Supplementary Information (SI) for Physical Chemistry Chemical Physics.

# Supporting Information

## Probing Histidine Tautomer by Theoretical X-ray Absorption Spectroscopy for Biological and Pathological Studies

Hyundong Kim, Sompriya Chatterjee, Myounwoo Kim, Yeonsig Nam,* Shaul Mukamel,* and Jin Yong Lee*

Fig. S8 Theoretical N K-edge X-ray absorption spectra of fluoro-substituted δ (a) and ε tautomer (b) and π histidine (c) by RASPT2 (solid line) and RASSCF (dashed) in aqueous solution, respectively. The XAS spectra by RASPT2 come along with their analysis correspond to source nitrogen-delta (red), epsilon (green), backbone amine (blue) and substituent (grey) nitrogen-, respectively.

Fig. S9 Representative final orbital in the X-ray absorption transition with corresponding peak in theoretical N K-edge X-ray absorption spectra of δ (a) and ε tautomer (b) and π histidine (c) in aqueous solution, respectively. The white dots indicate the sites of excitation which remark core holes.

Fig. S10 Representative final orbital in the X-ray absorption transition with corresponding peak in theoretical N K-edge X-ray absorption spectra of methyl-substituted δ (a) and ε histidine (b) in aqueous solution, respectively. Peaks from backbone amine are not assigned. The white dots indicate the sites of excitation which remark core holes.

Fig. S11 Representative final orbital in the X-ray absorption transition with corresponding peak in theoretical N K-edge X-ray absorption spectra of amino-substituted δ (a) and ε histidine (b) in aqueous solution, respectively. Peaks from backbone amine are not assigned. The white dots indicate the sites of excitation which remark core holes.

Fig. S12 Representative final orbital in the X-ray absorption transition with corresponding peak in theoretical N K-edge X-ray absorption spectra of fluoro-substituted δ (a) and ε histidine (b) in aqueous solution, respectively. Peaks from backbone amine are not assigned. The white dots indicate the sites of excitation which remark core holes.

# Background signal

Expected background signal was obtained by theoretical study on other amino acids which contain notable nitrogen atom in their structure. RASSCF/RASPT2 study was conducted on arginine, asparagine, glutamine, lysine, proline and tryptophan molecule in the same way as the study of histidine.[S1-4] Their zwitterions were optimized by using B3LYP/aug-cc-pVDZ and C-PCM with ORCA.[S5] Information about active space and computed roots are described in Table S1.

Table S1 Computational details of RASSCF/RASPT2 for bacoground signal.

| | Active space (orbitals, electrons) | Number of roots |
|---|---|---|
| Arginine | (10, 10) | 10 |
| Asparagine | (6, 8) | 6 |
| Glutamine | (8, 8) | 8 |
| Lysine | (8, 8) | 8 |
| Proline | (8, 8) | 10 |
| Tryptophan | (12, 14) | 10 |

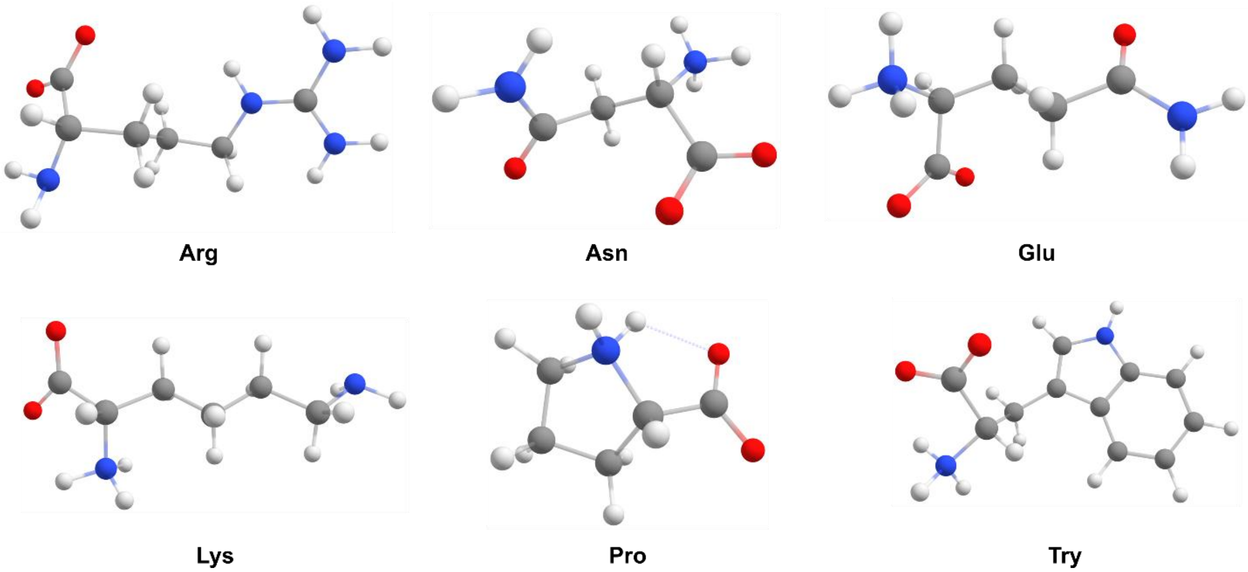
Arg
Asn
Glu
Lys
Pro
Try

Fig. S1 Optimized zwitterion of arginine (Arg), asparagine (Asn), glutamine (Glu), lysine (Lys), proline (Pro) and tryptophan (Try).

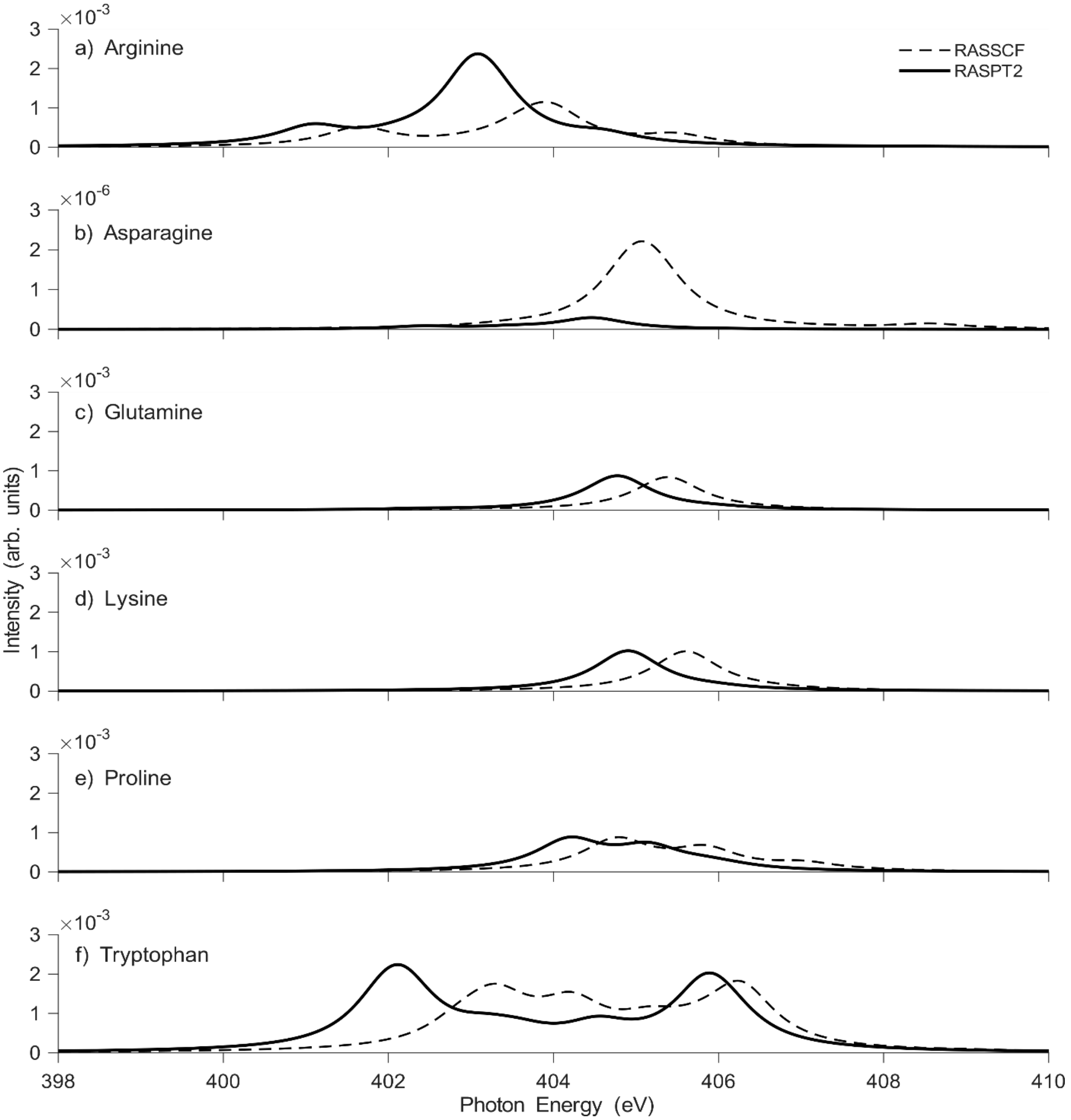

Fig. S2 Theoretical N K-edge X-ray absorption spectra of arginine (a), asparagine (b), glutamine (c), lysine (d), proline (e) and tryptophan (f) by RASPT2 (solid line) and RASSCF (dashed) in aqueous solution, respectively.

