# Supporting Information

# Assessing Lysine and Cysteine Reactivities for Designing Targeted Covalent Kinase Inhibitors

Ruibin Liu,[†] Zhi Yue,[†,‡] Cheng-Chieh Tsai,[†] and Jana Shen[*,†]

† *Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD 21201*

‡ *Current address: Department of Chemistry, University of Chicago, Chicago, IL 60637*

E-mail: jana.shen@rx.umaryland.edu

## Supplementary Tables

Table S1: Calculated p$K_a$'s of the catalytic lysines as well as reactive lysines and cysteines in the kinase dataset studied by continuous CpHMD

| PDB ID | Kinase name | Kinase group | Inhibitor type[a] | p$K_a$ catalytic Lys[b] | p$K_a$ other Lys[b] | p$K_a$ reactive Cys[c] |
|---|---|---|---|---|---|---|
| 2G2F | ABL1 | TK | II | 8.5 (K271) | - | 8.0 (C305) |
| 5U8L | EGFR | TK | II | 7.3 (K745) | 8.2 (K852) | 8.1 (C781) |
| 5UG8 | EGFR(mt) | TK | II | 7.1 (K745) | - | 8.1 (C781) |
| 4MXC | MET | TK | II | 7.7 (K1110) | - | - |
| 3V5Q | NTRK3 | TK | II | 9.7 (K572) | - | - |
| 5K9I | c-SRC | TK | II | 8.1 (K295) | - | 6.8 (C277) |
| 4TPT | LIMK2 | TKL | III | 9.1 (K360) | - | 5.8 (C365) |
| 4ITJ | RIPK1 | TKL | III | 6.7 (K45) | - | 8.3 (C53) |
| 4USD | LOK | STE | II | 9.0 (K65) | - | 7.6 (C206) |
| 4ZLO | PAK1 | STE | II | 9.7 (K299) | - | 5.0 (C411)[d] |
| 3NAX | PDK1 | AGC | II | 7.0 (K111) | - | 6.0 (C260) |
| 2YZA | AMPKa2 | CAMK | II | 8.3 (K45) | - | - |
| 1G3N | CDK6 | CMGC | None | 6.1 (K43) | 6.0 (K147) | 7.9 (C15) |
| 4JAI | Aurora | Other | II | 7.9 (K162) | - | 5.4 (C290)[d] |
| 4YZ9 | IRE1$\alpha$ | Other | II | 8.8 (K599) | - | 5.8 (C715)<br>7.9 (C747)[d] |
| 2XNM | NEK2 | Other | II | 7.8 (K37) | - | - |
| 4X7N | PEK | Other | II | 7.6 (K622) | 7.3 (K151) | 5.2 (C261) |
| 6FDZ | ULK3 | Other | II | 9.2 (K44) | - | - |

[a] Inhibitor type is defined in the main text. Inhibitors were removed in the simulations.

[b] The first 0.2 ns (per replica) was discarded in the lysine p$K_a$ calculations.

[c] For the cysteine p$K_a$ calculations, the first 1 ns (per replica) was discarded due to the need for longer equilibration time and slower convergence. We note, since the cysteine p$K_a$ obtained from the second half of the simulation was always lower than that from the first half, the conclusion regarding reactivity was not affected. A large-scale validation study for cysteine p$K_a$ calculations will be conducted in the future.

[d] These residues were missing in the crystal structures and their positions were built using SWISS-MODEL.[1]

Table S2: Calculated and experimental cysteine and lysine p$K_a$'s in the validation systems

| Protein | PDB | Residue | Expt[a] | CpHMD | Propka[b] |
|---|---|---|---|---|---|
| Kinase | 1I0E | C283 | 5.6 | 5.5 | 10.4 |
| V74K | 3RUZ | K74 | 7.4 | 6.8 | 9.1 |
| V99K | 4HMI | K99 | 6.5 | 5.8 | 7.5 |
| L125K | 3C1E | K125 | 6.2 | 5.6 | 7.3 |

[a] The experimental p$K_a$ of Cys283 in creatine kinase is taken from ref.[2] The p$K_a$'s of lysines in the SNase mutants are taken from ref.[3]

[b] Propka3.1[4] was used.

Table S3: Calculated p$K_a$'s for lysines in the V74K mutant of SNase and cysteine in creatine kinase

| SNase (PDB: 3RUZ) | | Creatine kinase (PDB: 1I0E) | |
|---|---|---|---|
| Residue | p$K_a$ | Residue | p$K_a$ |
| K9 | 11.5 | C74 | $> 11$ |
| K16 | 10.1 | C146 | 10.8 |
| K24 | 8.7 | C254 | $> 11$ |
| K28 | 11.1 | C283 | 5.5 |
| K53 | 10.8 | | |
| K63 | 10.7 | | |
| K64 | 11.0 | | |
| K70 | 10.9 | | |
| K71 | 10.7 | | |
| K74 | 7.2 | | |
| K78 | 10.5 | | |
| K84 | 10.6 | | |
| K97 | 10.9 | | |
| K110 | 10.9 | | |
| K116 | 10.4 | | |
| K127 | 10.6 | | |
| K133 | 11.3 | | |
| K134 | 10.6 | | |
| K136 | 11.2 | | |

