Supporting Information For

Insights into ionic liquid-enhanced membrane protein stability through machine learning and molecular simulation

Ju Liu[a], Guiming Zhang[bc], Cheng Song[d], Yanlei Wang[b], Jing Ren[e]*, and Hongyan He[a]*

[a] *Center of Ionic Liquids and Green Energy, Beijing Key Laboratory of Solid State Battery and Energy Storage Process, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China*

[b]*School of Chemistry and Life Resources, Renmin University of China, Beijing, 100872, China*

[c]*College of Chemistry and Life Science, Beijing University of Technology, 100124, China*

[d]*School of Population and Health, Renmin University of China, Beijing, 100872, China*

[e]*Department of Plastic and Reconstructive Surgery, the First Medical Center, Chinese PLA General Hospital, Beijing 100853, China*

Email: ylwang17@ruc.edu.cn; and eggy119@163.com

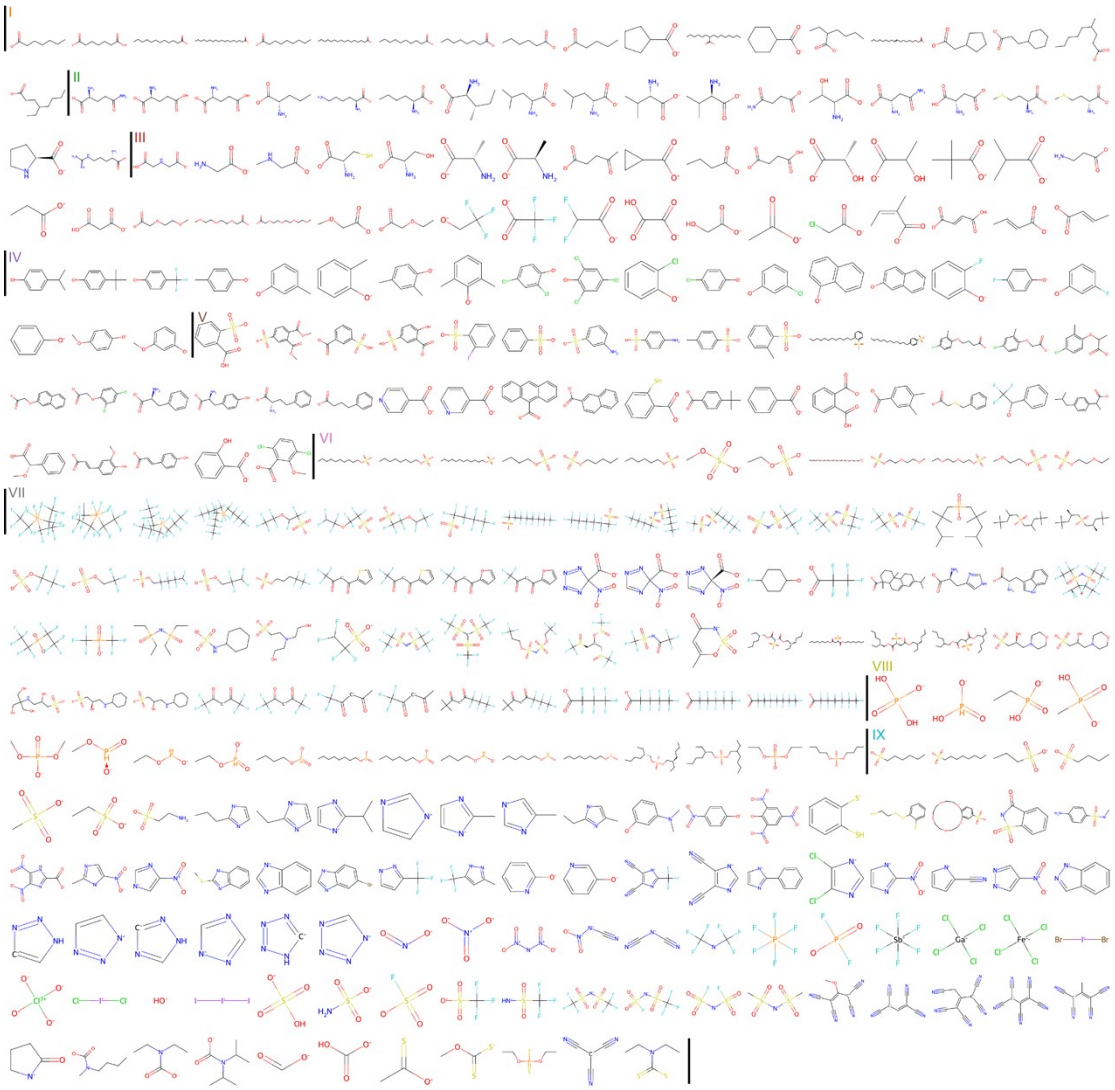

Fig. S1 The anions are arranged according to hierarchical clustering analysis.

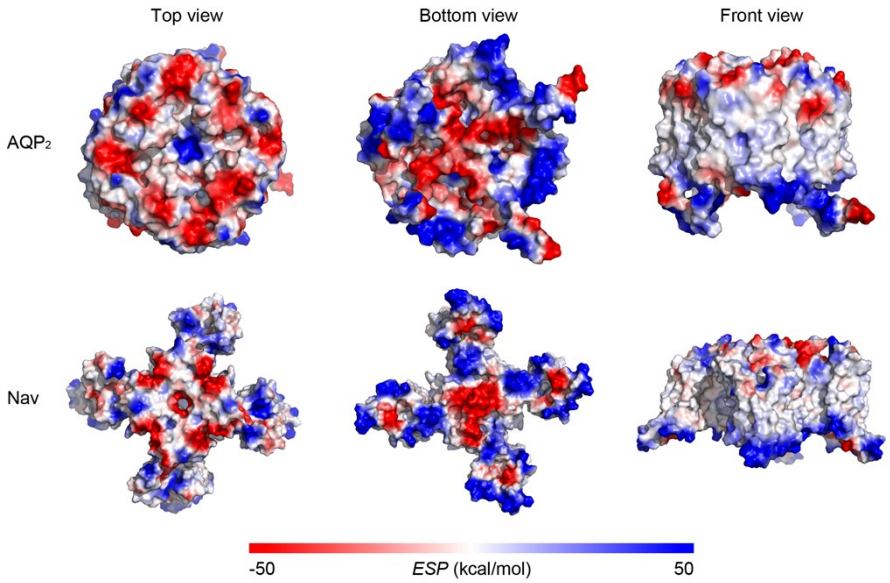

Fig. S2 Electrostatic potential distribution of the AQP_2 and Nav.

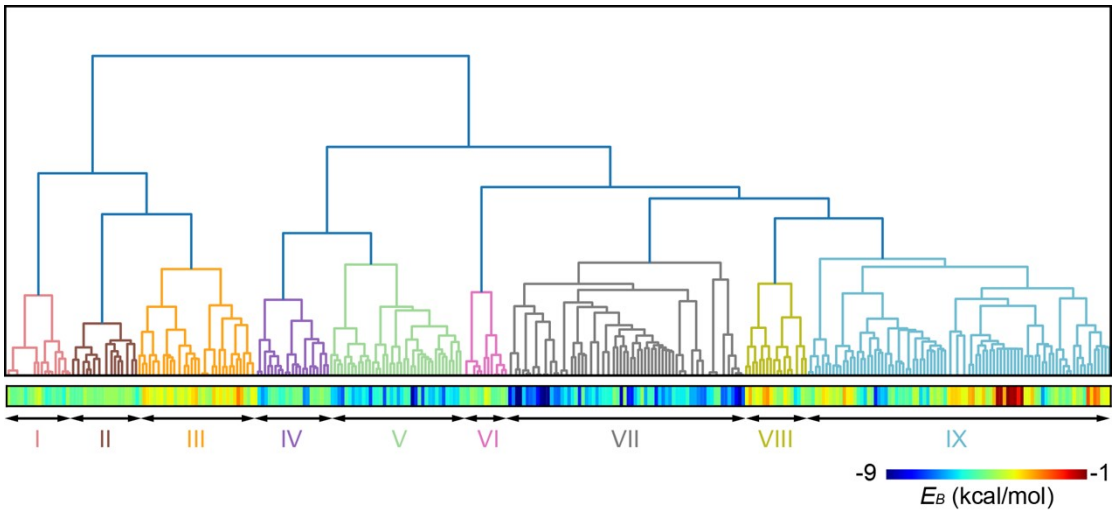

Fig. S3 Hierarchical clustering dendrogram of anions, and binding energy between Nav and anions.