# Optimized geometries

Table S2 Computed optimized structure of the histidines and their substituted compounds. The following gives the xyz coordinates associated with the optimized structure.

| | | | |
|---|---|---|---|
| 1-1. δ His | | | |
| N | -3.60904 | -0.71987 | -1.4425 |
| H | -4.28514 | -1.43935 | -1.74631 |
| C | -3.09839 | -1.01084 | -0.0566 |
| H | -2.42328 | -0.12761 | 0.244606 |
| C | -4.26177 | -1.05286 | 0.937965 |
| C | -2.22873 | -2.28659 | -0.09949 |
| H | -5.03716 | -1.76209 | 0.597431 |
| H | -3.85778 | -1.45425 | 1.880079 |
| C | -4.85848 | 0.306292 | 1.170243 |
| O | -1.99435 | -2.774 | -1.24024 |
| N | -6.07773 | 0.492983 | 1.788435 |
| C | -4.40289 | 1.584548 | 0.906995 |
| C | -6.30731 | 1.833642 | 1.878035 |
| N | -5.31119 | 2.528732 | 1.353714 |
| O | -1.80525 | -2.70341 | 1.008623 |
| H | -4.07527 | 0.197662 | -1.50432 |
| H | -7.20635 | 2.242815 | 2.331481 |
| H | -2.83118 | -0.74958 | -2.11784 |
| H | -6.7017 | -0.2413 | 2.129381 |
| H | -3.47224 | 1.879429 | 0.428215 |
| 1-2. ε His | | | |
| C | -2.26129 | -2.29359 | -0.09943 |
| C | -3.09118 | -0.99215 | -0.05913 |
| C | -4.24949 | -1.01879 | 0.943281 |
| C | -4.35685 | 1.587107 | 0.923018 |
| C | -4.88204 | 0.330778 | 1.158532 |
| C | -6.38449 | 1.754834 | 1.784386 |
| O | -1.71345 | -2.62587 | 0.983246 |
| O | -2.19996 | -2.89869 | -1.20575 |
| H | -2.40241 | -0.12599 | 0.21949 |
| H | -2.84934 | -0.61563 | -2.11678 |
| H | -3.40555 | 1.925579 | 0.523126 |
| H | -3.83936 | -1.39935 | 1.891307 |
| H | -4.17897 | 0.162735 | -1.46409 |
| H | -4.21178 | -1.48245 | -1.77071 |
| H | -5.0199 | -1.74023 | 0.621044 |
| H | -5.26847 | 3.495972 | 1.283917 |
| H | -7.29026 | 2.218733 | 2.166286 |
| N | -3.62017 | -0.70391 | -1.43941 |
| N | -5.32891 | 2.476975 | 1.328851 |
| N | -6.15264 | 0.452178 | 1.69513 |
| 1-3. π His | | | |
| C | -2.20061 | -2.24794 | -0.0906 |
| C | -3.10822 | -0.98595 | -0.08652 |
| C | -4.25165 | -1.00886 | 0.926959 |
| C | -4.4003 | 1.613228 | 0.928219 |
| C | -4.85971 | 0.340872 | 1.166312 |
| C | -6.38433 | 1.799504 | 1.907584 |
| O | -1.83785 | -2.64853 | -1.23235 |
| O | -1.88815 | -2.71869 | 1.02702 |
| H | -2.41434 | -0.09417 | 0.163598 |
| H | -2.87824 | -1.1443 | -2.12324 |

|  |  |  |  |
|---|---|---|---|
| H | -3.48135 | 1.967955 | 0.47544 |
| H | -3.83216 | -1.38977 | 1.871308 |
| H | -3.84532 | 0.190673 | -1.72098 |
| H | -4.45996 | -1.35879 | -1.67252 |
| H | -5.03775 | -1.7201 | 0.617399 |
| H | -5.31126 | 3.512266 | 1.354131 |
| H | -6.71291 | -0.25947 | 2.086683 |
| H | -7.28618 | 2.215843 | 2.348462 |
| N | -3.61845 | -0.78738 | -1.49068 |
| N | -5.36237 | 2.485879 | 1.395743 |
| N | -6.09435 | 0.500464 | 1.77353 |

2-1. δ $CH_3$-His

|  |  |  |  |
|---|---|---|---|
| N | -4.22857 | -1.18505 | -1.34702 |
| H | -5.0052 | -1.80253 | -1.05852 |
| C | -3.1475 | -1.12491 | -0.30142 |
| H | -2.40591 | -0.28853 | -0.65486 |
| C | -3.73992 | -0.77526 | 1.072614 |
| C | -2.3651 | -2.45863 | -0.32922 |
| H | -4.34705 | -1.62282 | 1.435006 |
| H | -2.88838 | -0.67934 | 1.759535 |
| C | -4.57025 | 0.474204 | 1.108354 |
| O | -2.61499 | -3.24196 | -1.28912 |
| N | -5.95621 | 0.438472 | 1.12349 |
| C | -4.24714 | 1.819977 | 1.128988 |
| C | -6.40872 | 1.718036 | 1.159851 |
| N | -5.40466 | 2.582358 | 1.159902 |
| O | -1.51385 | -2.62249 | 0.58272 |
| H | -4.61999 | -0.25351 | -1.55253 |
| H | -7.46718 | 1.964302 | 1.181008 |
| H | -3.8386 | -1.59357 | -2.20966 |
| H | -6.53469 | -0.40437 | 1.137326 |
| C | -2.89035 | 2.454237 | 1.124139 |
| H | -2.10018 | 1.69616 | 1.194424 |
| H | -2.72288 | 3.032661 | 0.202481 |
| H | -2.77772 | 3.147602 | 1.970741 |

2-2. ε $CH_3$-His

|  |  |  |  |
|---|---|---|---|
| N | -4.35258 | -1.2616 | -1.28793 |
| H | -5.18524 | -1.59221 | -0.77671 |
| C | -3.16338 | -1.09958 | -0.37708 |
| H | -2.49257 | -0.29197 | -0.84218 |
| C | -3.60825 | -0.6821 | 1.032397 |
| C | -2.34662 | -2.41406 | -0.43061 |
| H | -4.13354 | -1.52777 | 1.506972 |
| H | -2.69665 | -0.49743 | 1.615098 |
| C | -4.51155 | 0.517101 | 1.059126 |
| O | -2.65477 | -3.23021 | -1.34835 |
| N | -5.88813 | 0.366928 | 0.929289 |
| C | -4.19027 | 1.856323 | 1.184152 |
| C | -6.38815 | 1.590386 | 0.982519 |
| N | -5.40415 | 2.517199 | 1.134541 |
| O | -1.42583 | -2.53787 | 0.416076 |
| H | -4.61122 | -0.38733 | -1.7649 |
| H | -7.44038 | 1.856085 | 0.923958 |
| H | -4.09467 | -1.99118 | -1.97807 |
| H | -5.54212 | 3.526658 | 1.213487 |
| C | -2.88663 | 2.572897 | 1.322272 |
| H | -2.07328 | 1.850937 | 1.46326 |
| H | -2.65736 | 3.168621 | 0.424756 |
| H | -2.89126 | 3.257535 | 2.183817 |

