**Appendix. Supplementary material**

**Supplementary Tables**

**Table A.1.** The calculated quantum chemical parameters and molecular volume for neutral amino acids at DFT/B3LYP/6-311+G(d) level of theory in gas and aqueous phases, including the experimental values of the IE.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Amino acid** | **Phase a** | **EHOMO**  **(eV)** | **ELUMO**  **(eV)** | **∆E**  **(eV)** | **χ**  **(eV)** | **η**  **(eV)** | **ΔN** | **μ**  **(D)** | **TNCMP** | **TNCNBO** | **MV**  **(cm3 mol-1)** | **IE**  **(%)** |
| **Val** | G | -6.714 | -0.432 | 6.282 | 3.573 | 3.141 | 0.144 | 2.208 | -3.607 | -3.603 | 90.178 | **-20** |
| A | -6.941 | -0.484 | 6.458 | 3.713 | 3.229 | 0.119 | 3.278 | -3.815 | -3.670 | 98.899 |
| **Gly** | G | -6.871 | -0.463 | 6.408 | 3.667 | 3.204 | 0.127 | 2.135 | -1.771 | -2.399 | 70.622 | **-1** |
| A | -7.090 | -0.394 | 6.696 | 3.742 | 3.348 | 0.110 | 3.335 | -1.969 | -2.465 | 64.203 |
| **Ala** | G | -6.997 | -0.621 | 6.376 | 3.809 | 3.188 | 0.105 | 4.537 | -2.204 | -2.782 | 61.154 | **25** |
| A | -7.046 | -0.441 | 6.605 | 3.744 | 3.303 | 0.111 | 3.331 | -2.494 | -2.884 | 80.129 |
| **Asp** | G | -7.163 | -0.693 | 6.470 | 3.928 | 3.235 | 0.085 | 3.272 | -2.316 | -3.967 | 100.643 | **28** |
| A | -7.161 | -0.781 | 6.380 | 3.971 | 3.190 | 0.080 | 3.249 | -2.750 | -4.076 | 86.910 |
| **Pro** | G | -6.425 | -0.636 | 5.790 | 3.531 | 2.895 | 0.164 | 4.073 | -2.479 | -3.011 | 82.565 | **30** |
| A | -6.468 | -0.487 | 5.981 | 3.478 | 2.991 | 0.168 | 5.267 | -2.609 | -3.101 | 86.689 |
| **Leu** | G | -6.782 | -0.341 | 6.441 | 3.562 | 3.220 | 0.143 | 2.119 | -3.975 | -3.974 | 131.011 | **34** |
| A | -6.980 | -0.461 | 6.519 | 3.721 | 3.260 | 0.116 | 3.078 | -4.142 | -4.043 | 137.813 |
| **Phe** | G | -6.648 | -0.619 | 6.029 | 3.634 | 3.015 | 0.140 | 1.805 | -4.340 | -3.646 | 113.645 | **36** |
| A | -6.742 | -0.647 | 6.095 | 3.695 | 3.048 | 0.129 | 2.736 | -4.709 | -3.759 | 149.325 |
| **Glu** | G | -7.009 | -0.501 | 6.507 | 3.755 | 3.254 | 0.111 | 1.283 | -2.841 | -4.395 | 99.332 | **38** |
| A | -7.080 | -0.594 | 6.486 | 3.837 | 3.243 | 0.099 | 3.663 | -2.934 | -4.477 | 106.514 |
| **Gln** | G | -6.890 | -0.435 | 6.455 | 3.663 | 3.228 | 0.127 | 2.674 | -3.210 | -4.520 | 126.766 | **40** |
| A | -7.049 | -0.521 | 6.528 | 3.785 | 3.264 | 0.106 | 5.046 | -3.540 | -4.619 | 123.841 |
| **Asn** | G | -6.989 | -0.595 | 6.395 | 3.792 | 3.197 | 0.108 | 5.045 | -2.617 | -4.105 | 89.314 | **45** |
| A | -7.054 | -0.592 | 6.462 | 3.823 | 3.231 | 0.102 | 5.330 | -2.995 | -4.216 | 86.704 |
| **Arg** | G | -6.140 | -0.592 | 5.547 | 3.366 | 2.774 | 0.201 | 2.298 | -4.376 | -5.422 | 127.300 | **46** |
| A | -6.276 | -0.512 | 5.764 | 3.394 | 2.882 | 0.188 | 4.993 | -4.850 | -5.578 | 119.328 |
| **The** | G | -6.746 | -0.416 | 6.330 | 3.581 | 3.165 | 0.142 | 4.137 | -2.815 | -3.580 | 98.049 | **50** |
| A | -7.081 | -0.406 | 6.675 | 3.744 | 3.338 | 0.110 | 5.829 | -3.152 | -3.689 | 99.067 |
| **Lys** | G | -6.612 | -0.347 | 6.265 | 3.480 | 3.133 | 0.160 | 2.973 | -3.861 | -4.387 | 118.493 | **56** |
| A | -6.612 | -0.347 | 6.265 | 3.480 | 3.132 | 0.160 | 2.974 | -3.862 | -4.489 | 107.991 |
| **Cys** | G | -6.679 | -0.757 | 5.922 | 3.718 | 2.961 | 0.129 | 2.533 | -1.988 | -2.782 | 103.522 | **58** |
| A | -6.793 | -0.691 | 6.102 | 3.742 | 3.051 | 0.121 | 2.318 | -2.303 | -2.867 | 76.831 |
| **Try** | G | -5.714 | -0.707 | 5.007 | 3.211 | 2.504 | 0.254 | 2.355 | -5.720 | -4.211 | 130.686 | **67** |
| A | -5.955 | -0.906 | 5.049 | 3.431 | 2.525 | 0.208 | 8.112 | -5.720 | -4.377 | 121.150 |
| **Met** | G | -6.186 | -0.524 | 5.662 | 3.355 | 2.831 | 0.199 | 2.269 | -3.300 | -3.818 | 128.269 | **80** |
| A | -6.274 | -0.562 | 5.712 | 3.418 | 2.856 | 0.186 | 3.329 | -3.532 | -3.877 | 99.435 |
| **Tyr** | G | -6.173 | -0.638 | 5.535 | 3.405 | 2.767 | 0.194 | 2.648 | -5.121 | -4.223 | 126.487 | **82** |
| A | -6.256 | -0.697 | 5.559 | 3.476 | 2.780 | 0.181 | 2.313 | -5.536 | -4.348 | 148.781 |