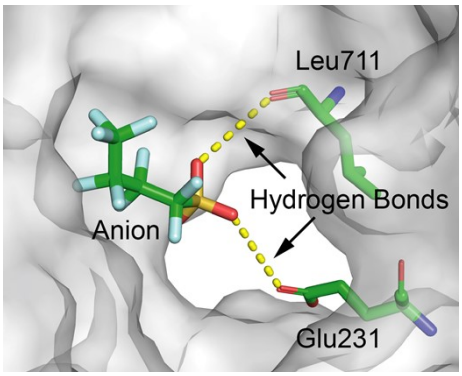

Fig. S4 The representative anion of cluster VII formed hydrogen bonds with AQP_2 protein at the binding site.

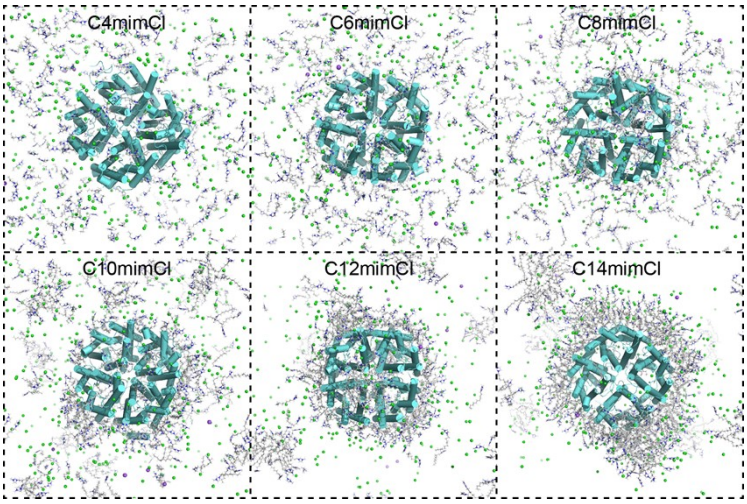

Fig. S5 Screenshot of simulations for [$C_n$mim]Cl with different chain length.