# Supplementary Figures

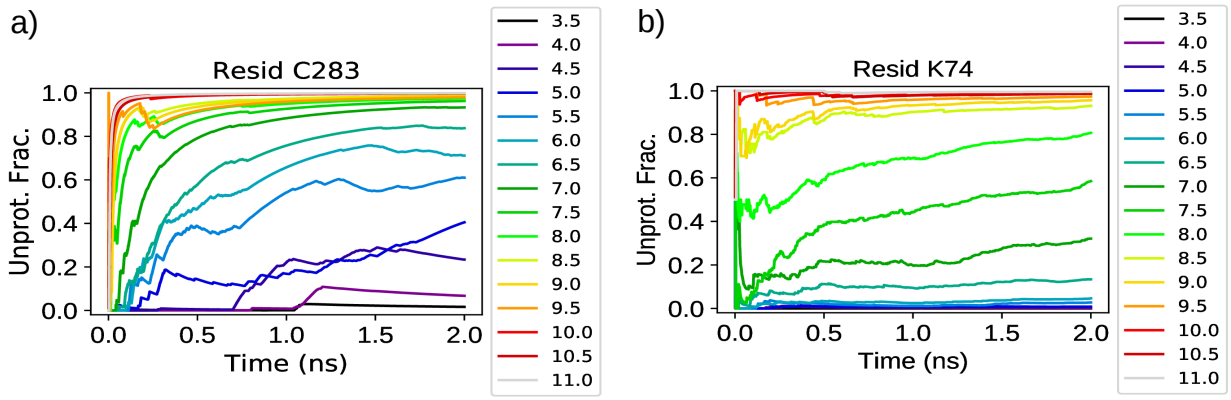


Figure S1: **Data from the CpHMD simulations of creatine kinase and V74K SNase.** a) Cumulatively calculated unprotonated fraction versus time at different pH for Cys283 in creatine kinase (PDB: 1I0E). b) Cumulatively calculated unprotonated fraction versus time at different pH for Lys74 in V74K SNase (PDB: 3RUZ).

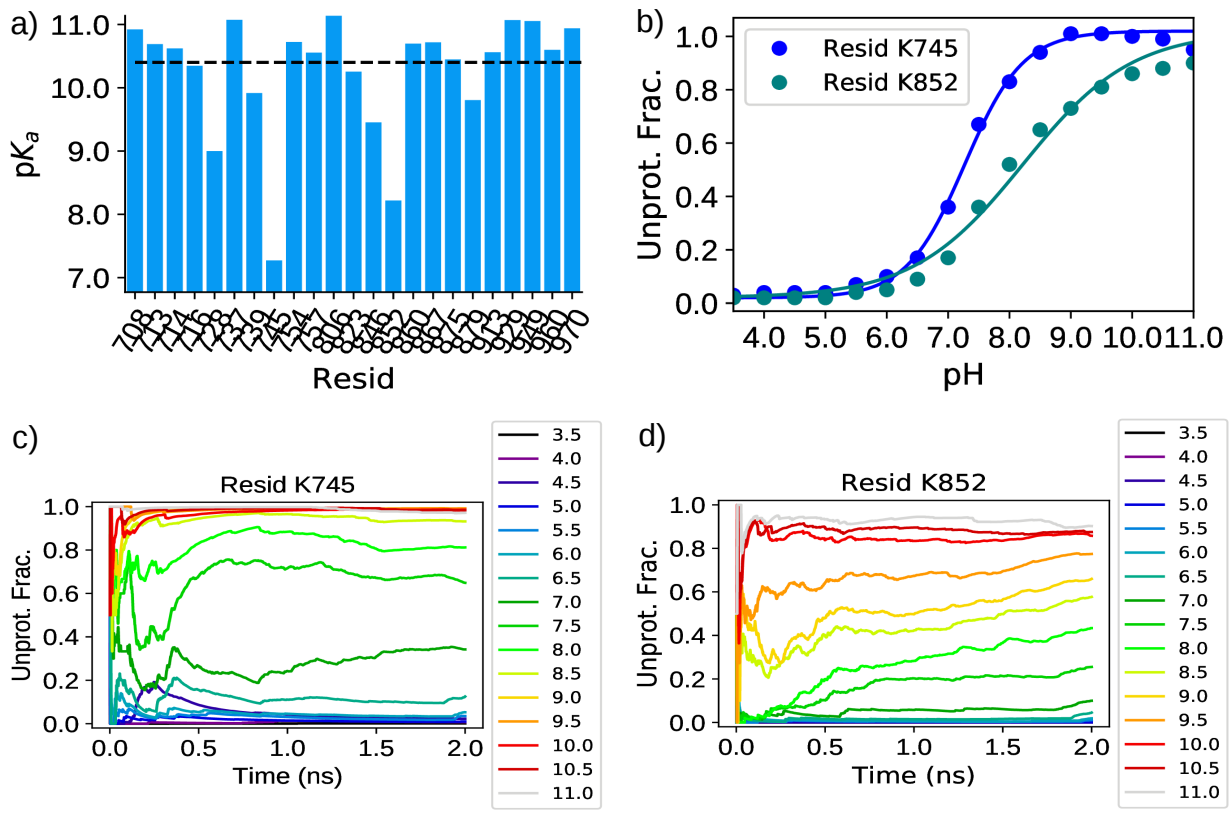


Figure S2: **Data from the CpHMD simulations of the wild-type EGFR kinase (PDB: 5U8L)**. a) Calculated p$K_a$'s for all lysines. The dashed line represents the model p$K_a$ of lysine. b) Titration plots (unprotonated fractions at different pH) for the reactive lysines. c) and d) Cumulatively calculated unprotonated fraction versus time at different pH for the reactive lysines.