| | | | |
|---|---|---|---|
| 2-3. π $CH_3$-His | | | |
| N | -4.12511 | -1.12472 | -1.39797 |
| H | -4.97238 | -1.64247 | -1.11173 |
| C | -3.13426 | -1.00408 | -0.27252 |
| H | -2.40353 | -0.15271 | -0.57453 |
| C | -3.83473 | -0.78985 | 1.071308 |
| C | -2.25219 | -2.2901 | -0.30133 |
| H | -4.48142 | -1.65711 | 1.288116 |
| H | -3.05248 | -0.77518 | 1.843138 |
| C | -4.65415 | 0.458731 | 1.157884 |
| O | -2.31361 | -2.96809 | -1.36701 |
| N | -6.03441 | 0.465304 | 1.303633 |
| C | -4.26515 | 1.774634 | 1.114013 |
| C | -6.48487 | 1.720241 | 1.348423 |
| N | -5.42481 | 2.525402 | 1.232531 |
| O | -1.54252 | -2.50327 | 0.709315 |
| H | -4.41872 | -0.21768 | -1.78821 |
| H | -7.51951 | 2.029613 | 1.464809 |
| H | -3.65649 | -1.70299 | -2.1222 |
| H | -5.47801 | 3.55195 | 1.23672 |
| H | -6.63161 | -0.36841 | 1.373013 |
| C | -2.90609 | 2.357079 | 0.963091 |
| H | -2.1461 | 1.591701 | 1.155108 |
| H | -2.75313 | 2.746537 | -0.05533 |
| H | -2.75287 | 3.186811 | 1.667296 |
| 3-1. δ $NH_2$-His | | | |
| N | -4.28573 | -1.40194 | -1.30878 |
| H | -5.13444 | -1.74892 | -0.85516 |
| C | -3.20801 | -1.02661 | -0.32362 |
| H | -2.70071 | -0.14969 | -0.74358 |
| C | -3.78253 | -0.68284 | 1.060929 |
| C | -2.17918 | -2.1999 | -0.31893 |
| H | -4.36305 | -1.5409 | 1.43321 |
| H | -2.91934 | -0.57321 | 1.729576 |
| C | -4.61645 | 0.560796 | 1.105491 |
| O | -2.35323 | -3.08019 | -1.21316 |
| N | -6.00869 | 0.567478 | 1.049148 |
| C | -4.2756 | 1.90101 | 1.198259 |
| C | -6.42977 | 1.851217 | 1.113542 |
| N | -5.40329 | 2.693323 | 1.201364 |
| O | -1.26906 | -2.14057 | 0.538721 |
| H | -4.55233 | -0.64198 | -1.93717 |
| H | -7.47847 | 2.127757 | 1.100018 |
| H | -3.86727 | -2.19284 | -1.84147 |
| H | -6.60532 | -0.24862 | 0.991472 |
| N | -2.99625 | 2.481768 | 1.211433 |
| H | -2.982 | 3.350957 | 1.737939 |
| H | -2.27675 | 1.859413 | 1.568205 |
| 3-2. ε $NH_2$-His | | | |
| N | -4.54889 | -1.51196 | -1.14348 |
| H | -5.32859 | -0.9265 | -0.77668 |
| C | -3.23509 | -1.05511 | -0.56633 |
| H | -2.84007 | -0.27782 | -1.22828 |
| C | -3.4649 | -0.48442 | 0.848531 |
| C | -2.26821 | -2.26929 | -0.53003 |
| H | -3.81887 | -1.30135 | 1.499063 |
| H | -2.48697 | -0.17336 | 1.233815 |
| C | -4.42667 | 0.666925 | 0.881549 |
| O | -2.8034 | -3.40932 | -0.63672 |

| | | | |
|---|---|---|---|
| N | -5.7165 | 0.539781 | 0.362628 |
| C | -4.26629 | 1.945003 | 1.385863 |
| C | -6.30846 | 1.705399 | 0.539574 |
| N | -5.47082 | 2.582665 | 1.166717 |
| O | -1.05254 | -1.99787 | -0.3663 |
| H | -4.55077 | -1.4639 | -2.16408 |
| H | -7.31478 | 1.972278 | 0.236879 |
| H | -4.65687 | -2.50369 | -0.88612 |
| N | -3.2324 | 2.575357 | 2.086637 |
| H | -3.02502 | 3.508491 | 1.737281 |
| H | -2.37848 | 2.02604 | 2.085813 |
| H | -5.69406 | 3.534337 | 1.432355 |
| 3-3. π $NH_2$-His | | | |
| N | -4.24859 | -1.38563 | -1.32154 |
| H | -5.09408 | -1.75747 | -0.88161 |
| C | -3.19619 | -0.97693 | -0.3243 |
| H | -2.7116 | -0.08336 | -0.73554 |
| C | -3.78489 | -0.66708 | 1.060583 |
| C | -2.12562 | -2.11617 | -0.31217 |
| H | -4.37007 | -1.52796 | 1.415443 |
| H | -2.93132 | -0.5595 | 1.741371 |
| C | -4.62413 | 0.570594 | 1.125343 |
| O | -2.25385 | -2.98429 | -1.22447 |
| N | -6.01743 | 0.570319 | 1.075792 |
| C | -4.26148 | 1.900444 | 1.237666 |
| C | -6.49362 | 1.807531 | 1.158695 |
| N | -5.43729 | 2.632336 | 1.251443 |
| O | -1.23784 | -2.03665 | 0.56518 |
| H | -4.52679 | -0.63624 | -1.95794 |
| H | -7.53464 | 2.101542 | 1.1612 |
| H | -3.79345 | -2.1645 | -1.845 |
| H | -5.50266 | 3.642558 | 1.326008 |
| H | -6.60074 | -0.25706 | 1.003175 |
| N | -3.02619 | 2.512775 | 1.242668 |
| H | -2.9737 | 3.361565 | 1.797866 |
| H | -2.25305 | 1.889746 | 1.450835 |
| 4-1. δ F-His | | | |
| N | -4.08524 | -1.04013 | -1.35884 |
| H | -4.99958 | -1.42703 | -1.11051 |
| C | -3.14183 | -0.94948 | -0.18668 |
| H | -2.54427 | -0.04359 | -0.34529 |
| C | -3.87628 | -0.86088 | 1.159595 |
| C | -2.17176 | -2.1698 | -0.29964 |
| H | -4.54943 | -1.72454 | 1.267317 |
| H | -3.10727 | -0.9563 | 1.934242 |
| C | -4.63681 | 0.409365 | 1.381729 |
| O | -2.24279 | -2.81688 | -1.38553 |
| N | -5.98068 | 0.600234 | 1.088931 |
| C | -4.26955 | 1.635837 | 1.885363 |
| C | -6.32166 | 1.870506 | 1.420887 |
| N | -5.28018 | 2.536619 | 1.916582 |
| O | -1.40378 | -2.36333 | 0.669085 |
| H | -4.23561 | -0.14796 | -1.83348 |
| H | -7.32316 | 2.264761 | 1.291674 |
| F | -3.03886 | 1.979973 | 2.329636 |
| H | -3.61922 | -1.72248 | -1.99358 |
| H | -6.60713 | -0.09902 | 0.706733 |
| 4-2. ε F-His | | | |
| N | -4.24068 | -1.08823 | -1.27702 |

| | | | |
|---|---|---|---|
| H | -4.47769 | -2.09046 | -1.32918 |
| C | -3.09088 | -0.96855 | -0.31107 |
| H | -2.47949 | -0.11923 | -0.6321 |
| C | -3.62972 | -0.73342 | 1.113198 |
| C | -2.26214 | -2.28148 | -0.37887 |
| H | -4.21659 | -1.61392 | 1.419381 |
| H | -2.76696 | -0.66666 | 1.785259 |
| C | -4.46339 | 0.505329 | 1.23774 |
| O | -2.84088 | -3.27365 | -0.90501 |
| N | -5.59706 | 0.722121 | 0.457632 |
| C | -4.32038 | 1.591351 | 2.058659 |
| C | -6.0929 | 1.894752 | 0.815235 |
| N | -5.34282 | 2.465192 | 1.801108 |
| O | -1.11357 | -2.23016 | 0.125183 |
| H | -5.04892 | -0.5165 | -0.96833 |
| H | -5.50376 | 3.35633 | 2.257326 |
| H | -6.97442 | 2.37349 | 0.40393 |
| F | -3.40249 | 1.875431 | 2.992381 |
| H | -3.97415 | -0.79259 | -2.21821 |

4-3. π F-His

| | | | |
|---|---|---|---|
| C | -2.22141 | -2.25328 | -0.26423 |
| C | -3.08057 | -0.96127 | -0.17018 |
| C | -4.10128 | -0.96863 | 0.96746 |
| C | -4.29515 | 1.65152 | 1.080347 |
| C | -4.74477 | 0.363845 | 1.204258 |
| C | -6.36046 | 1.798353 | 1.822031 |
| F | -3.12236 | 2.118775 | 0.680345 |
| H | -2.35845 | -0.10076 | -0.03036 |
| H | -3.06702 | -0.93665 | -2.24054 |
| H | -3.56943 | -1.28784 | 1.87682 |
| H | -4.1414 | 0.202481 | -1.61436 |
| H | -4.5278 | -1.4229 | -1.63859 |
| H | -4.89051 | -1.71595 | 0.773483 |
| H | -5.22849 | 3.547802 | 1.466109 |
| H | -6.68963 | -0.26609 | 1.872804 |
| H | -7.31008 | 2.195319 | 2.171313 |
| N | -3.75798 | -0.7483 | -1.49592 |
| N | -5.29123 | 2.518155 | 1.460852 |
| N | -6.04638 | 0.512177 | 1.669559 |
| O | -1.78629 | -2.70641 | 0.821283 |
| O | -2.02046 | -2.68928 | -1.43084 |

