

## Modeling pKas of unfolded proteins to probe structural models of unfolded state

Nayere Tajielyato and Emil Alexov\*

*Department of Physics and Astronomy  
Clemson University, Clemson, SC 29630, USA*

*\*ealexov@clemson.edu*

Received 21 May 2019

Accepted 27 May 2019

Published 24 June 2019

Modeling unfolded states of proteins has implications for protein folding and stability. Since in unfolded state proteins adopt multiple conformations, any experimentally measured quantity is ensemble averaged, therefore the computed quantity should be ensemble averaged as well. Here, we investigate the possibility that one can model an unfolded state ensemble with the coil model approach, algorithm such as “flexible-meccano” [Ozenne V *et al.*, Flexible-meccano: A tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables, *Bioinformatics* **28**:1463–1470, 2012], developed to generate structures for intrinsically disordered proteins. We probe such a possibility by using generated structures to calculate pKas of titratable groups and compare with experimental data. It is demonstrated that even with a small number of representative structures of unfolded state, the average calculated pKas are in very good agreement with experimentally measured pKas. Also, predictions are made for titratable groups for which there is no experimental data available. This suggests that the coil model approach is suitable for generating 3D structures of unfolded state of proteins. To make the approach suitable for large-scale modeling, which requires limited number of structures, we ranked the structures according to their solvent accessible surface area (SASA). It is shown that in the majority of cases, the top structures with smallest SASA are enough to represent unfolded state.

**Keywords:** Unfolded state; pKas; electrostatics; Poisson–Boltzmann; dielectric constant; conformations.

### 1. Introduction

The unfolded state and intrinsically disordered proteins (IDPs) or regions (IDRs) adopt a conformational landscape which is still not very well understood. Generating the “right” conformational ensemble of an unfolded state is not an easy task and needs knowledge of how the protein was unfolded, including what denaturants were

\*Corresponding author.

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

used and what the conditions were.<sup>1</sup> In some cases, the unfolded state ensembles contain hydrophobic clusters and/or well-defined secondary structure elements.<sup>2-4</sup> The importance of knowledge of unfolded state is demonstrated in modeling folding free energy, the difference of the energy of folded and unfolded states. Consequently, the structural information of the unfolded state is valuable for mechanistic and thermodynamic studies of folding.

Ionizable side chains represent about 30% of the proteins' residues and are crucial for stability and solubility of proteins.<sup>5-7</sup> The pKa values of ionizable residues determine the pH-dependence of protein folding and binding energies.<sup>8-13</sup> In some cases, modeling the pH-dependence of stability requires knowledge of pKas in unfolded state.<sup>14-17</sup> While it is frequently assumed that the pKas in an unfolded state are unperturbed (equal to intrinsic pKas), there are also studies that show the role of electrostatic interactions between charged residues in the denatured state.<sup>18,19</sup> Additionally, some studies described the impact of changes in the protonation state of ionizable residues in the formation of hydrophobic clusters in denatured state.<sup>20</sup> This indicates that some titratable groups may have perturbed pKas in their unfolded state.

Experimentally, the pKa values of titratable residues in the unfolded state were measured for some proteins using NMR spectroscopy, site-directed mutagenesis and CD temperature-denaturation measurements.<sup>21-24</sup> Most of these experiments have indicated that the pKa values of acidic residues are lowered in unfolded state compared with intrinsic pKas.<sup>21,22,25</sup> These studies are used in this work to benchmark DelPhiPKa along with 3D models of unfolded states.

Previous attempts to model pKas of unfolded states were applied to either molecular dynamics (MD) simulation to generate structures of unfolded state<sup>14</sup> or Gaussian-chain model.<sup>15-17</sup> MD with artificially increased van der Walls atomic radii was used to "unfold" the proteins and obtained structures to calculate pKas of unfolded state.<sup>14</sup> Alternatively, the electrostatic interactions in unfolded state were estimated with Gaussian-chain model and then were used to compute the corresponding pKas.<sup>15-17</sup> In our work, we took a different approach, taking advantage of advances made in the field of IDPs.