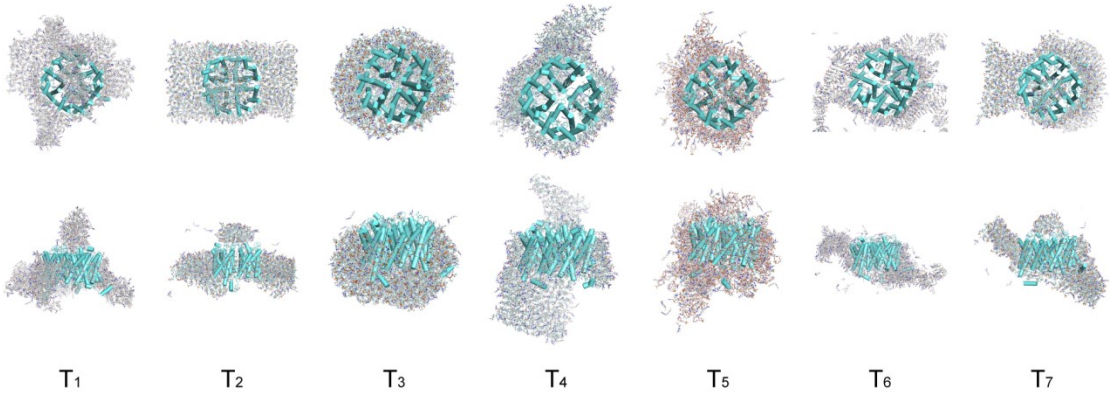

Fig. S6 Screenshot of simulations for the top 7 anions with the highest E_B

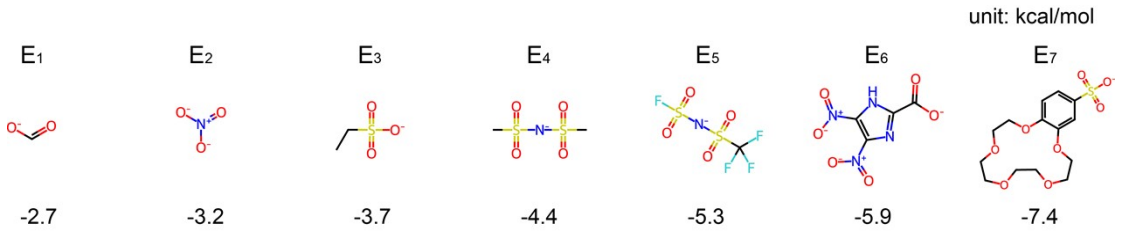

Fig. S7 Structures of the 7 anions with different E_B from same cluster.

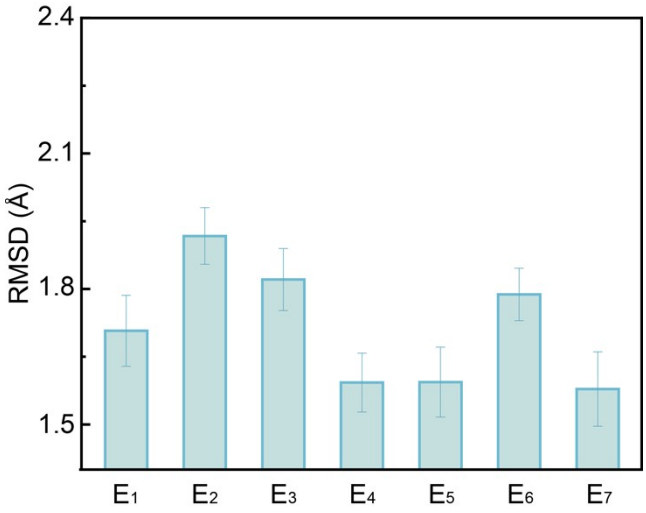

Fig. S8 The average values of RMSD of the protein in aqueous solutions containing $[C_6mim]^+$ based ILs with different anions, derived from the IX cluster with different E_B.

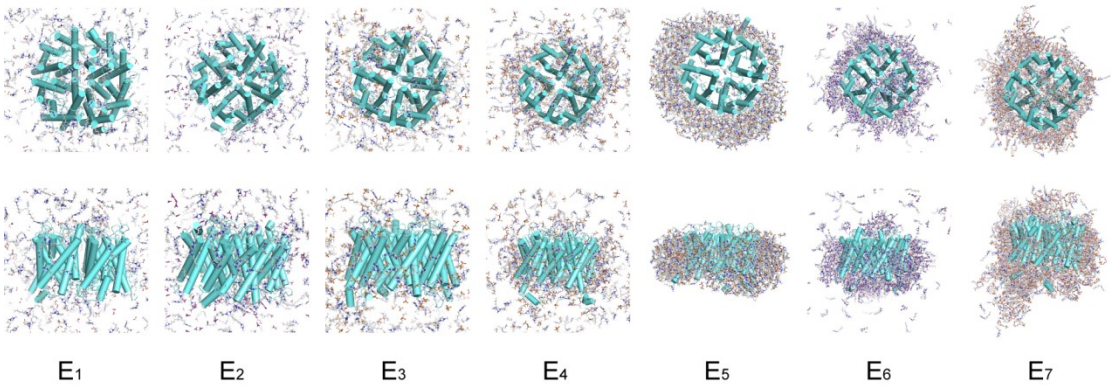

Fig. S9 Screenshots of simulations for the 7 anions from the same cluster but with different E_B