# X-ray absorption spectra in gas phase

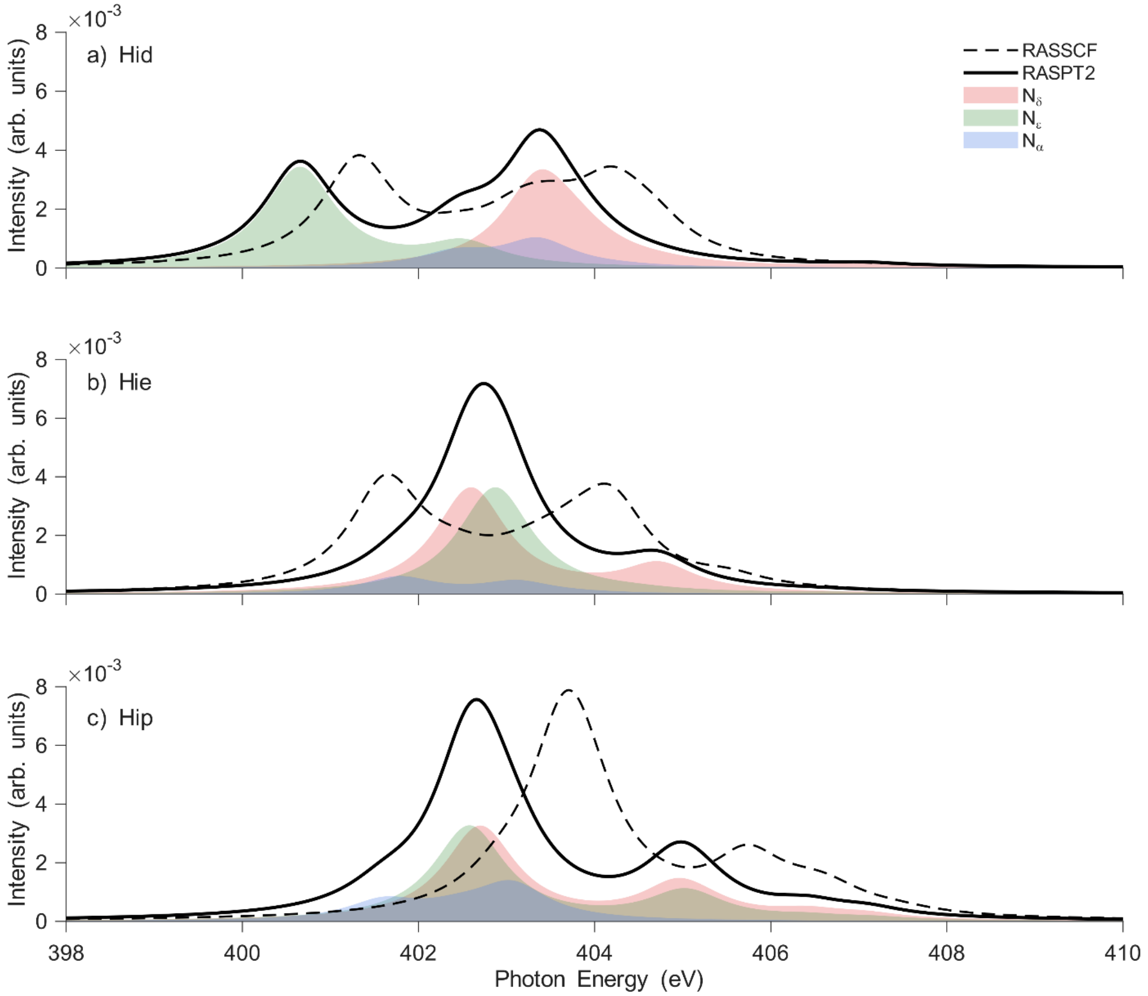

Fig. S3 Theoretical N K-edge X-ray absorption spectra of neutral δ (a) and ε tautomer (b) and π histidine by RASPT2 (solid line) and RASSCF (dashed) in gas phase, respectively. The XAS spectra by RASPT2 come along with their analysis correspond to source nitrogen-delta (red), epsilon (green) and backbone amine (blue) nitrogen-, respectively.

## Active space orbitals for the ground state calculations

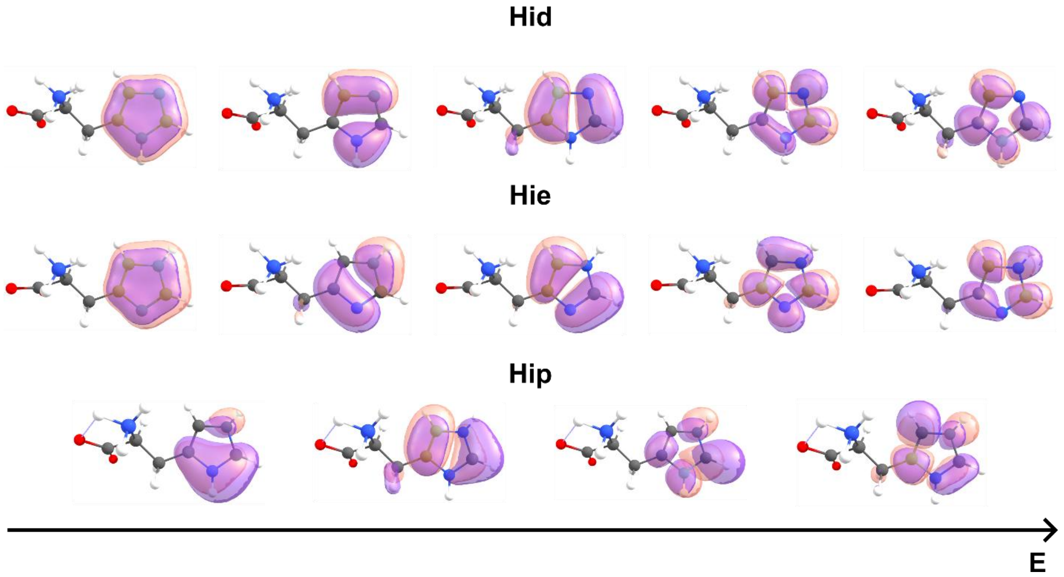

Fig. S4 π bonding and antibonding molecular orbitals along the imidazole ring of δ (top) and ε tautomer (middle) and π histidine (bottom) considered in (10,1/9/0) active space for the ground state RASSCF calculations. Orbitals in the active that shows no π character are not shown here. Isovalue of 0.3 a.u. is used to plot the molecular orbitals.

## X-ray absorption spectra – major peaks

Table S3 Calculated main N K-edge from the imidazole ring for δ and ε tautomers with RASPT2.

| | Energy (eV) | | | | | |
|---|---|---|---|---|---|---|
| | δ tautomer | | ε tautomer | | $\Delta E_{\delta-\varepsilon}$ | |
| | $N_\delta$ | $N_\varepsilon$ | $N_\delta$ | $N_\varepsilon$ | Pyridinic | Pyrrolic |
| His | 403.9 | 401.1 | 400.6 | 403.1 | 0.5 | 0.8 |
| $CH_3$-His | 403.0 | 400.6 | 400.2 | 403.3 | 0.4 | 0.3 |
| $NH_2$-His | 403.0 | 400.9 | 400.5 | 403.5 | 0.4 | 0.5 |
| F-His | 402.6 | 400.9 | 400.8 | 404.6 | 0.1 | 2.0 |

# Mulliken charge analysis

Table S4 Mulliken charge on nitrogen and substituent (substituted C or F) of histidine and derivatives. a. Muliken charge data is adopted from nonequilibrium RASSCF with different reaction field.