A great amount of effort has been invested to develop methods to predict the conformational ensembles of IDPs. A straightforward approach to model the conformational ensemble of unfolded state is the MD simulation.<sup>26</sup> Another approach, the coil models, alternatively describes the conformational ensemble of IDPs by sampling different conformational states of individual residues using the coil libraries based on the experimental information of residue-specific  $\{\varphi, \psi\}$  angles.<sup>27-30</sup> Several studies showed that the results from coil modeling accurately reproduce experimental data such as J-coupling,<sup>27</sup> residual dipolar couplings (RDCs)<sup>28,30</sup> and small-angle X-ray scatterings (SAXS) curves for both IDPs<sup>28,30</sup> and unfolded proteins.<sup>31</sup> Thus, in our work the unfolded proteins are generated via coil models and we use the conformational ensembles generated by a particular software, the flexible-meccano software.<sup>30</sup>

Our interest in modeling unfolded state stems from the importance of unfolded state in calculating folding free energy changes due to amino acid mutations, since this is related to understanding effects of disease-causing nonsynonymous single nucleotide polymorphisms (nsSNPs).<sup>32–35</sup> All available methods based on energy difference between two states, folded and unfolded, use an approximation and consider only a few residues as representative of unfolded state.<sup>36–39</sup> This is clearly a severe assumption that omits residual interactions in unfolded state. With this work, we would like to suggest that one can explicitly generate representative full length structures of unfolded state using the tools for modeling IDPs, the flexible-meccano software.<sup>30</sup>

In this work, we use the flexible-meccano software<sup>30</sup> to generate ensemble of structures for unfolded state of selected proteins for which there is experimental data of pK<sub>a</sub>s in unfolded conditions. These structures are used to compute pK<sub>a</sub>s of titratable residues with DelPhiPKa<sup>40,41</sup> and results compared with experimental data. To provide some guidance how to select minimal number of representative structures (to be applicable for large-scale calculations), we investigate the possibility of the generated structures ranked with respect to solvent accessibility surface area (SASA). Furthermore, we make predictions for pK<sub>a</sub>s of unfolded state for titratable groups that are not reported in the literature.

## 2. Methods

The main purpose of this investigation is to probe if structures generated with flexible-meccano can be used as representative structure for modeling unfolded state. Such a possibility is tested via comparing experimental pK<sub>a</sub>s of unfolded state and predicted pK<sub>a</sub>s with DelPhiPKa using structures generated with flexible-meccano. However, one can use this approach to predict pK<sub>a</sub>s delivered via unfolding experiments and to infer the pH-dependence of folding as well.

**Dataset:** We searched for proteins with experimentally measured pK<sub>a</sub> values of titratable residues in the unfolded state. This resulted in the following cases: Turkey Ovomucoid Third Domain (OMTKY3),<sup>25</sup> Barnase,<sup>21</sup> chymotrypsin inhibitor 2 (CI2),<sup>22</sup> protein G B1 domain (PGB1)<sup>23</sup> and Hen Egg white Lysozyme (HEWL).<sup>42</sup>

**Unfolded state:** The flexible-meccano algorithm<sup>30</sup> was used to generate ensemble of structures for the unfolded state. This method builds multiple different copies of the polypeptide chain based on the random sampling backbone dihedral angle potential wells assuring that these conformations do not self-overlap (Fig. S1). This method was shown to successfully reproduce the experimental data such as RDC and SAXS by averaging over the built-ensembles.<sup>30</sup> Thus, we generated 1000 structures for each protein in our data set using default parameters of flexible-meccano software.

**pK<sub>a</sub> calculations:** The pK<sub>a</sub> calculations of titratable residues were performed by DelPhiPKa which is a Poisson–Boltzmann-based approach.<sup>41</sup> Most of the calculations were done using default values of DelPhiPKa parameters. However, to study the effect of D<sub>2</sub>O (deuterium oxide or heavy water) and H<sub>2</sub>O in case of thermal denaturation, we considered a range of solvent dielectric constants of 74, 75, 78.5 and

80 since most of these experimental works were done for the H<sub>2</sub>O/D<sub>2</sub>O at 25°C in which H<sub>2</sub>O and D<sub>2</sub>O have dielectric constants of 78.39 and 78.06.<sup>43</sup> A dielectric constant of 4 and 8 was considered for each protein to test the sensitivity of results. The charge and radius parameters were assigned using PARSE.<sup>44</sup> The salt concentration was set to a value that matches the experimental conditions. Therefore, the salt concentrations for OMTKY3, Barnase, C12, PGB1 and HEWL were 0.2, 0.05, 0.05, 0.001 and 0.144 M, respectively.