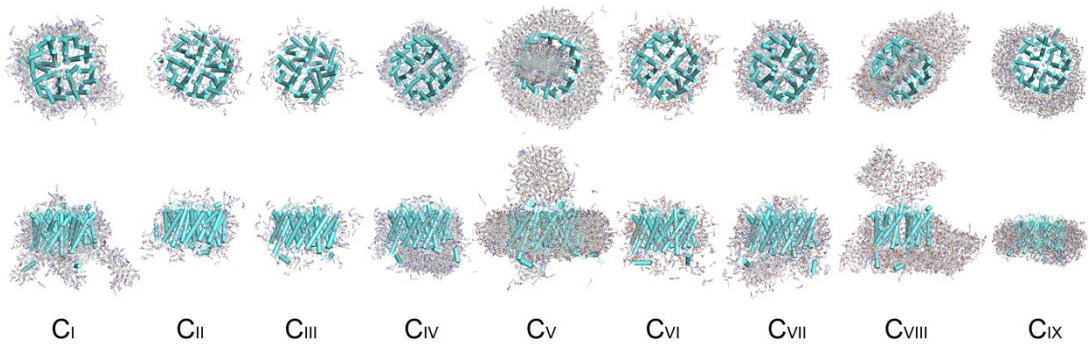

Fig. S10 Screenshots of simulations for the 9 anions with the same E_B but from different clusters

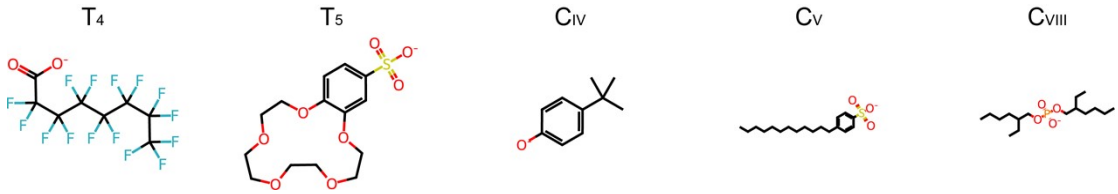

Fig. S11 Structures of the five screened ILs with the lowest RMSD.

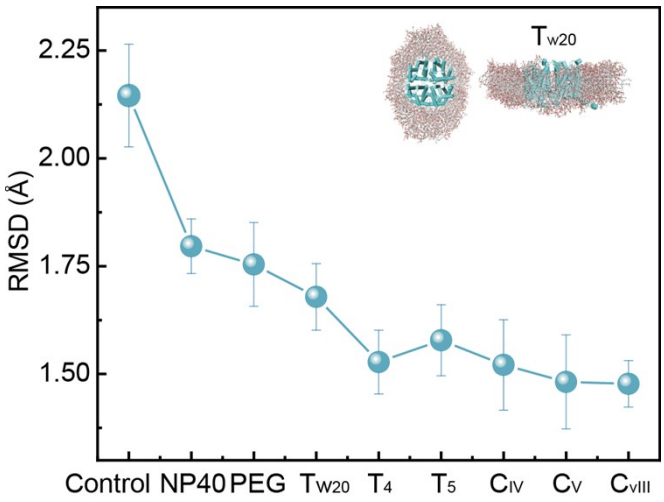

Fig. S12 The five screened ILs were compared with conventional protein stabilizers regarding their impact on stability of AQP_2. The inset plots represent the simulation screenshots depicting the adsorption states of Tw20 on the protein surface.