| | Na | Nd | Ne | Substituent |
|---|---|---|---|---|
| δ His – $N_\delta$ peak | | | | |
| Initial state | 0.7018 | 0.0776 | -1.0813 | - |
| Final state | 0.6947 | -0.1132 | -1.0831 | - |
| δ His – $N_\varepsilon$ peak | | | | |
| Initial state | 0.7018 | 0.0775 | -1.0819 | - |
| Final state | 0.6984 | 0.0910 | -1.3059 | - |
| ε His – $N_\delta$ peak | | | | |
| Initial state | 0.6962 | -0.9531 | -0.0391 | - |
| Final state | 0.6846 | -1.1511 | -0.0556 | - |
| ε His – $N_\varepsilon$ peak | | | | |
| Initial state | 0.6962 | -0.9531 | -0.0391 | - |
| Final state | 0.6742 | -0.8799 | -0.2001 | - |
| δ $CH_3$-His – $N_\delta$ peak | | | | |
| Initial state | 0.7077 | 0.2204 | -0.9025 | 0.6655 |
| Final state | 0.5762 | 0.0642 | -0.7744 | 0.7632 |
| δ $CH_3$-His – $N_\varepsilon$ peak | | | | |
| Initial state | 0.7077 | 0.2204 | -0.9022 | 0.6653 |
| Final state | 0.6847 | 0.2524 | -1.1994 | 0.7212 |
| ε $CH_3$-His – $N_\delta$ peak | | | | |
| Initial state | 0.6459 | -0.8230 | 0.1304 | 0.5012 |
| Final state | 0.6015 | -1.1063 | 0.1803 | 0.5162 |
| ε $CH_3$-His – $N_\varepsilon$ peak | | | | |
| Initial state | 0.6459 | -0.8232 | 0.1304 | 0.5011 |
| Final state | 0.5695 | -0.6764 | 0.0324 | 0.6113 |
| δ F-His – $N_\delta$ peak | | | | |
| Initial state | 0.5913 | 0.2064 | -0.8822 | -0.7674 |
| Final state | 0.5729 | 0.1804 | -1.1033 | -0.772 |
| δ F-His – $N_\varepsilon$ peak | | | | |
| Initial state | 0.5913 | 0.2063 | -0.8824 | -0.7674 |
| Final state | 0.5082 | 0.0266 | -0.8418 | -0.7534 |
| ε F-His – $N_\delta$ peak | | | | |
| Initial state | 0.6215 | -0.7946 | 0.0546 | -0.7093 |
| Final state | 0.5988 | -1.0668 | 0.0613 | -0.7057 |
| ε F-His – $N_\varepsilon$ peak | | | | |
| Initial state | 0.6215 | -0.7947 | 0.0546 | -0.7093 |
| Final state[a] | 0.5995 | -0.7035 | -0.0343 | -0.7356 |

# X-ray absorption spectra – both RASSCF/PT2

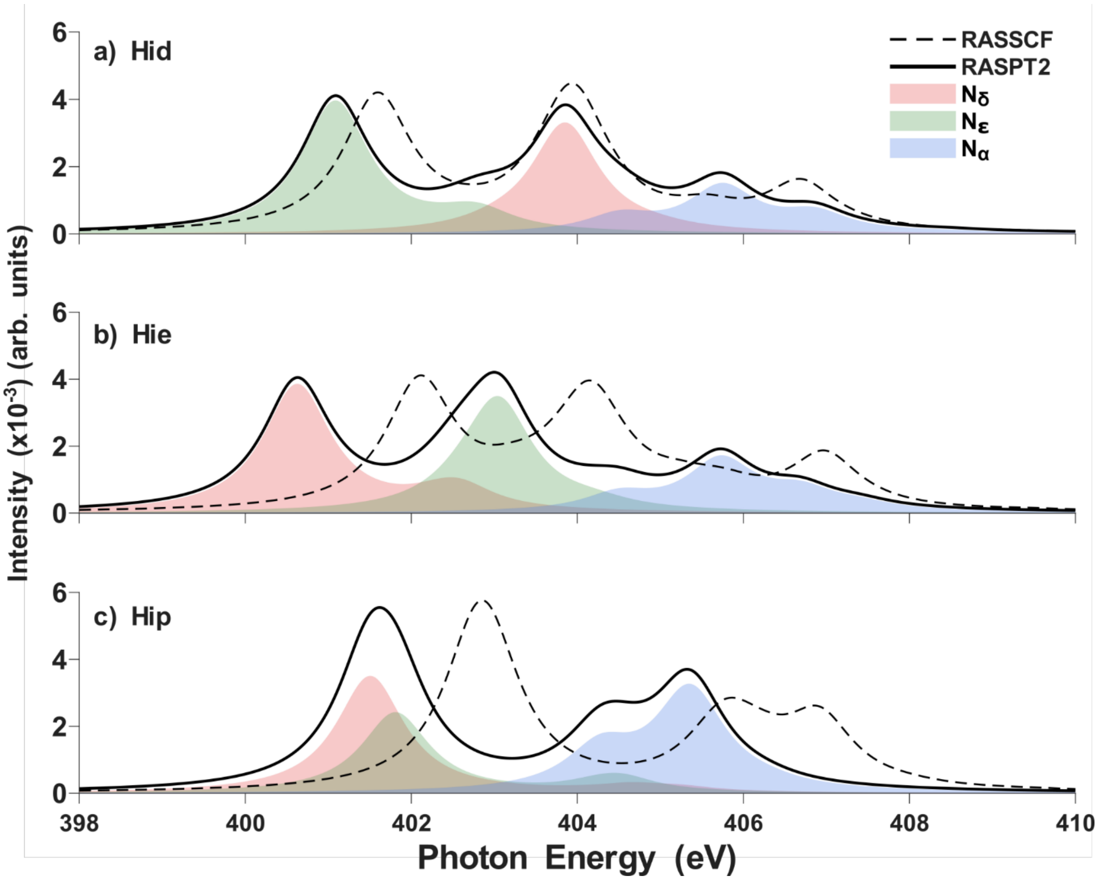

Fig. S5 Theoretical N K-edge X-ray absorption spectra of δ (a) and ε tautomer (b) and π histidine (c) by RASPT2 (solid line) and RASSCF (dashed) in aqueous solution, respectively. The XAS spectra by RASPT2 come along with their analysis correspond to source nitrogen-delta (red), epsilon (green) and backbone amine (blue) nitrogen-, respectively.

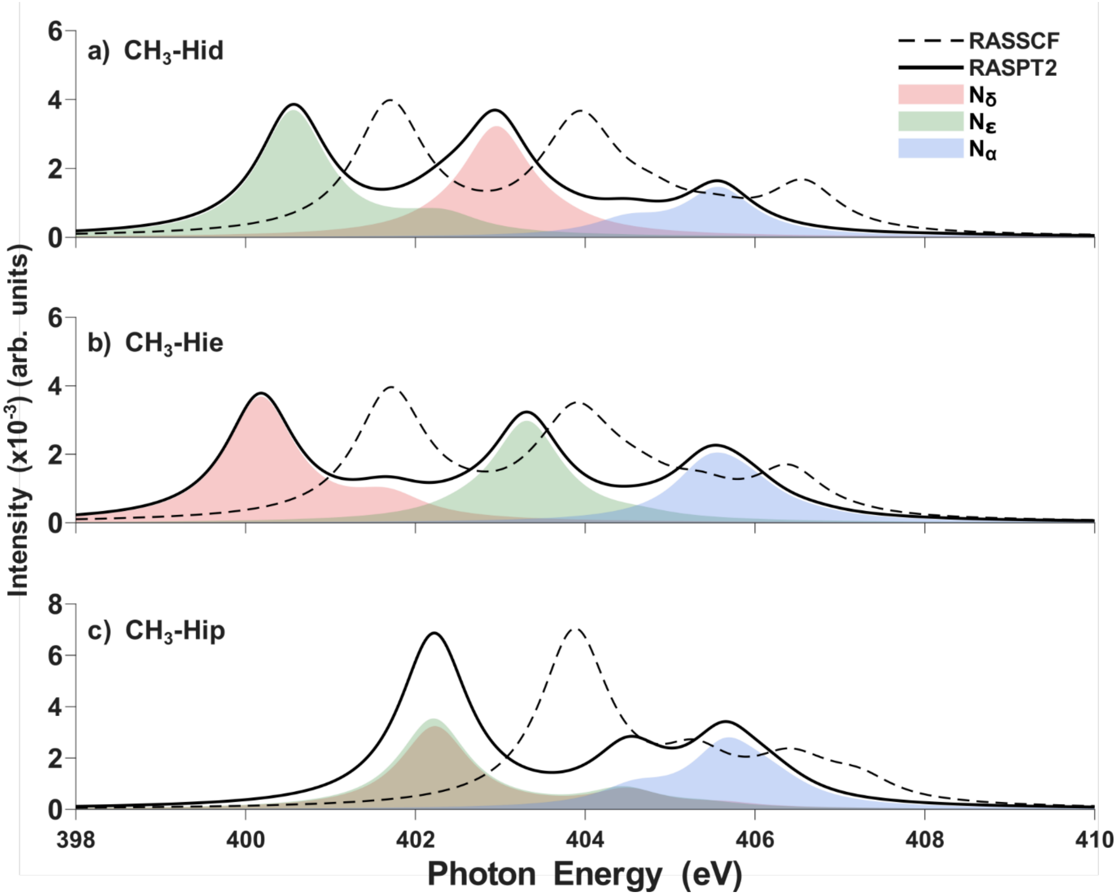

Fig. S6 Theoretical N K-edge X-ray absorption spectra of methyl-substituted δ (a) and ε tautomer (b) and π histidine (c) by RASPT2 (solid line) and RASSCF (dashed) in aqueous solution, respectively. The XAS spectra by RASPT2 come along with their analysis correspond to source nitrogen-delta (red), epsilon (green), backbone amine (blue) and substituent (grey) nitrogen-, respectively.