Solvent accessible surface area: SASA was calculated by VMD<sup>45</sup> for each structure in structural ensembles with the solvent probe radius of 1.4 Å. It was used to rank the structures according min/max SASA.

### 3. Results and Discussions

Here, we report DelPhiPKa calculated pKas of ionizable groups using flexible-mecano generated structures. For each of the proteins, we generated 1000 structures and carried pKas calculations. At the same time, we address the possibility to reduce the number of needed structures by selecting the top 15, either with maximum (very extended structures) or minimum (relatively compact structures) SASA. We also investigate the effect of different values of internal and external dielectric constants. Lysines and argamines are excluded from our investigation, because there is no experimental data.

#### 3.1. OMTKY3

A study of the pH dependence of OMTKY3 stability showed that among all the titratable residues, the pKas of Asp7 and Asp27 are perturbed and therefore they contribute to the pH dependence of OMTKY3 stability.<sup>25</sup> The pKa of Asp7 and Asp27 in unfolded state were reported as 3.9 and 3.6.

We calculated the pKa values of titratable residues using all 1000 structures and considering ionic concentration of 200 mM and solvent and solute dielectric constant of 80 and 8, respectively (Table 1). It can be seen that predicted values for Asp7 and Asp27 are 3.87 and 3.56, which are almost the same as the ones found in the experimental work.<sup>25</sup> Additionally, we predicted the pKas of all other acidic groups

Table 1. Predicted pKa values of titratable residues in unfolded state of OMTKY3. (1), (2) and (3) are the average pKa's of all 1000 snapshots, 15 structures with minimum SASA, and 15 structures with maximum SASA. The standard deviations are given in parenthesis.

Unfolded state set	Asp7	Asp27	Glu10	Glu19	Glu43	His52
Predicted pKa <sup>(1)</sup>	3.89(0.3)	3.69(0.29)	3.97(0.24)	3.84(0.34)	3.89(0.31)	6.31(0.49)
Predicted pKa with minimum SASA <sup>(2)</sup>	3.96(0.44)	3.71(0.30)	3.9(0.15)	3.51(1.00)	3.83(0.39)	6.13(0.92)
Predicted pKa with maximum SASA <sup>(3)</sup>	3.96(0.16)	3.7(0.22)	3.96(0.07)	3.9(0.1)	3.96(0.05)	6.30(0.37)

and His52, for which there is no experimental data. To see the effect of reducing the number of structures used in the modeling, the pKa values were re-calculated using only 15 snapshots with maximum and minimum SASA. As it can be seen, the only residue where pKa is affected by the accessible surface area selection is Glu19. For the minimum SASA set, the pKa value of this residue is perturbed from the model compound value by more than 0.5 units.

The thermal denaturation of OMTKY3 was done in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$ .<sup>45</sup> To consider the effect of  $\text{D}_2\text{O}$  on the pKas of unfolded state, we calculated the pKas of titratable residues using solvent dielectric constant of 74, 75 and 78.5 (Table S1). Moreover, to investigate the effect of protein (solute) dielectric constant on the pKa values in unfolded state, we performed calculation using protein dielectric constant of 4 as well (Table S2). Our results show that the effect of  $\text{D}_2\text{O}$  on pKa values of titratable residues in unfolded state is negligible (Table S1). Moreover, the results using dielectric constant of protein of 4 and 8 are practically the same. Therefore, the calculations of pKas in unfolded state are not sensitive to these parameters (Table S2).

### 3.2. Barnase

The thermal unfolding experiments of barnase suggested that pKa values of all acidic residues, Asp and Glu, on average are 0.4 lower than of the model component values.<sup>21</sup> Thus, there is no experimental data for individual pKas, but just a tendency that acid pKas are perturbed on average by 0.4 pH units. To see if our modeling can reproduce this trend, we calculated the individual pKa values for all Asp, Glu and His residues for each structure in unfolded ensemble (1000 structures) and then averaging over all pKas of Asp and Glu separately (Table 2). The calculations were done at ionic concentration of 0.05, solvent dielectric constant of 80 and protein (solute) dielectric constant of 8. The results show that pKas of all Asp and Glu are lower than the ones in model component (model component values for our calculations are Asp = 3.9, Glu = 4.0 and His = 6.5). Moreover, the averaged pKas of Asp and Glu are 0.27 and 0.37, respectively, lower than the model pKas, results which are in good agreement with the experimental observations. The pKa value of His18 is provided as well for future comparison.