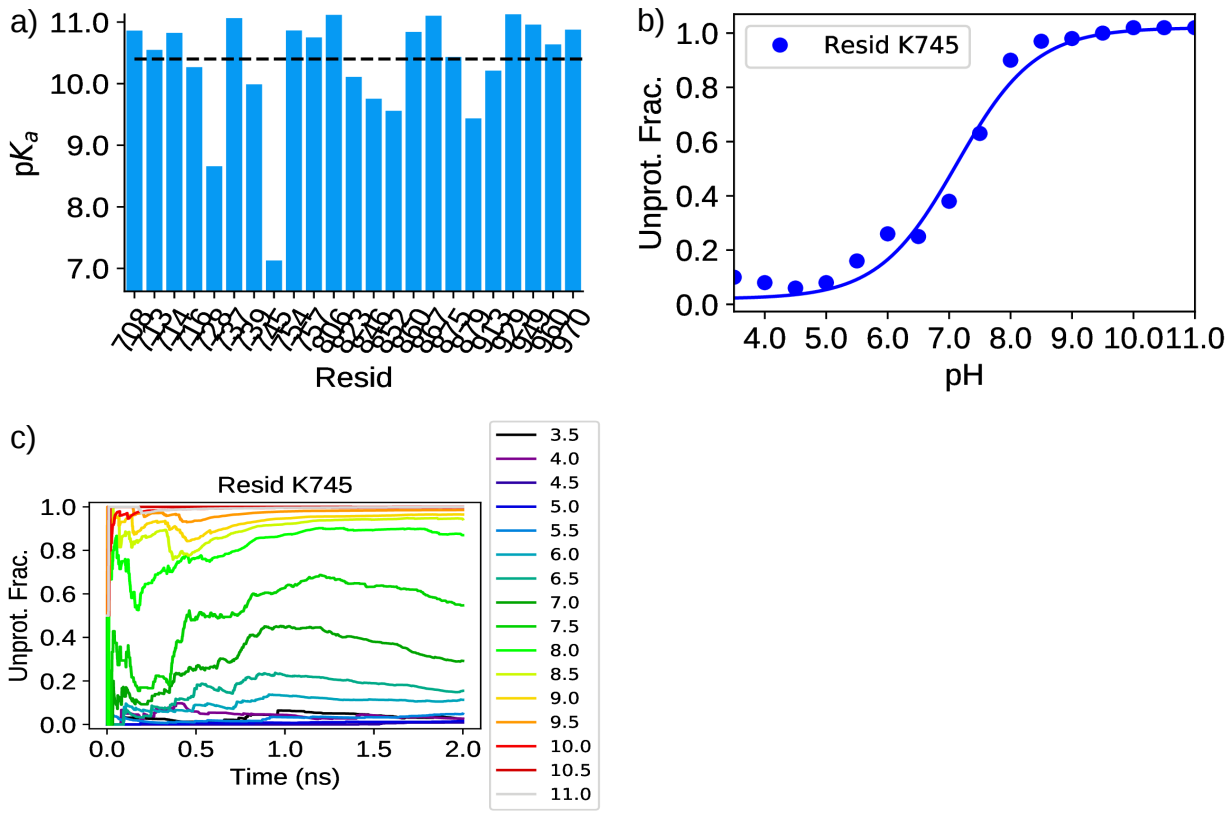


Figure S3: **Data from the CpHMD simulations of the mutant EGFR kinase (PDB: 5UG8).** a) Calculated p$K_a$'s for all lysines. The dashed line represents the model p$K_a$ of lysine. b) Titration plots (unprotonated fractions at different pH) for the reactive lysines. c) Cumulatively calculated unprotonated fraction versus time at different pH for the reactive lysine.

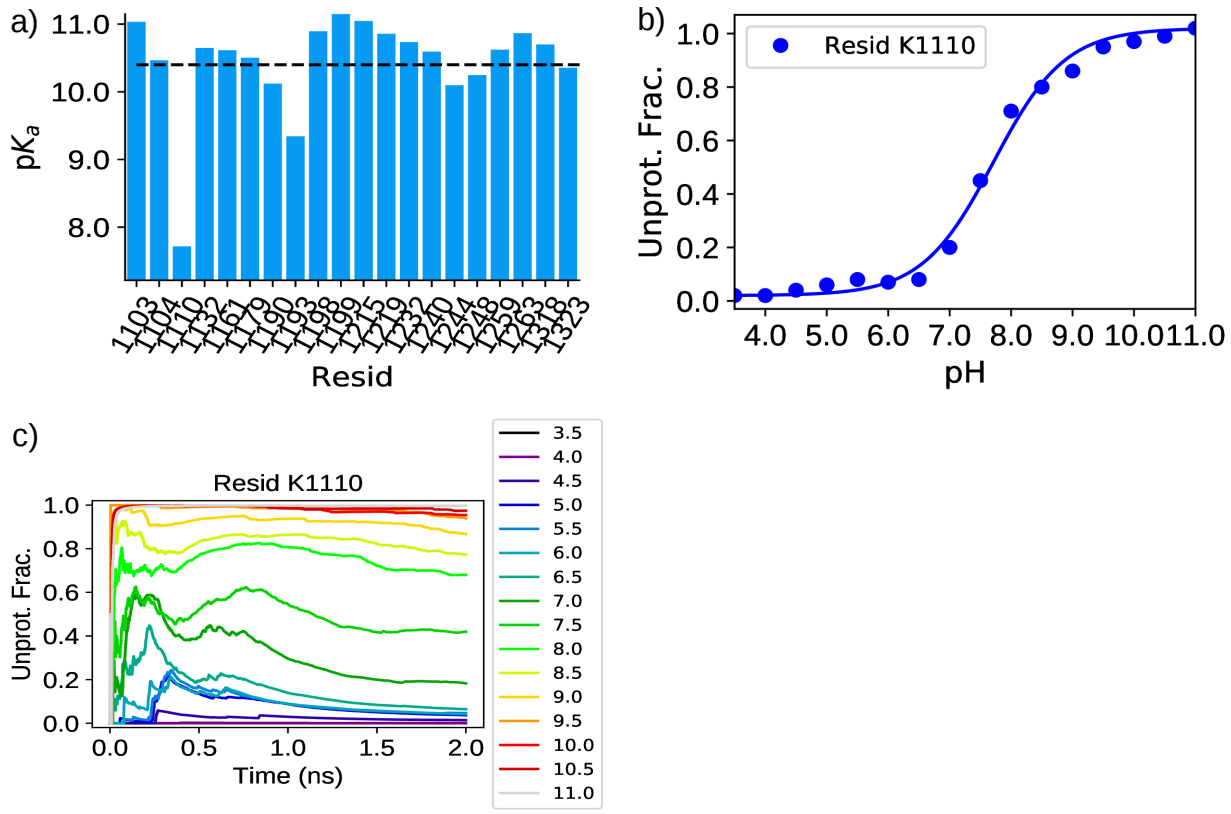


Figure S4: **Data from the CpHMD simulations of the MET kinase (PDB: 4MXC).** a) Calculated p$K_a$'s for all lysines. The dashed line represents the model p$K_a$ of lysine. b) Titration plots (unprotonated fractions at different pH) for the reactive lysines. c) Cumulatively calculated unprotonated fraction versus time at different pH for the reactive lysine.