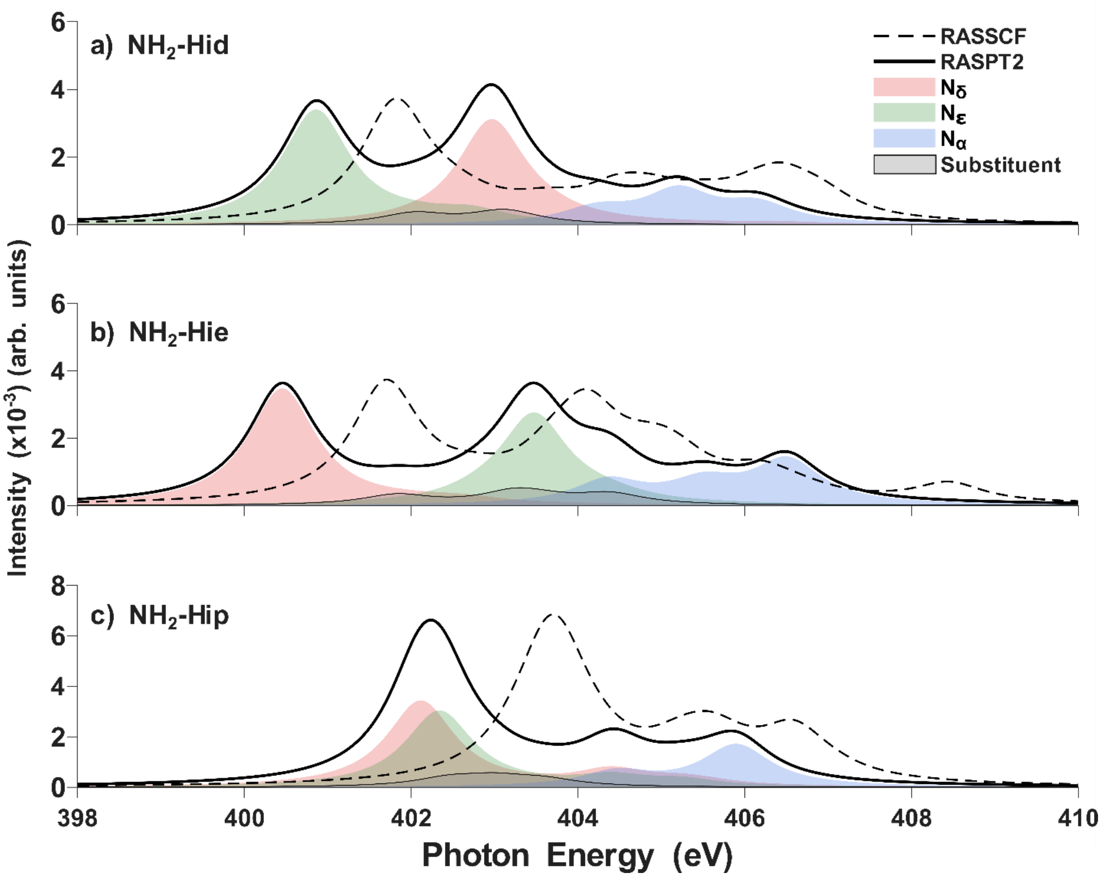

Fig. S7 Theoretical N K-edge X-ray absorption spectra of amino-substituted δ (a) and ε tautomer (b) and π histidine (c) by RASPT2 (solid line) and RASSCF (dashed) in aqueous solution, respectively. The XAS spectra by RASPT2 come along with their analysis correspond to source nitrogen-delta (red), epsilon (green), backbone amine (blue) and substituent (grey) nitrogen-, respectively.

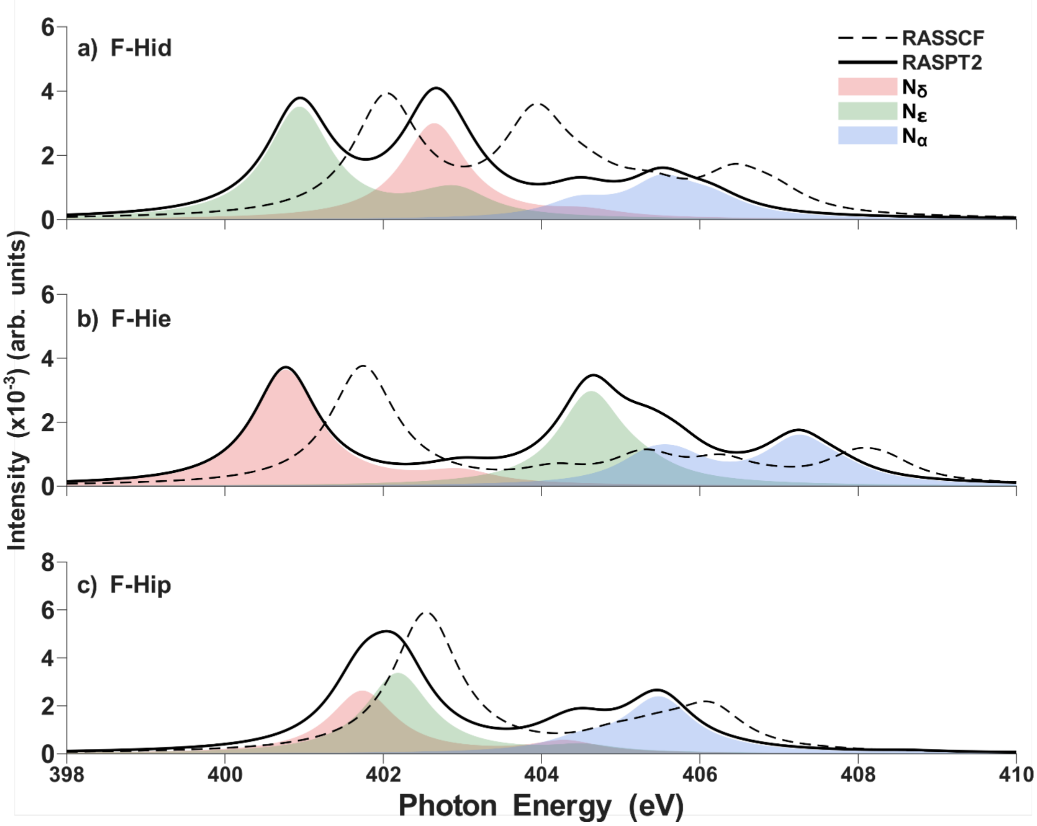

Fig. S8 Theoretical N K-edge X-ray absorption spectra of fluoro-substituted δ (a) and ε tautomer (b) and π histidine (c) by RASPT2 (solid line) and RASSCF (dashed) in aqueous solution, respectively. The XAS spectra by RASPT2 come along with their analysis correspond to source nitrogen-delta (red), epsilon (green), backbone amine (blue) and substituent (grey) nitrogen-, respectively.