Similarly, as above, we repeated the modeling using only 15 structures that have minimum or maximum of SASA out of 1000 snapshots (Table 2). The result of the structures with the minimum SASA shows that the average  $\Delta\text{pKa}$  values of Asp and Glu are 0.47 and 0.57, which are better predictions compared to the ones with the maximum SASA, 0.19 and 0.29. This speaks in favor of models that are more compact, allowing for residual interactions in unfolded state.

To test the sensitivity of results, with respect to the value of internal and external dielectric constants, we carried calculations with in solvent dielectric constant of 74, 75, 78.5 as well (Table S3). No significant effect was observed in the range of dielectric constant values. Similarly, repeating the modeling with internal dielectric constant of 4 had negligible effect of predicted pKas (Table S4).

Table 2. Predicted pKa values of Asp and Glu residues in unfolded state. The  $\Delta pK_a$  shift is ( $pK_a$  model compound — calculated  $pK_a$ ). (1), (2) and (3) are the average  $pK_a$  of all 1000 snapshots, 15 structures with minimum SASA and 15 structures with maximum SASA. The standard deviations are presented in parenthesis. The last two columns are the average  $pK_a$  values for Asp and Glu residues, respectively.

Unfolded state set	Asp8	Asp12	Asp22	Asp44	Asp54	Asp75	Asp86	Asp93	Asp101	Glu60	Glu73	His18	His102	Avg.-Asp	Avg.-Glu	
Predicted $pK_a^{(1)}$	3.81(0.33)	3.75(0.40)	3.58(0.34)	3.66(0.45)	3.62(0.42)	3.60(0.48)	3.51(0.4)	3.66(0.34)	3.41(0.43)	3.7(0.43)	3.54(0.47)	3.66(0.43)	6.48(0.59)	6.33(0.6)	3.62(0.13)	3.63(0.18)
$\Delta pK_a^{(1)}$ shift	0.08	0.14	0.31	0.23	0.27	0.29	0.38	0.23	0.48	0.29	0.45	0.33	0.01	0.16	0.28	0.37
Predicted $pK_a$ with minimum SASA <sup>(2)</sup>	3.60(0.6)	3.75(0.19)	3.43(0.59)	3.51(0.28)	3.47(0.87)	3.20(0.59)	3.47(0.34)	3.61(0.3)	2.98(1.07)	3.64(0.61)	3.06(1.10)	3.73(0.21)	6.37(0.57)	5.93(1.61)	3.45(0.17)	3.47(0.15)
$\Delta pK_a^{(2)}$ shift	0.29	0.14	0.46	0.38	0.42	0.69	0.42	0.28	0.91	0.35	0.94	0.26	0.12	0.56	0.45	0.53
Predicted $pK_a$ with maximum SASA <sup>(3)</sup>	3.83(0.24)	3.82(0.25)	3.56(0.34)	3.83(0.08)	3.66(0.33)	3.80(0.23)	3.59(0.31)	3.78(0.23)	3.35(0.32)	3.88(0.2)	3.69(0.20)	3.74(0.13)	6.48(0.3)	6.39(0.49)	3.69(0.06)	3.77(0.08)
$\Delta pK_a^{(3)}$ shift	0.06	0.07	0.33	0.06	0.23	0.09	0.30	0.11	0.54	0.11	0.3	0.25	0.01	0.10	0.21	0.23

Table 3. Predicted pKa shifts for Asp and Glu residues in unfolded state. The  $\Delta pK_a$  shift is ( $pK_a$  model compound — calculated  $pK_a$ ). (1), (2) and (3) are the average  $pK_a$  of all 1000 snapshots, 15 structures with minimum SASA and 15 structures with maximum SASA. The standard deviations are presented in parenthesis. The last two columns are the average  $pK_a$  values for Asp and Glu residues, respectively.