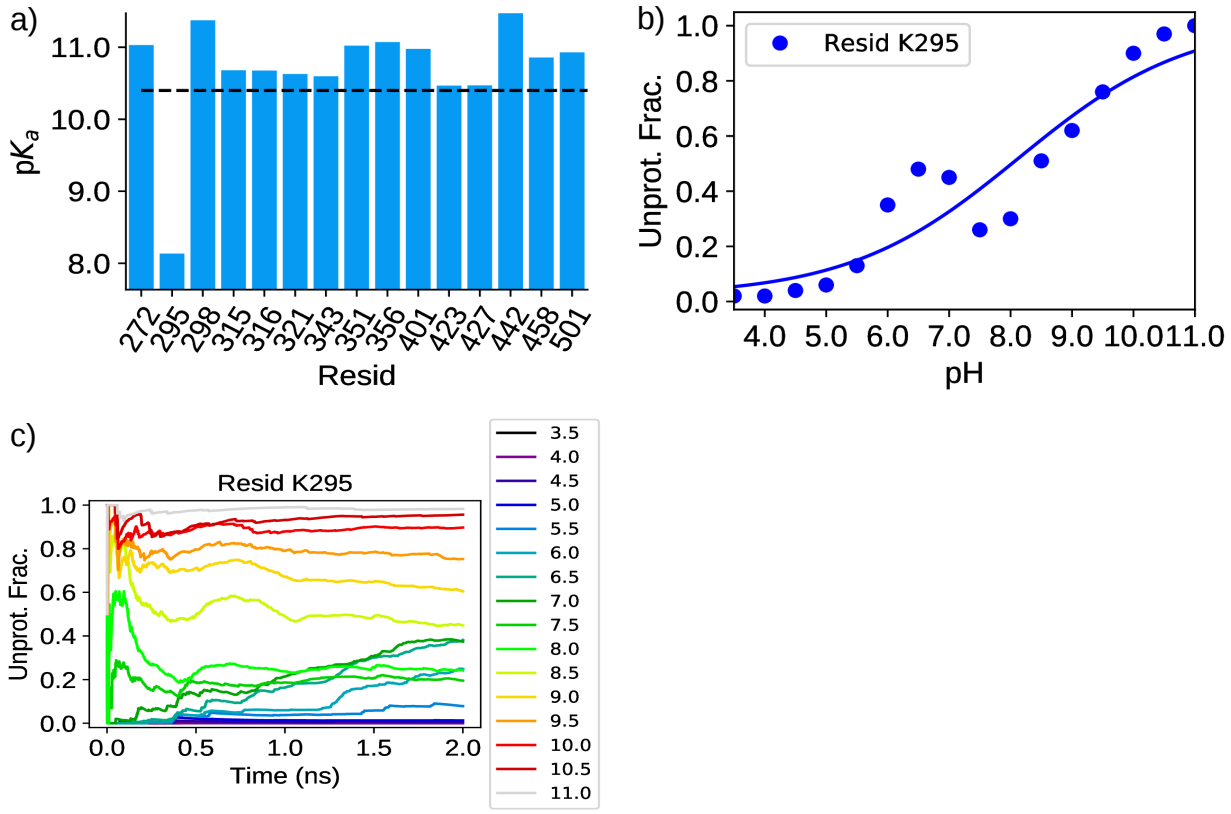


Figure S5: **Data from the CpHMD simulations of the c-SRC kinase (PDB: 5K9I).** a) Calculated p$K_a$'s for all lysines. The dashed line represents the model p$K_a$ of lysine. b) Titration plots (unprotonated fractions at different pH) for the reactive lysines. c) Cumulatively calculated unprotonated fraction versus time at different pH for the reactive lysine.

