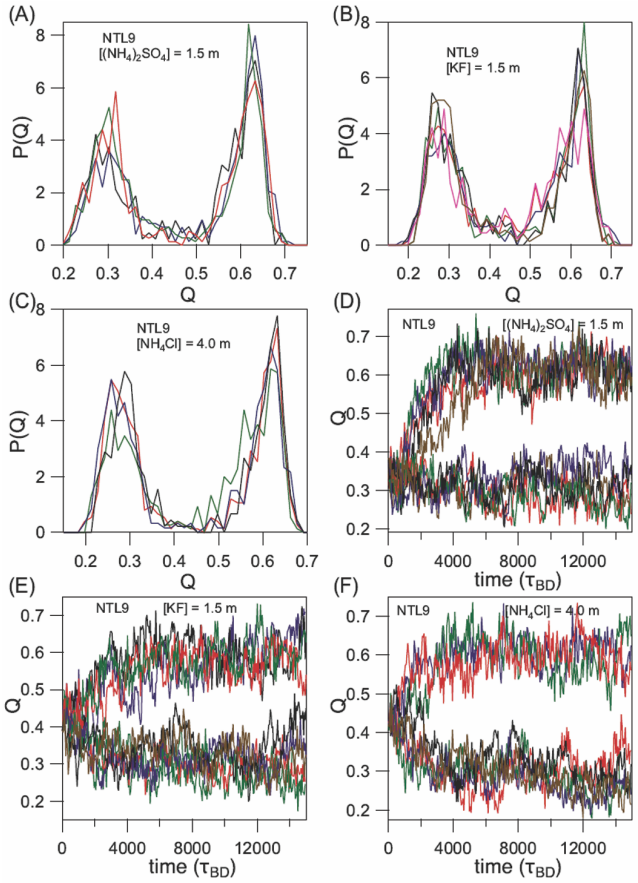

Fig. 6. P_{fold} analysis of NTL9 at T = 370 K and (A) [$(NH_4)_2SO_4$] = 1.5 m, (B) [KF] = 1.5 m, and (C) [NH_4Cl] = 4.0 m. Probability distribution of final Q values, $P(Q)$, computed from five hundred short Brownian dynamics trajectories spawned using a protein conformation from the TSE of NTL9 as the starting initial conformation. $P(Q)$ for different TS conformations are shown in different colors. Representative Brownian dynamics simulation trajectories spawned using a protein structure from the TSE obtained at T = 370 K and (D) [$(NH_4)_2SO_4$] = 1.5 m, (E) [KF] = 1.5 m, and (F) [NH_4Cl] = 4.0 m as the initial starting conformation show that the trajectories land with almost equal probability in both the NBA and UBA states.

ries spawned using the structures in the TSE at the chosen salt concentration land in either NBA or UBA state (Fig. 3). For each salt, P_{fold}, analysis is performed for 370 putative TS structures, initially generated at T = 323 K and [salt] = 0 m with fraction of native contacts, Q, between 0.41 and 0.45. The number of protein conformations identified as TS structures in the presence of NH_4Cl, KCl and $(NH_4)_2SO_4$ are eight, seven and seven, respectively. The TSE of *lac*-DBD is homogeneous (Fig. 4) and the fraction of native-like contacts (Q_{TS}) calculated for the TSE in presence of salts is ≈ 0.42, which is comparable to Q_{TS} obtained in the absence of salts[18].

To illustrate the movement of TS towards the unfolded state in the presence of protective salts, we took the protein conformations in the TSE obtained in the presence of salt, and ran P_{fold} again for these conformations in zero salt concentration (Fig. 5A,B,C). The P_{fold} value in the absence of salt decreases to lower value implying that the conformations are closer to the UBA state and less likely to fold in the absence of protective salt. The average of P_{fold} value obtained for TSE of *lac*-DBD in presence of NH_4Cl, KCl and $(NH_4)_2SO_4$ are 0.496, 0.487 and 0.494, respectively, whereas, corresponding P_{fold} values in absence of salt are 0.396, 0.25 and 0.284, respectively. Decrease in P_{fold} value in absence of salt is consistent with Hammond's postulate, which says that the TSE should shift towards the UBA state in the presence of protective salts.

3.3. Salt effects on the TSE of NTL9

The salt concentrations we chose for NH_4Cl, KF and $(NH_4)_2SO_4$ to calculate the TSE for NTL9 are 4 m, 1.5 m and 1.5 m, respectively. The putative TS structures for P_{fold} analysis are initially generated at T = 370 K and [salt] = 0 m with $0.38 \leq Q \leq 0.43$. The number of protein conformations satisfying $0.4 \leq P_{fold} \leq 0.6$ in the presence of $(NH_4)_2SO_4$, KF and NH_4Cl are four, six and four, respectively. Representative short simulation trajectories initiated using the protein conformations in the TSE land with almost equal probability in the UBA or NBA state (Fig. 6). The protein conformations in the TSE are structurally heterogeneous (Fig. 7). Average fraction of native contacts present in the TSE (Q_{TS}) of NTL9 in the presence of $(NH_4)_2SO_4$, KF and NH_4Cl are ≈0.39, ≈ 0.39 and ≈0.40, respectively. These values are comparable to Q_{TS} of NTL9 at [salt] = 0 m, which is ≈0.40.