Unfolded state set	Asp23	Asp45	Glu04	Glu07	Glu14	Glu15	Glu26	Glu41	Glu52	Glu55	Avg.-Asp	Avg.-Glu	
Predicted $pK_a^{(1)}$	3.709(0.25)	3.448(0.36)	3.78(0.28)	3.83(0.42)	3.79(0.41)	3.74(0.55)	3.73(0.37)	3.73(0.42)	3.62(0.37)	3.73(0.39)	3.55(0.39)	3.59(0.12)	3.72(0.2)
$\Delta pK_a^{(1)}$ shift	0.19	0.42	0.22	0.17	0.21	0.26	0.26	0.27	0.27	0.38	0.45	0.31	0.28
Predicted $pK_a$ with minimum SASA <sup>(2)</sup>	3.77(0.18)	3.49(0.42)	3.6(0.38)	3.547(0.86)	3.812(0.30)	3.698(0.91)	3.348(0.87)	3.755(0.51)	3.64(0.36)	3.359(0.56)	3.63(0.17)	3.59(0.08)	
$\Delta pK_a^{(2)}$ shift	0.13	0.41	0.4	0.453	0.188	0.3	0.65	0.24	0.36	0.64	0.27	0.41	
Predicted $pK_a$ with maximum SASA <sup>(3)</sup>	3.64(0.27)	3.39(0.28)	3.81(0.25)	3.92(0.10)	3.83(0.16)	3.82(0.21)	3.83(0.13)	3.8(0.17)	3.65(0.23)	3.61(0.27)	3.51(0.09)	3.78(0.11)	
$\Delta pK_a^{(3)}$ shift	0.26	0.51	0.19	0.08	0.17	0.18	0.17	0.2	0.35	0.39	0.39	0.22	

### 3.3. CI2

Thermal denaturation experiments of CI2 at ionic concentration of 50 mM demonstrated that the pKa values of acidic residues, Asp and Glu, in the unfolded state are on average 0.3 unit lower than their pKa values of the model compound.<sup>22</sup> Thus, we calculated the pKa values of all titratable residues, no histidine residue present in CI2, using 1000 snapshots of unfolded CI2. The DelPhiPka calculations were done at ionic concentration of 0.05 M, solvent dielectric constant of 80 and protein (solute) dielectric constant of 8. Additionally, we carried out pKas calculations using only for 15 snapshots with maximum and minimum SASA (Table 3).

We observe that all the  $\Delta pK_a$  shifts are positive, which means that the predicted pKa values are lower than the model pKas. Moreover, the average of  $\Delta pK_a$  value for Asp is 0.308 and for Glu is 0.22, which are close to the experimental values. Furthermore, we observe that the average  $\Delta pK_a$  of Asp and Glu are 0.329 and 0.37, respectively, for snapshots with minimum SASA. On the other hand, the predicted values for snapshots with maximum SASA are in less agreement with the experiment (Table 3). This again indicates that compact structures are better representatives for unfolded state.

As above, we do not observe significant effect of the internal or external dielectric constant values (Tables S5 and S6).

### 3.4. PGB1

The pH-dependent stability of PGB1 with the mutations T2Q, N8D, and N37D (PGB1-QDD) was determined by measuring the pKa values for unfolded state under native conditions.<sup>23</sup> The pKas values of acidic groups were measured in the presence of H<sub>2</sub>O/D<sub>2</sub>O at 25°C with the counterion concentration about 1 mM (Table 4).

We calculated the pKa values of the corresponding titratable residues (note that no histidine residue is present in PGB1) in the unfolded structures for solvent dielectric constant of 80 and protein dielectric constants of 8 (Table 4). We calculated the pKa values for 15 structures while having the minimum and maximum SASA as well. No clear trend is observed in this case to indicate what model of unfolded state is better for the pKas calculations.

Also, the predicted results for solvent dielectric constants of 74, 75 and 78.5 and the protein (solute) dielectric constants of 4 are shown in Tables S7 and S8. No significant impact of the dielectric constant values was found.

### 3.5. HEWL

The pKa values of several acidic residues in thermally unfolded state of HEWL for different pH and 50 mM Britton–Robinson buffer were reported.<sup>42</sup> For some of these residues, significant shifts from their model pKas were observed.<sup>42</sup>

We calculated the pKa of those corresponding acidic residues using salt concentration of 144 mM for 1000 snapshots of unfolded structure of HEWL with solvent

Table 4. Predicted pKa values of acidic residues in unfolded state of PGB1-QDD. The  $\Delta$ pKa shift is (pKa model compound — calculated pKa). (1), (2) and (3) are the average pKa of all 1000 snapshots, 15 structures with minimum SASA and 15 structures with maximum SASA. The standard deviations are presented in parenthesis.