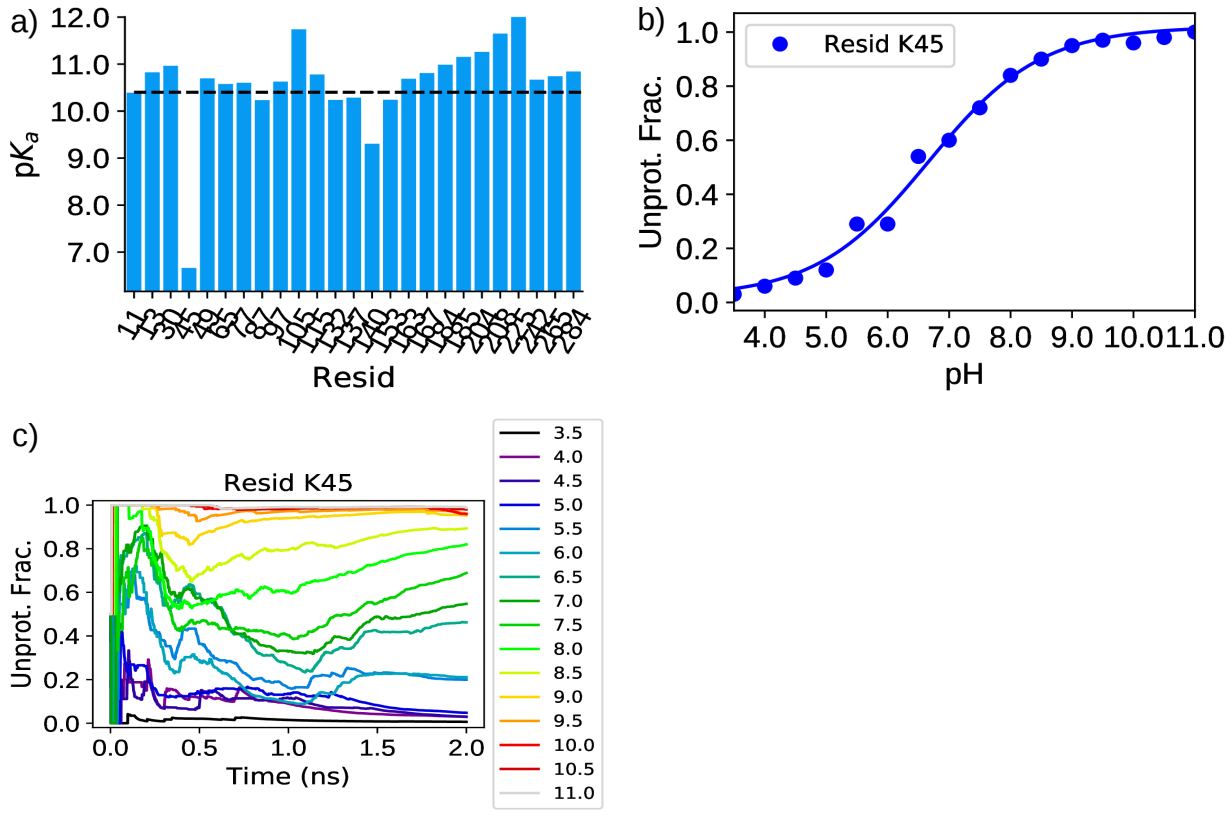


Figure S6: **Data from the CpHMD simulations of the RIPK1 kinase (PDB: 4ITJ).** a) Calculated p$K_a$'s for all lysines. The dashed line represents the model p$K_a$ of lysine. b) Titration plots (unprotonated fractions at different pH) for the reactive lysines. c) Cumulatively calculated unprotonated fraction versus time at different pH for the reactive lysine.

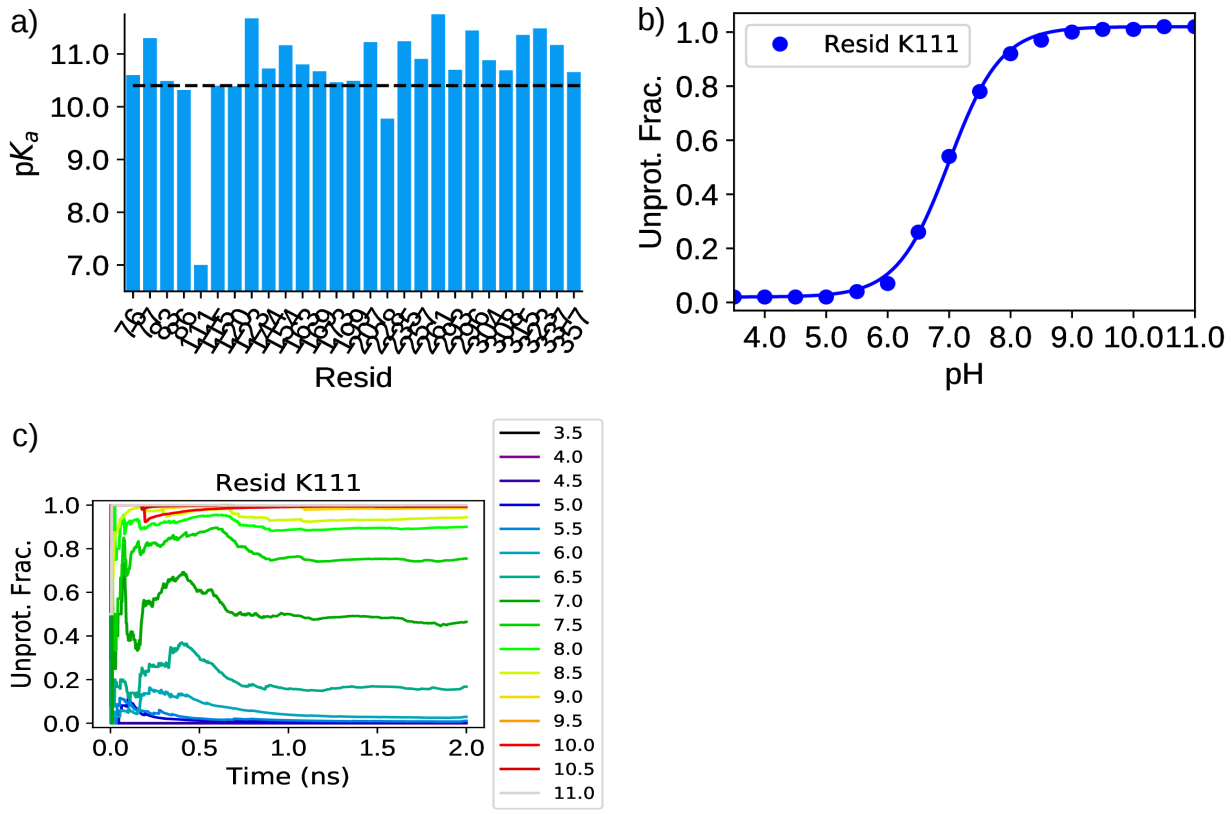


Figure S7: **Data from the CpHMD simulations of the PDK1 kinase (PDB: 3NAX).** a) Calculated p$K_a$'s for all lysines. The dashed line represents the model p$K_a$ of lysine. b) Titration plots (unprotonated fractions at different pH) for the reactive lysines. c) Cumulatively calculated unprotonated fraction versus time at different pH for the reactive lysine.