Again to illustrate the movement of TS towards the unfolded state in the presence of protective salts, we took the protein conformations in the TSE obtained in the presence of salt, and ran P_{fold} calculations for these conformations in zero salt concentration (Fig. 5D,E,F). The P_{fold} values obtained in the absence of protective salt are lower than the values obtained in the presence of salt. The average P_{fold}

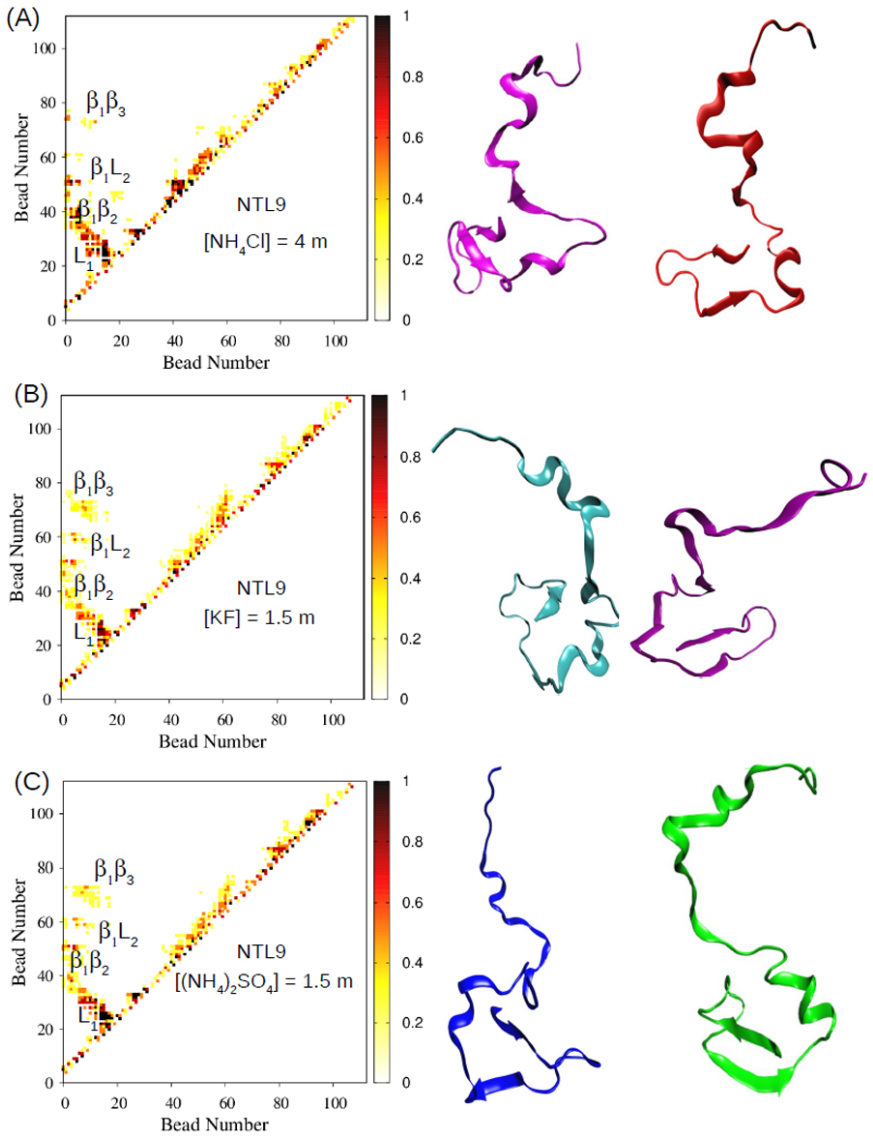

Fig. 7. Contact map of NTL9 TSE and representative heterogeneous conformations from the TSE obtained at T = 370 K and (A) [NH_4Cl] = 4.0 m, (B) [KF] = 1.5 m, and (C) [$(NH_4)_2SO_4$] = 1.5 m.

values for the TSE of NTL9 in the presence of $(NH_4)_2SO_4$, KF and NH_4Cl are 0.57, 0.56 and 0.55, respectively, whereas P_{fold} values for the TSE in the absence of salt are 0.3, 0.32 and 0.45, respectively. This suggests that the protein conformations in the TSE in the presence of protective salts move towards the UBA state following Hammond's postulate.