Unfolded state set	ASP08	ASP22	ASP36	ASP37	ASP40	ASP46	ASP47	GLU15	GLU19	GLU27	GLU42	GLU56
Experimental pKa	3.72	3.87	4.11	4.2	4.24	4.1	4.03	4.41	4.46	4.47	4.9	4.83
Experimental $\Delta$ pKa	0.28	0.13	-0.11	-0.2	-0.24	-0.1	-0.03	0.09	0.04	0.03	-0.4	-0.33
Predicted pKa <sup>(1)</sup>	3.485(0.42)	3.836(0.47)	3.835(0.56)	3.982(0.51)	4.058(0.44)	3.929(0.45)	3.905(0.56)	3.675(0.52)	3.909(0.44)	3.852(0.39)	4.208(0.45)	3.737(0.24)
$\Delta$ pKa <sup>(1)</sup> shift	0.414	0.063	0.064	-0.082	-0.158	-0.029	-0.005	0.324	0.09	0.147	-0.208	0.262
Predicted pKa with minimum SASA <sup>(2)</sup>	3.552(0.56)	3.942(0.57)	3.796(0.79)	3.458(1.12)	3.71(1.19)	3.829(0.68)	4.04(0.46)	3.436(1.08)	3.891(0.44)	3.349(1.29)	4.255(0.51)	3.573(0.47)
$\Delta$ pKa <sup>(2)</sup> shift	0.347	-0.042	0.1	0.44	0.19	0.07	-0.14	0.56	0.1	0.65	-0.255	0.42
Predicted pKa with maximum SASA <sup>(3)</sup>	3.553(0.38)	3.924(0.20)	3.73(0.47)	4.07(0.24)	4.18(0.28)	4.02(0.31)	3.99(0.25)	3.81(0.16)	3.949(0.09)	3.82(0.2)	4.18(0.12)	3.744(0.05)
$\Delta$ pKa <sup>(3)</sup> shift	0.367	-0.024	0.162	-0.171	-0.28	-0.126	-0.09	0.189	0.05	0.17	-0.18	0.255

Table 5. Predicted pKa values of titratable residues in unfolded state of HEWL. (1), (2) and (3) are the average pKa of all 1000 snapshots and 15 structures with minimum SASA, 15 structures with maximum SASA. The standard deviations are presented in parenthesis.

Unfolded state set	Asp18	Asp48	Asp52	Asp66	Asp87	Asp101	Asp119	Glu7	Glu35	His15
Experimental pKa	3.4(0.2)	2.9(0.3)	4(0.2)	3.1(0.1)	—	4.4(0.2)	3.6(0.2)	—	5(0.2)	—
Predicted pKa <sup>(1)</sup>	3.48(0.41)	3.76(0.33)	3.81(0.30)	3.63(0.46)	3.82(0.29)	3.69(0.31)	3.6(0.37)	3.73(0.41)	3.77(0.34)	6.21(0.58)
Predicted pKa with minimum SASA <sup>(2)</sup>	3.41(0.57)	3.71(0.24)	3.83(0.23)	3.29(1.0)	3.71(0.38)	3.75(0.29)	3.26(0.45)	3.79(0.23)	3.79(0.19)	6.33(0.39)
Predicted pKa with maximum SASA <sup>(3)</sup>	3.54(0.25)	3.77(0.16)	3.82(0.22)	3.79(0.17)	3.83(0.17)	3.85(0.22)	3.72(0.21)	3.87(0.07)	3.79(0.27)	6.4(0.22)

and protein dielectric constants of 80 and 8, respectively (Table 5). There is good correlation between the experimental and predicted pKa values for Asp18, Asp66, Asp101, Asp119, Asp35, Glu35 except Asp48 (Fig. S2). Moreover, we calculated the pKa values only for 15 snapshots with maximum and minimum SASA. The better correlation between the experimental estimated pKa values with the calculated ones is seen for the snapshots with the minimum SASA.

Moreover, the predicted results for solvent dielectric constants of 74, 75 and 78.5 and the protein (solute) dielectric constants of 4 are shown in Tables S9 and S10. Again, there is no significant difference observed for changes in these parameters.