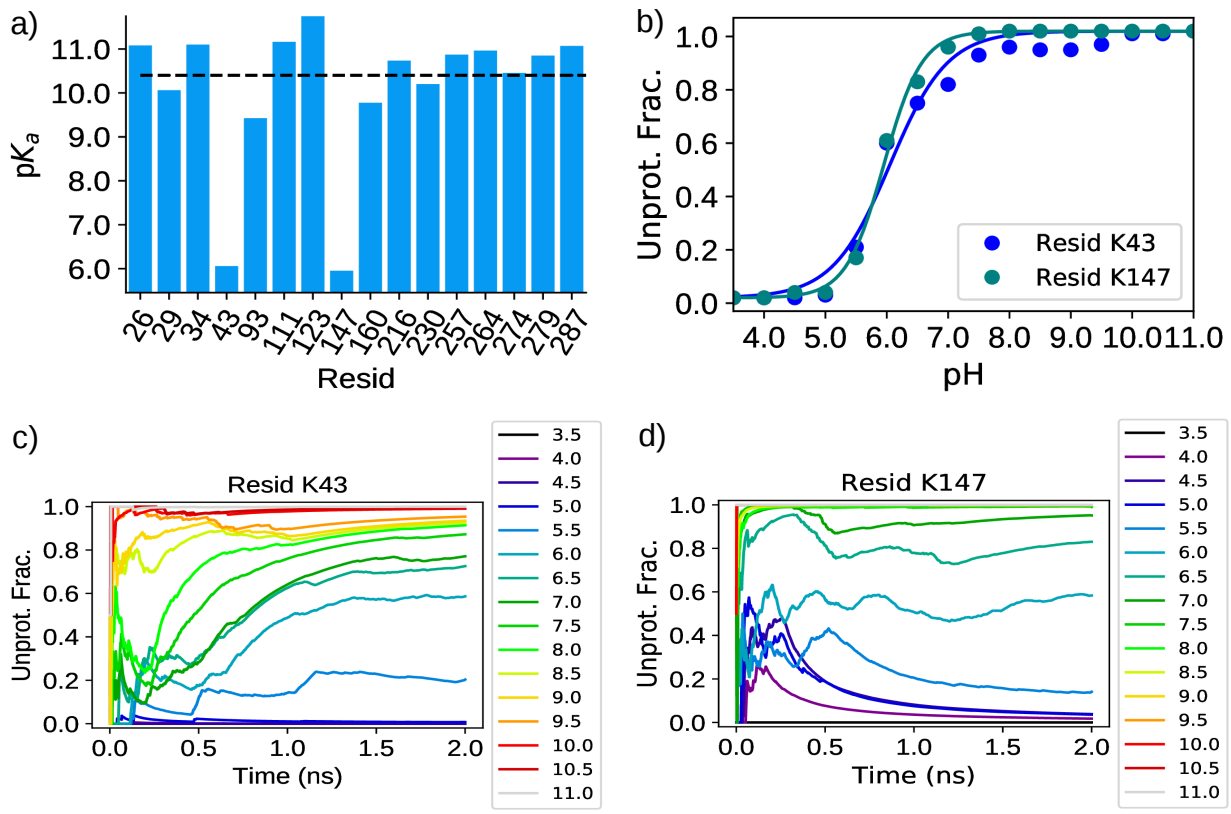


Figure S8: **Data from the CpHMD simulations of the CDK6 kinase (PDB: 1G3N).** a) Calculated p$K_a$'s for all lysines. The dashed line represents the model p$K_a$ of lysine. b) Titration plots (unprotonated fractions at different pH) for the reactive lysines. c) and d) Cumulatively calculated unprotonated fraction versus time at different pH for the reactive lysines.

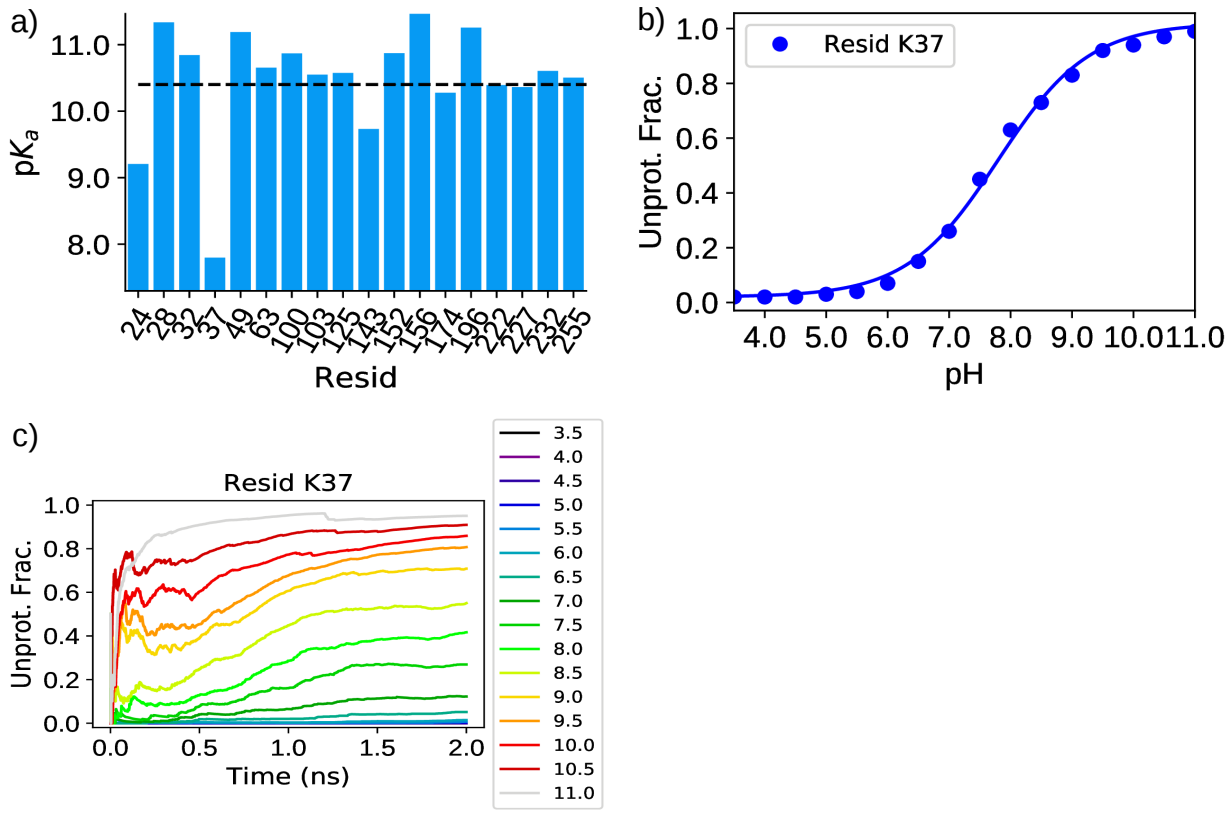


Figure S9: **Data from the CpHMD simulations of the NEK2 kinase (PDB: 2XNM).** a) Calculated p$K_a$'s for all lysines. The dashed line represents the model p$K_a$ of lysine. b) Titration plots (unprotonated fractions at different pH) for the reactive lysines. c) Cumulatively calculated unprotonated fraction versus time at different pH for the reactive lysine.