# Molecular orbital transitions for the X-ray absorption spectra

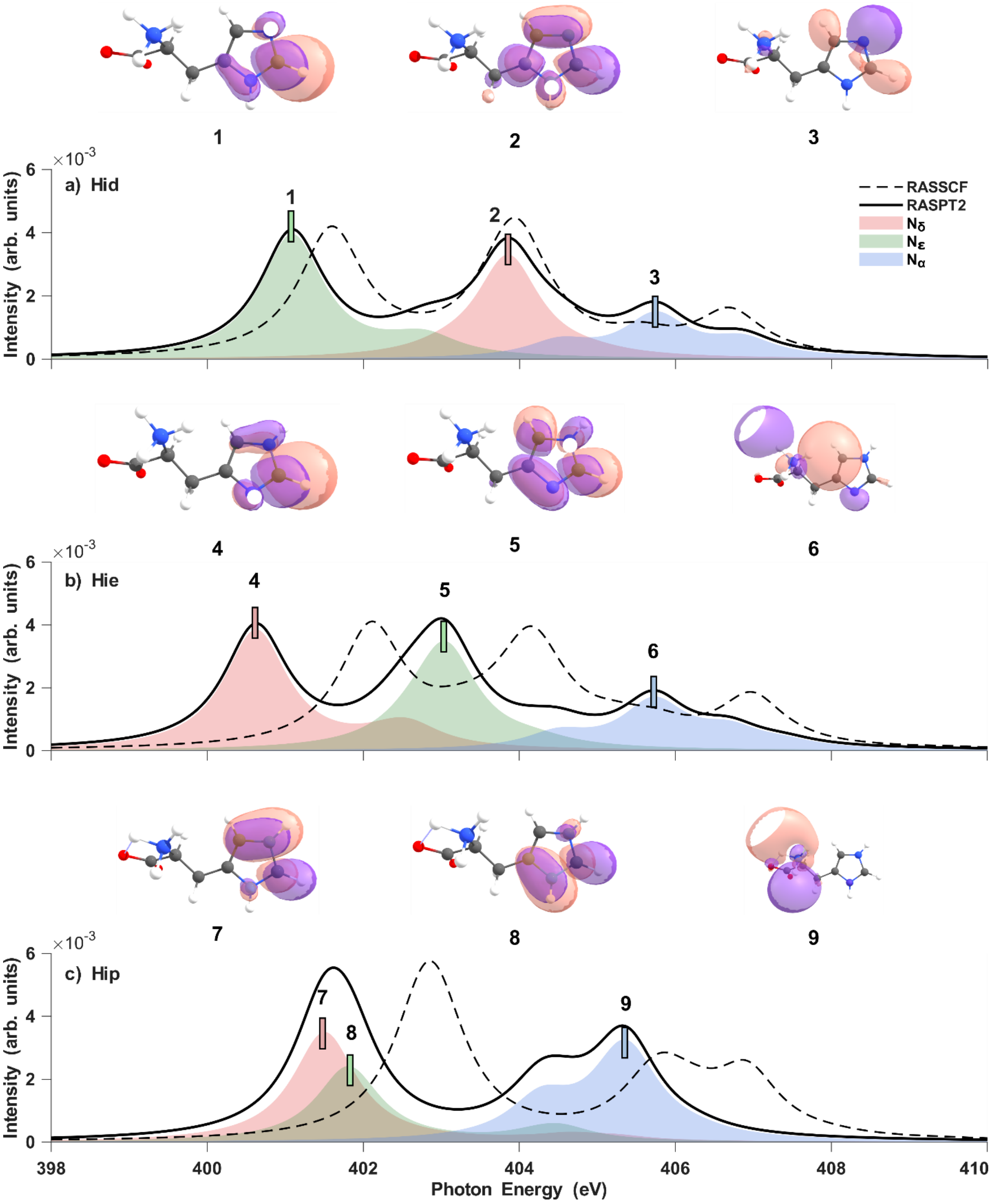

Fig. S9 Representative final orbital in the X-ray absorption transition with corresponding peak in theoretical N K-edge X-ray absorption spectra of δ (a) and ε tautomer (b) and π histidine (c) in aqueous solution, respectively. The white dots indicate the sites of excitation which remark core holes.

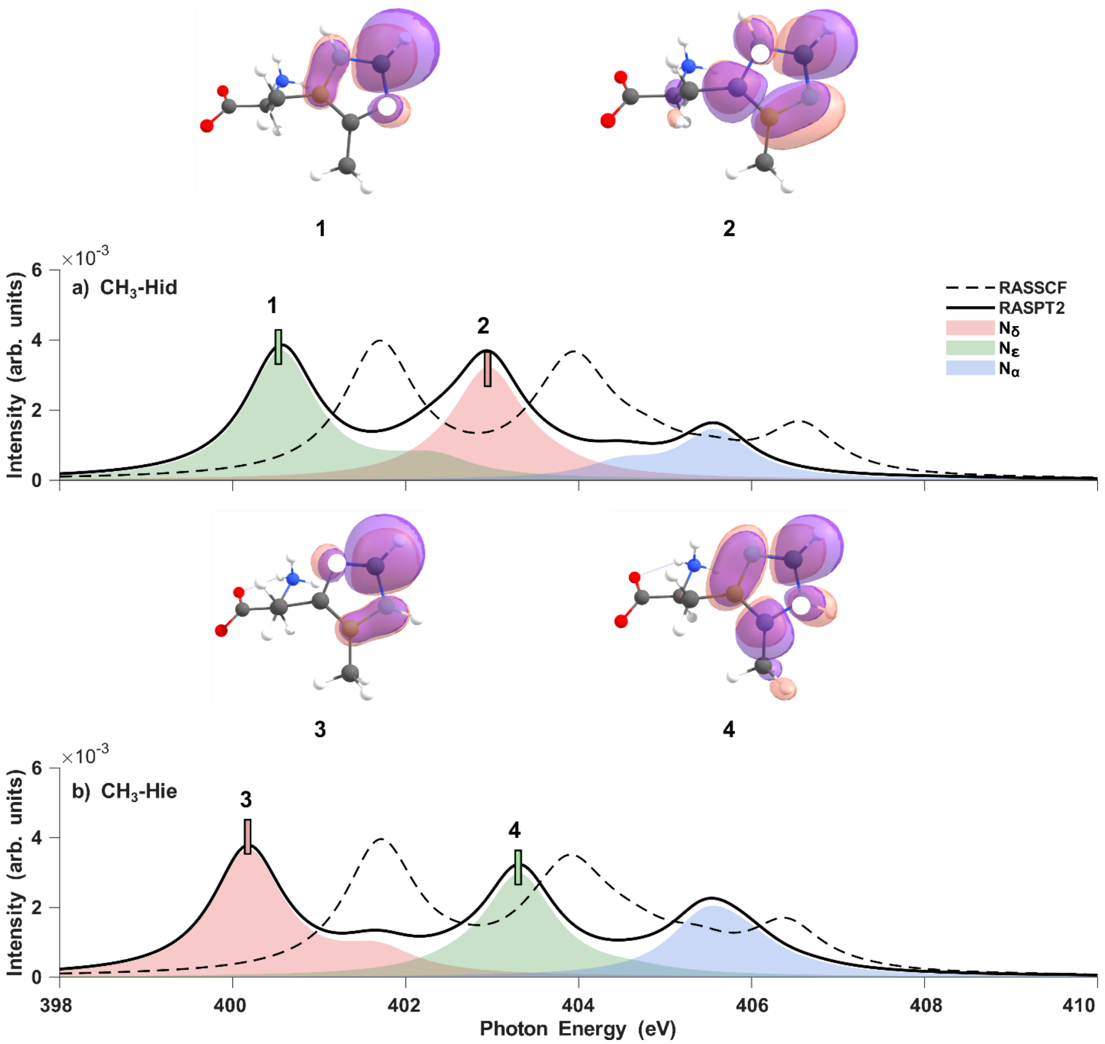

Fig. S10 Representative final orbital in the X-ray absorption transition with corresponding peak in theoretical N K-edge X-ray absorption spectra of methyl-substituted δ (a) and ε histidine (b) in aqueous solution, respectively. Peaks from backbone amine are not assigned. The white dots indicate the sites of excitation which remark core holes.