#### 4. Conclusion

The investigation showed that unfolded state can be modeled with the coil model, algorithm such as “flexible-meccano”, if one is concerned about pKa values of acidic groups. Even more, the number of snapshots representing are not many and one can be successful using only 15 structures with minimum SASA. This reduces the computational cost of modeling.

Apart from testing the possibility to model unfolded state to compute pK<sub>a</sub>s of titratable groups, we also delivered predictions for pK<sub>a</sub>s that are not experimentally measured yet.

While we attempted to justify the usage of “flexible-meccano” for modeling pK<sub>a</sub>s in unfolded state, it is tempting to generalize the observation and to consider that generated ensemble can be used for any other type of modeling. Thus, we speculate that snapshots can be used to model unfolded state in cases of computing folding free energy and folding free energy caused by mutations.

#### Acknowledgments

The work was supported by a grant from NIH, Grant Number R01GM093937. We thank Morgan Willard for proofreading the paper.

#### References

1. Cho JH, Sato S, Raleigh DP, Thermodynamics and kinetics of non-native interactions in protein folding: A single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state, *J Mol Biol* **338**:827–837, 2004.
2. Pace CN, Laurents DV, Erickson RE, Urea denaturation of barnase: pH dependence and characterization of the unfolded state, *Biochemistry* **31**:2728–2734, 1992.
3. Robic S *et al.*, Role of residual structure in the unfolded state of a thermophilic protein, *Proc Natl Acad Sci USA* **100**:11345–11349, 2003.
4. Religa TL *et al.*, Solution structure of a protein denatured state and folding intermediate, *Nature* **437**:1053–1056, 2005.
5. Bashford D, Macroscopic electrostatic models for protonation states in proteins, *Front Biosci* **9**:1082–1099, 2004.

6. Cherny VV *et al.*, Properties of single voltage-gated proton channels in human eosinophils estimated by noise analysis and by direct measurement, *J Gen Physiol* **121**:615–628, 2003.
7. Warshel A, Russell ST, Calculations of electrostatic interactions in biological systems and in solutions, *Q Rev Biophys* **17**:283–422, 1984.
8. Mitra RC, Zhang Z, Alexov E, In silico modeling of pH-optimum of protein-protein binding, *Proteins* **79**:925–936, 2011.
9. Talley K, Alexov E, On the pH-optimum of activity and stability of proteins, *Proteins* **78**:2699–2706, 2010.
10. Isvoran A, Craescu CT, Alexov E, Electrostatic control of the overall shape of calmodulin: Numerical calculations, *Eur Biophys J* **36**:225–237, 2007.
11. Shosheva A *et al.*, Experimental and numerical study of the poplar plastocyanin isoforms using Tyr as a probe for electrostatic similarity and dissimilarity, *Biochim Biophys Acta* **1698**:67–75, 2004.
12. Alexov E, Numerical calculations of the pH of maximal protein stability. The effect of the sequence composition and three-dimensional structure, *Eur J Biochem* **271**:173–185, 2004.
13. Tajielyato N, Alexov E, Processivity vs. beating: Comparing cytoplasmic and axonemal dynein microtubule binding domain association with microtubule, *Int J Mol Sci* **20**:2019.
14. Elcock AH, Realistic modeling of the denatured states of proteins allows accurate calculations of the pH dependence of protein stability, *J Mol Biol* **294**:1051–1062, 1999.
15. Zhou HX, Residual charge interactions in unfolded staphylococcal nuclease can be explained by the Gaussian-chain model, *Biophys J* **83**:2981–2986, 2002.
16. Zhou HX, A Gaussian-chain model for treating residual charge-charge interactions in the unfolded state of proteins, *Proc Natl Acad Sci USA* **99**:3569–3574, 2002.
17. Zhou HX, Direct test of the Gaussian-chain model for treating residual charge-charge interactions in the unfolded state of proteins, *J Am Chem Soc* **125**:2060–2061, 2003.
18. Cho JH, Raleigh DP, Electrostatic interactions in the denatured state and in the transition state for protein folding: Effects of denatured state interactions on the analysis of transition state structure, *J Mol Biol* **359**:1437–1446, 2006.
19. Trefethen JM *et al.*, Charge-charge interactions in the denatured state influence the folding kinetics of ribonuclease Sa, *Protein Sci* **14**:1934–1938, 2005.
20. Guzman-Casado M *et al.*, Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of *Thermus thermophilus* ribonuclease H, *J Mol Biol* **329**:731–743, 2003.
21. Oliveberg M, Arcus VL, Fersht AR, pKA values of carboxyl groups in the native and denatured states of barnase: The pKA values of the denatured state are on average 0.4 units lower than those of model compounds, *Biochemistry* **34**:9424–9433, 1995.
22. Tan YJ *et al.*, Perturbed pKA-values in the denatured states of proteins, *J Mol Biol* **254**:980–992, 1995.
23. Lindman S *et al.*, pK(a) values for the unfolded state under native conditions explain the pH-dependent stability of PGB1, *Biophys J* **99**:3365–3373, 2010.
24. Meng W, Raleigh DP, Analysis of electrostatic interactions in the denatured state ensemble of the N-terminal domain of L9 under native conditions, *Proteins* **79**:3500–3510, 2011.
25. Swint-Kruse L, Robertson AD, Hydrogen bonds and the pH dependence of ovomucoid third domain stability, *Biochemistry* **34**:4724–4732, 1995.
26. Piana S *et al.*, Water dispersion interactions strongly influence simulated structural properties of disordered protein states, *J Phys Chem B* **119**:5113–5123, 2015.