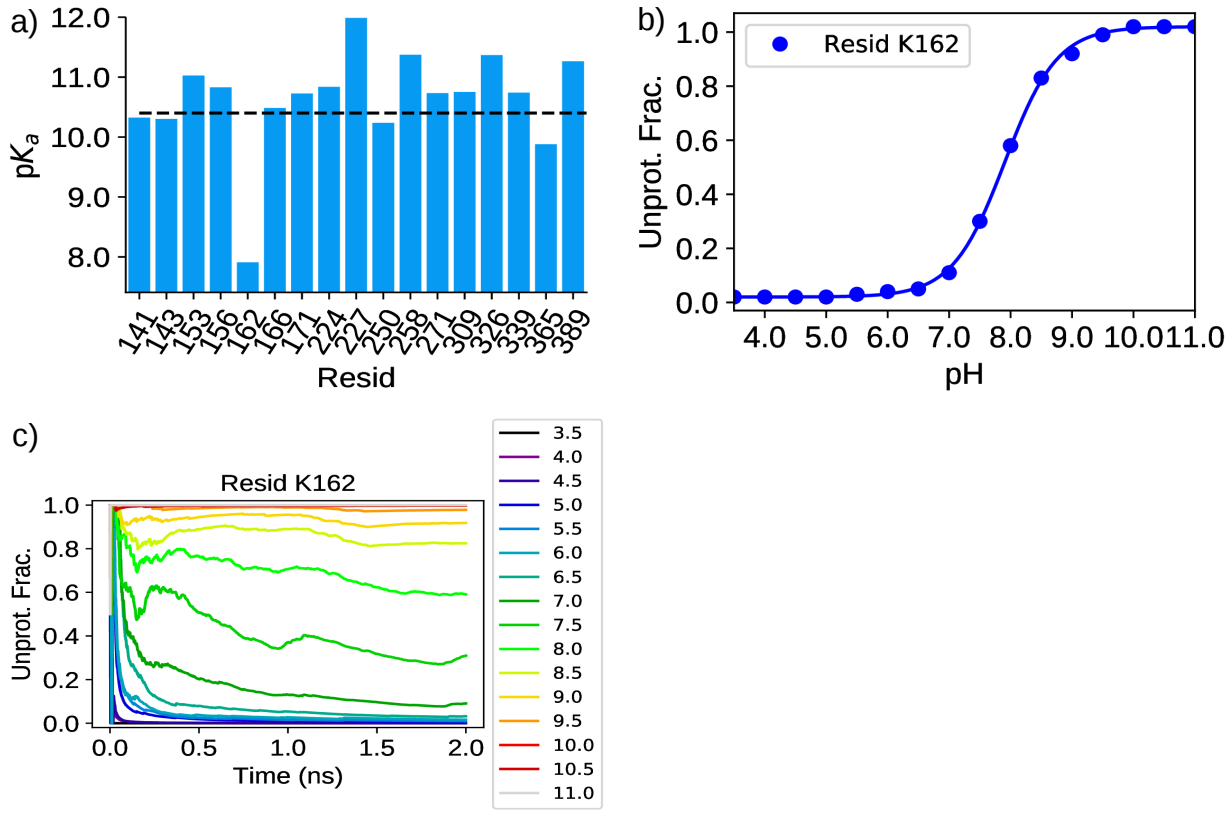


Figure S10: **Data from the CpHMD simulations of the Aurora kinase (PDB: 4JAI).** a) Calculated p$K_a$'s for all lysines. The dashed line represents the model p$K_a$ of lysine. b) Titration plots (unprotonated fractions at different pH) for the reactive lysines. c) Cumulatively calculated unprotonated fraction versus time at different pH for the reactive lysine.