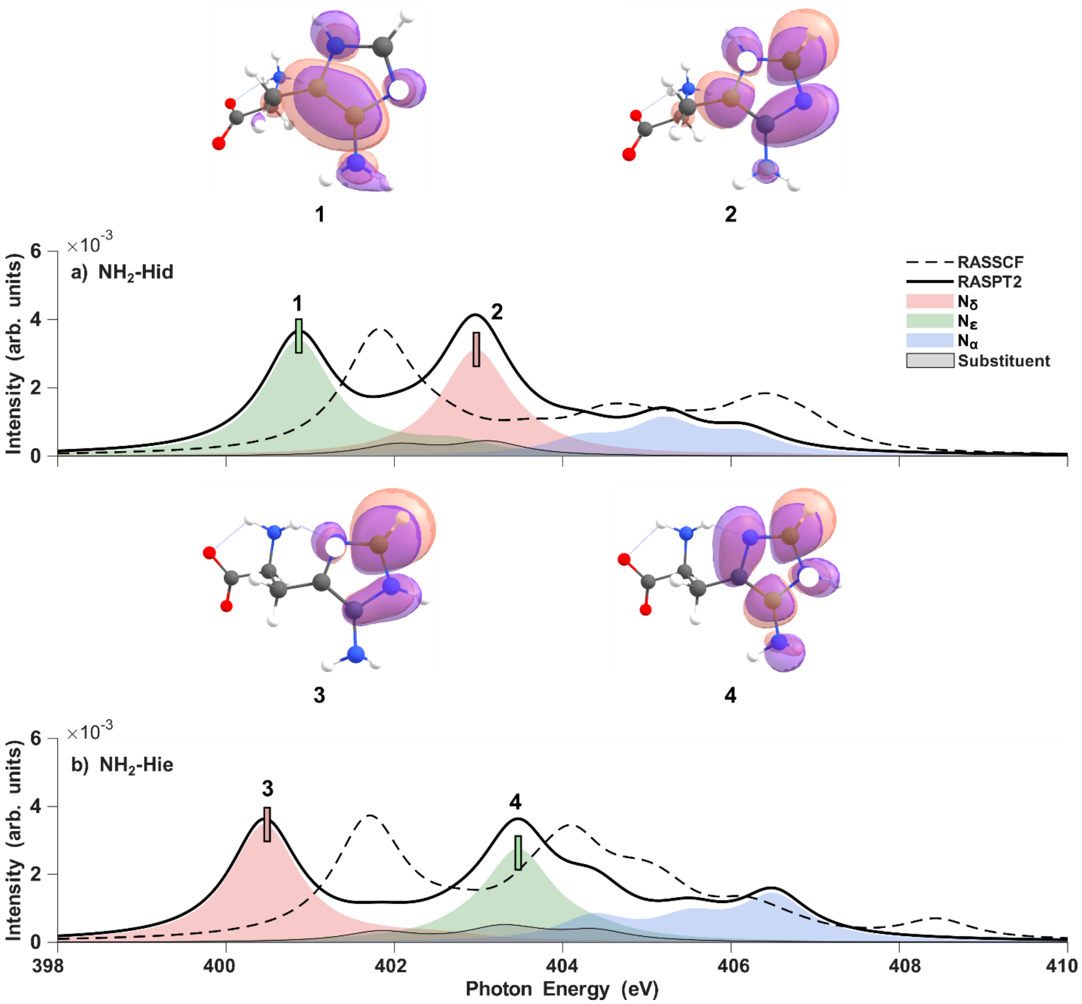

Fig. S11 Representative final orbital in the X-ray absorption transition with corresponding peak in theoretical N K-edge X-ray absorption spectra of amino-substituted δ (a) and ε histidine (b) in aqueous solution, respectively. Peaks from backbone amine are not assigned. The white dots indicate the sites of excitation which remark core holes.

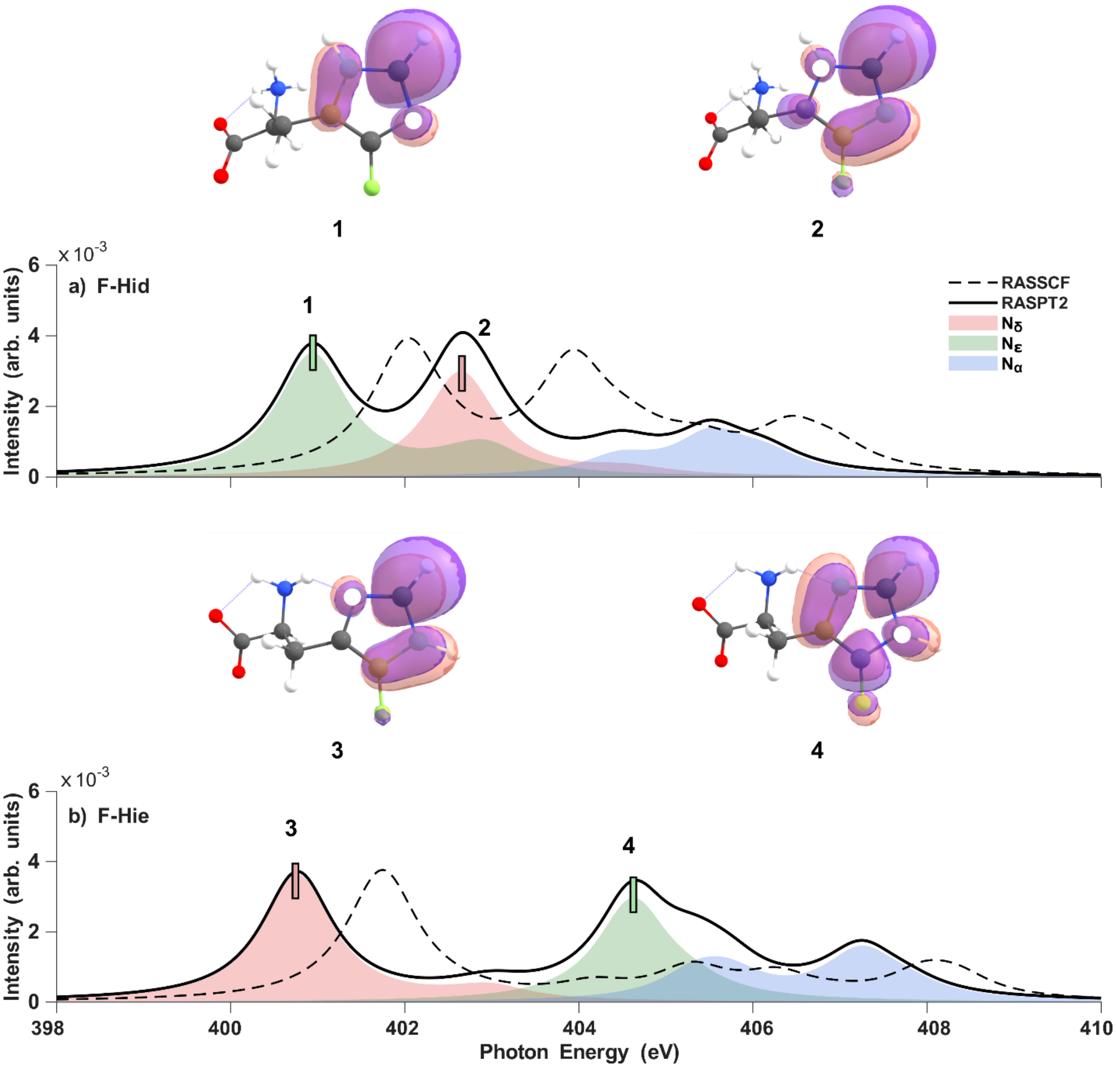

Fig. S12 Representative final orbital in the X-ray absorption transition with corresponding peak in theoretical N K-edge X-ray absorption spectra of fluoro-substituted δ (a) and ε histidine (b) in aqueous solution, respectively. Peaks from backbone amine are not assigned. The white dots indicate the sites of excitation which remark core holes.

# Reference

S1. G. L. Manni, I. F. Galván, A. Alavi, F. Aleotti, F. Aquilante, J. Autschbach, D. Avagliano, A. Baiardi, J. J. Bao, S. Battaglia, L. Birnoschi, A. Blanco-González, S. I. Bokarev, R. Broer, R. Cacciari, P. B. Calio, R. K. Carlson, R. C. Couto, L. Cerdán, L. F. Chibotaru, N. F. Chilton, J. R. Church, I. Conti, S. Coriani, J. Cuéllar-Zuquin, R. E. Daoud, N. Dattani, P. Decleva, C. de Graaf, M. G. Delcey, L. De Vico, W. Dobrautz, S. S. Dong, R. Feng, N. Ferré, M. Filatov(Gulak), L. Gagliardi, M. Garavelli, L. González, Y. Guan, M. Guo, M. R. Hennefarth, M. R. Hermes, C. E. Hoyer, M. Huix-Rotllant, V. K. Jaiswal, A. Kaiser, D. S. Kaliakin, M. Khamesian, D. S. King, V. Kochetov, M. Krośnicki, A. A. Kumaar, E. D. Larsson, S. Lehtola, M.-B. Lepetit, H, Lischka, P. L. Ríos, M. Lundberg, D. Ma, S. Mai, P. Marquetand, I. C. D. Merritt, F. Montorsi, M. Mörchen, A. Nenov, V. H. A. Nguyen, Y. Nishimoto, M. S. Oakley, M. Olivucci, M. Oppel, D. Padula, R. Pandharkar, Q. M. Phung, F. Plasser, G. Raggi, E. Rebolini, M. Reiher, I. Rivalta, D. Roca-Sanjuán, T. Romig, A. A. Safari, A. Sánchez-Mansilla, A. M. Sand, I. Schapiro, T. R. Scott, J. Segarra-Martí, F. Segatta, D.-C. Sergentu, P. Sharma, R. Shepard, Y. Shu, J. K. Staab, T. P. Straatsma, L. K. Sørensen, B. N. C. Tenorio, D. G. Truhlar, L. Ungur, M. Vacher, V. Veryazov, T. A. Voß, O, Weser, D. Wu, X. Yang, D. Yarkony, C. Zhou, J. P. Zobel, R. Lindh, *J. Chem. Theory Comput.*, 2023, **19**, 6933-6991

S2. B. Mennucci, J. Tomasi, R. Cammi, J. R. Cheeseman, M. J. Frisch, F. J. Devlin, S. Gabriel, P. J. Stephens, *J. Phys. Chem. A*, 2002, **106**, 6102-6113

S3. T. H. Dunning, *J. Chem. Phys*., 1989, **90**, 1007-1023

S4. D. E. Woon, T. H. Dunning, *J. Chem. Phys.*, 1995, **103**, 4572-4585

S5. F. Neese, *WIRES Comput. Molec. Sci.*, 2022, **12**, e1606