27. Smith LJ *et al.*, Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations, *J Mol Biol* **255**:494–506, 1996.
28. Bernado P, Svergun DI, Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering, *Mol Biosyst* **8**:151–167, 2012.
29. Feldman HJ, Hogue CW, A fast method to sample real protein conformational space, *Proteins* **39**:112–131, 2000.
30. Ozenne V *et al.*, Flexible-meccano: A tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables, *Bioinformatics* **28**:1463–1470, 2012.
31. Bernado P *et al.*, A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering, *Proc Natl Acad Sci USA* **102**:17002–17007, 2005.
32. Kucukkal TG *et al.*, Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins, *Curr Opin Struct Biol* **32**:18–24, 2015.
33. Kucukkal TG *et al.*, Computational and experimental approaches to reveal the effects of single nucleotide polymorphisms with respect to disease diagnostics, *Int J Mol Sci* **15**:9670–9717, 2014.
34. Peng Y, Alexov E, Basu S, Structural perspective on revealing and altering molecular functions of genetic variants linked with diseases, *Int J Mol Sci* **20**:2019.
35. Steff S *et al.*, Molecular mechanisms of disease-causing missense mutations, *J Mol Biol* **425**:3919–3936, 2013.
36. Getov I, Petukh M, Alexov E, SAAFEC: Predicting the effect of single point mutations on protein folding free energy using a knowledge-modified MM/PBSA approach, *Int J Mol Sci* **17**:512, 2016.
37. Dehouck Y *et al.*, PoPMuSiC 2.1: A web server for the estimation of protein stability changes upon mutation and sequence optimality, *BMC Bioinf* **12**:151, 2011.
38. Gilis D, Rooman M, PoPMuSiC, an algorithm for predicting protein mutant stability changes: Application to prion proteins, *Protein Eng* **13**:849–856, 2000.
39. Kwasigroch JM *et al.*, PoPMuSiC, rationally designing point mutations in protein structures, *Bioinformatics* **18**:1701–1702, 2002.
40. Wang L, Li L, Alexov E, pKa predictions for proteins, RNAs, and DNAs with the Gaussian dielectric function using DelPhi pKa, *Proteins* **83**:2186–2197, 2015.
41. Wang L, Zhang M, Alexov E, DelPhiPKa web server: Predicting pKa of proteins, RNAs and DNAs, *Bioinformatics* **32**:614–615, 2016.
42. Bradley J *et al.*, Highly perturbed pKa values in the unfolded state of hen egg white lysozyme, *Biophys J* **102**:1636–1645, 2012.
43. Vidulich RN and Wilson DJ, The environmental setting as a factor in social influence, *J Soc Psychol* **71**:247–255, 1967.
44. Holm L, Sander C, Parser for protein folding units, *Proteins* **19**:256–268, 1994.
45. Humphrey W, Dalke A, Schulten K, VMD: Visual molecular dynamics, *J Mol Graph* **14**:33–38, 27–28 1996.