4. Conclusions

Using molecular dynamics simulations and a coarse-grained protein model[18] built based on the experimentally measured free energies for the transfer of chemical groups from water to salt solutions we computed the TSE of two small single domain proteins, *lac*-DBD and NTL9 in the presence and absence of protective salts. The simulations show that the protein conformations in the TSE can be either homogeneous with all the protein conformations in the TSE having a similar folded topology or heterogeneous with the structures having different topology. We further show that the salts induce very subtle changes in the protein conformations in the TSE compared to the structures in the TSE obtained in the absence of salt. Although the changes in the protein conformations in the TSE obtained in the presence

and absence of protective salts are difficult to infer directly, the P_{fold} calculations show that the ensemble slightly moves towards the protein unfolded state following Hammond's postulate.

Acknowledgement

GR acknowledges funding from Science and Engineering Research Board (EMR/2016/001356) and Nano mission, Department of Science and Technology, India. HM acknowledges research fellowship from Indian Institute of Science-Bangalore. The computations are performed using the TUE and Cray XC40 cluster at IISc.

References

1. S. E. Jackson and A. R. Fersht, *Biochemistry*, 1991, **30(43)**, 10428.
2. S. E. Jackson, *Fold Des.*, 1998, **3**, R81.
3. A. Matouschek and A. R. Fersht, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 7814.
4. A. Matouschek, D. E. Otzen, L. S. Itzhaki, S. E. Jackson and A. R. Fersht, *Biochemistry*, 1995, **34**, 13656.
5. G. S. Hammond, *J. Am. Chem. Soc.*, 1955, **77**, 334.
6. C. R. Matthews, *Method Enzymol.*, 1987, **154**, 498.
7. A. R. Fersht, A. Matouschek and L. Serrano, *J. Mol. Biol.*, 1992, **224(3)**, 771.
8. B. A. Krantz and T. R. Sosnick, *Nat. Struct. Biol.*, 2001, **8(12)**, 1042.
9. B. A. Krantz, R. S. Dothager and T. R. Sosnick, *J. Mol. Biol.*, 2004, **337(2)**, 463.
10. T. R. Sosnick, B. A. Krantz, R. S. Dothager and M. Baxa, *Chem. Rev.*, 2006, **106(5)**, 1862.
11. T. Y. Yoo, A. Adhikari, Z. Xia, T. Huynh, K. F. Freed, R. Zhou and T. R. Sosnick, *J. Mol. Biol.*, 2012, **420(3)**, 220.
12. H. M. Went and S. E. Jackson, *Protein Eng.*, 2005, **18(5)**, 229.
13. D. E. Kim, C. Fisher and D. Baker, *J. Mol. Biol.*, 2000, **298(5)**, 971.
14. H. Maity and G. Reddy, *J. Am. Chem. Soc.*, 2016, **138(8)**, 2609.
15. G. Reddy and D. Thirumalai, *J. Phys. Chem. B*, 2015, **119(34)**, 11358.
16. E. P. O'Brien, G. Ziv, G. Haran, B. R. Brooks and D. Thirumalai, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 13403.
17. Z. Liu, G. Reddy, E. P. O'Brien and D. Thirumalai, *Proc. Natl. Acad. Sci. USA*, 2011, **108(19)**, 7787.
18. H. Maity, A. N. Muttathukattil and G. Reddy, *J. Phys. Chem. Lett.*, 2018, **9(17)**, 5063.
19. C. B. Hyeon, R. I. Dima and D. Thirumalai, *Structure*, 2006, **14**, 1633.
20. M. Record and C. Anderson, *Biophys. J.*, 1995, **68**, 786.
21. E. Courtenay, M. Capp, R. Saecker and M. Record, *Proteins: Struct. Funct. Genet.*, 2000, **41**, 72.
22. A. R. Fersht, *Proc. Natl. Acad. Sci. USA*, 1995, **92(24)**, 10869.
23. C. G. Kalodimos, N. Biris, A. M. J. J. Bonvin, M. M. Levandoski, M. Guennuegues, R. Boelens and R. Kaptein, *Science*, 2004, **305(5682)**, 386.
24. D. L. Luisi, B. Kuhlman, K. Sideras, P. A. Evans and D. P. Raleigh, *J. Mol. Biol.*, 1999, **289(1)**, 167.
25. R. Du, V. Pande, A. Grosberg, T. Tanaka and E. Shakhnovich, *J. Chem. Phys.*, 1998, **108**, 334.
26. D. L. Ermak and J. A. Mccammon, *J. Chem. Phys.*, 1978, **69(4)**, 1352.
27. L. M. Pegram, T. Wendor, R. Erdmann, I. Shkel, D. Bellissimo, D. J. Felitsky and M.T. Record (Jr.), *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 7716.
28. R. Sengupta, A. Pantel, X. Cheng, I. Shkel, I. Peran, N. Stenzoski, D. P. Raleigh and M. T. Record (Jr.), *Biochemistry*, 2016, **55(15)**, 2251.
29. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14(1)**, 33.