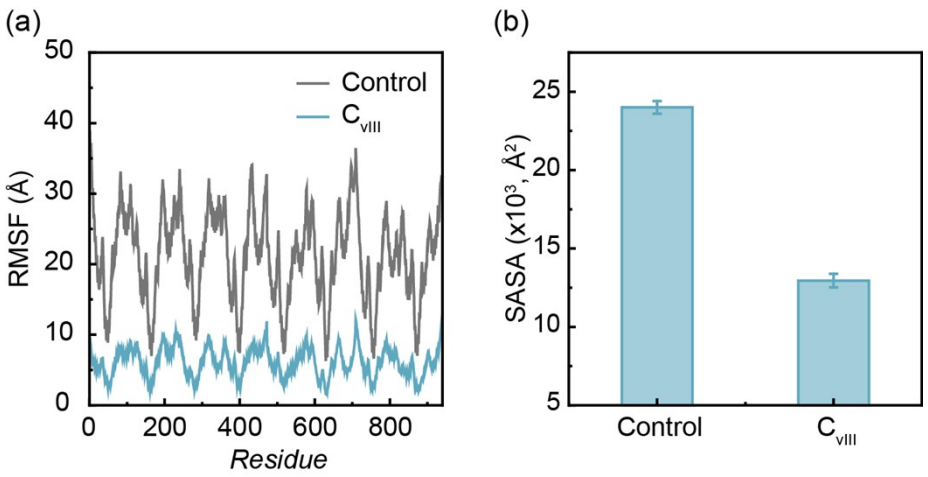

Fig. S13 (a) Root mean square fluctuation (RMSF) and (b)solvent accessible surface area (SASA) of hydrophobic core of protein in the control group and CIII system.

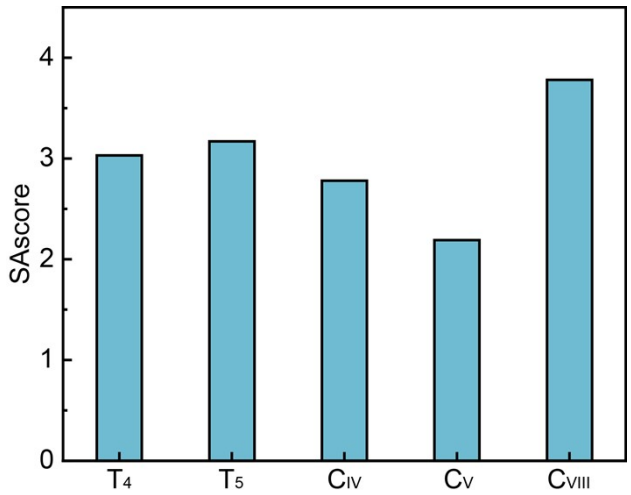

Fig. S14 The synthetic accessibility score (SAscore) assesses the ease of synthesis for five screened ILs. The SAscore ranges from 1 (easy to synthesize) to 10 (very difficult to synthesize).

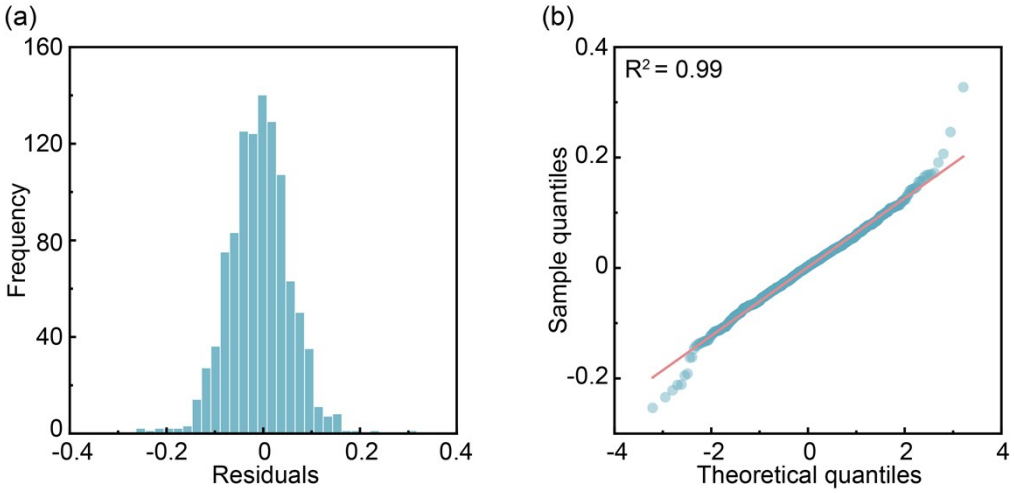

Fig. S15 (a) The histogram and (b) quantile-quantile plot illustrate the residuals of the model. The histogram shows that the residuals are symmetrically distributed around zero, and most points in the quantile-quantile plot are close to a 45° diagonal line, indicating that the residuals are approximately normally distributed. Therefore, there is no need to apply any transformations to the original data.