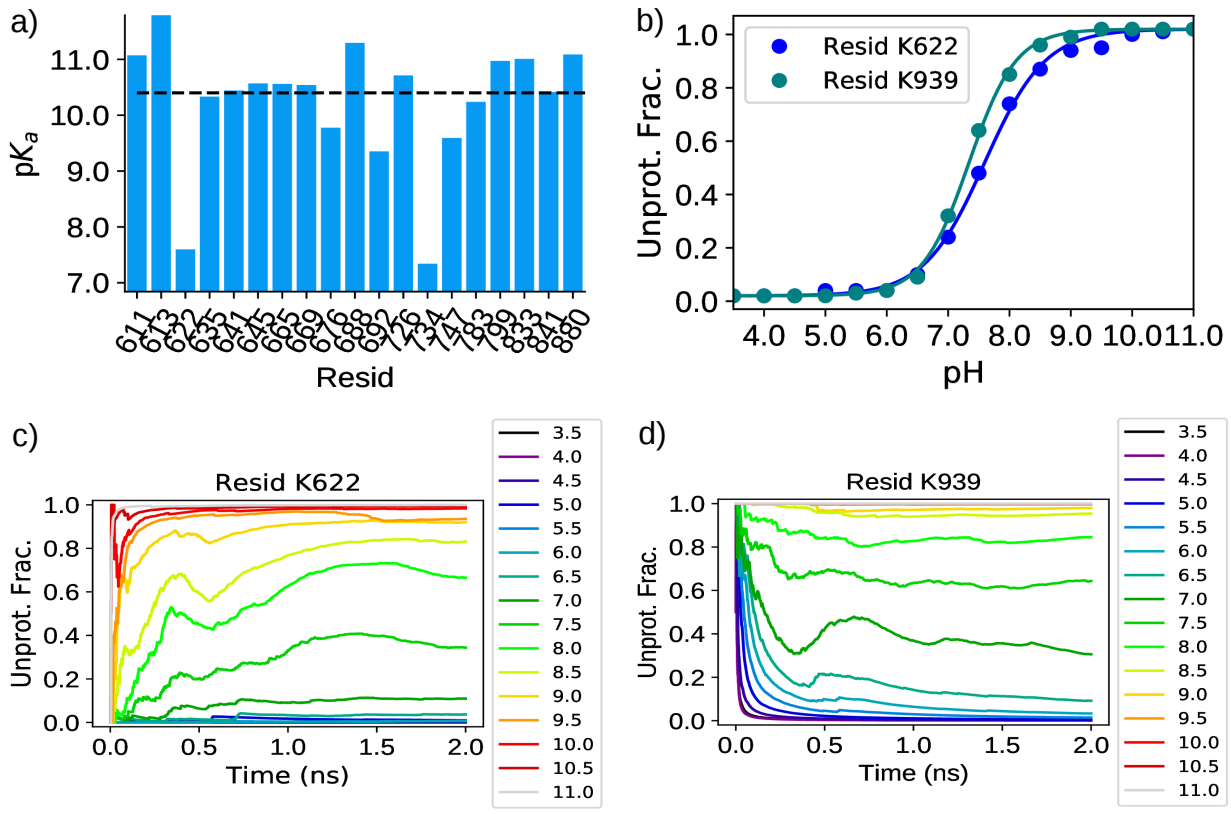


Figure S11: **Data from the CpHMD simulations of the PEK kinase (PDB: 4X7N).** a) Calculated p$K_a$'s for all lysines. The dashed line represents the model p$K_a$ of lysine. b) Titration plots (unprotonated fractions at different pH) for the reactive lysines. c) Cumulatively calculated unprotonated fraction versus time at different pH for the reactive lysines.

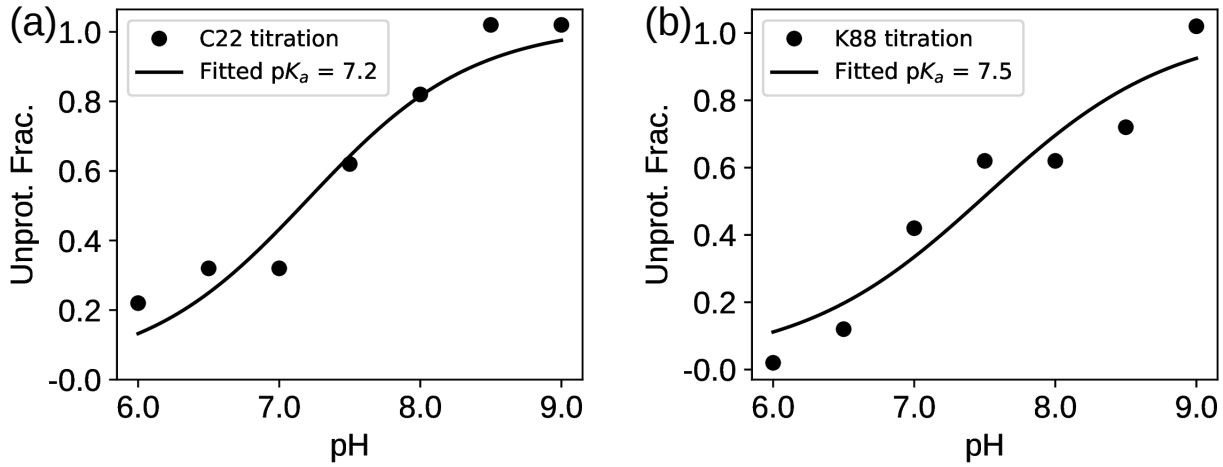


Figure S12: **Data from the preliminary CpHMD simulations to compare with the hyper-reactive lysine/cysteine residues discovered by chemical proteomic experiments.**[5–7] Titration plots for a) C22 of the MAP3 kinase ZAK (PDB: 5X5O) and b) K88 (K71 in the PDB file) of the Adenosine kinase ADK (PDB: 1BX4). These simulations were performed on a GPU card using our newly developed GPU-accelerated CpHMD implementation (Harris and Shen, manuscript submitted). Instead of pH replica-exchange, single-pH simulations were used. Otherwise, the simulation setup was identical to the one described in the main text. For ZAK, 5 10-ns simulations were run at each of the following pH values: 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, and 9.0. For ADK, 6 10-ns simulations were run at each of the following pH values: 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, and 9.0. For each simulation, an unprotonated fraction was calculated from the second half of the trajectory (5–10 ns). A data point represents the average unprotonated fraction calculated from multiple simulations at the same pH. The fitting p$K_{\mathrm{a}}$ was obtained from the best fit of the average unprotonated fractions at different pH to the generalized Henderson-Hasselbalch equation.

# References


(1) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, 296–303.

(2) Wang, P.-F.; McLeish, M. J.; Kneen, M. M.; Lee, G.; Kenyon, G. L. An Unusually Low p$K_a$ for Cys282 in the Active Site of Human Muscle Creatine Kinase. *Biochemistry* **2001**, *40*, 11698–11705.

(3) Isom, D. G.; Castañeda, C. A.; Cannon, B. R.; García-Moreno E., B. Large shifts in p$K_{\mathrm{a}}$ values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5260–5265.

(4) Søndergaard, C. R.; Mats H. M. Olsson, M. R.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p$K_a$ Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.

(5) Wang, C.; Weerapana, E.; Blewett, M. M.; Cravatt, B. F. A chemoproteomic platform to quantitatively map targets of lipid-derived electrophiles. *Nat. Methods* **2014**, *11*, 79–88.

(6) Backus, K. M.; Correia, B. E.; Lum, K. M.; Forli, S.; Horning, B. D.; González-Páez, G. E.; Chatterjee, S.; Lanning, B. R.; Teijaro, J. R.; Olson, A. J.; Wolan, D. W.; Cravatt, B. F. Proteome-wide covalent ligand discovery in native biological systems. *Nature* **2016**, *534*, 570–574.

(7) Hacker, S. M.; Backus, K. M.; Lazear, M. R.; Forli, S.; Correia, B. E.; Cravatt, B. F. Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **2017**, *9*, 1181–1190.