a. “G” and “A” letters denote the gas and aqueous phases, respectively.

**EHOMO** (eV): Energy of the highest occupied molecular orbital.

**ELUMO** (eV): Energy of the lowest unoccupied molecular orbital.

**ΔE** (eV): Gap energy.

**η** (eV): Hardness.

**χ** (eV): Electronegativity.

**ΔN**: Electron-donating ability.

**μ** (D): Dipole moment.

**TNCMP**: Total negative charge obtained from *Mulliken* population analysis.

**TNCNBO**: Total negative charge obtained from Natural bond orbital analysis.

**MV** (cm3 mol-1): Molecular volume.

**Table A.2.** The calculated quantum chemical parameters and molecular volume for protonated amino acids at DFT/B3LYP/6-311+G(d) level of theory in gas and aqueous phases, including the experimental values of the IE.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Amino acid** | **Phase** | **EHOMO**  **(eV)** | **ELUMO**  **(eV)** | **∆E**  **(eV)** | **χ**  **(eV)** | **η**  **(eV)** | **ΔN** | **μ**  **(D)** | **TNCMP** | **TNCNBO** | **MV (cm3.mol-1)** | **IE**  **(%)** |
| **(Val-H)+** | G | -12.820 | -5.098 | 7.722 | 8.959 | 3.861 | -0.580 | 5.211 | -3.614 | -3.398 | 106.338 | **-20** |
| A | -8.881 | -1.199 | 7.682 | 5.040 | 3.841 | -0.073 | 7.122 | -3.235 | -2.761 | 101.022 |
| **(Gly-H)+** | G | -13.432 | -5.495 | 7.937 | 9.463 | 3.968 | -0.628 | 5.828 | -1.739 | -2.172 | 57.518 | **-1** |
| A | -8.972 | -1.051 | 7.922 | 5.012 | 3.961 | -0.067 | 7.526 | -1.148 | -1.545 | 45.086 |
| **(Ala-H)+** | G | -13.034 | -5.329 | 7.705 | 9.182 | 3.852 | -0.610 | 7.531 | -2.184 | -2.572 | 60.030 | **25** |
| A | -8.934 | -1.214 | 7.720 | 5.074 | 3.860 | -0.077 | 9.567 | -1.789 | -1.967 | 59.856 |
| **(Asp-H)+** | G | -11.869 | -5.685 | 6.184 | 8.777 | 3.092 | -0.695 | 9.222 | -2.564 | -3.739 | 77.992 | **28** |
| A | -8.555 | -1.437 | 7.118 | 4.996 | 3.559 | -0.072 | 12.127 | -2.109 | -3.176 | 85.487 |
| **(Pro-H)+** | G | -12.706 | -5.070 | 7.636 | 8.888 | 3.818 | -0.577 | 6.172 | -2.527 | -2.824 | 93.617 | **30** |
| A | -8.933 | -1.153 | 7.780 | 5.043 | 3.890 | -0.072 | 7.703 | -1.730 | -2.169 | 78.719 |
| **(Leu-H)+** | G | -12.547 | -5.115 | 7.432 | 8.831 | 3.716 | -0.585 | 6.445 | -3.949 | -3.772 | 106.467 | **34** |
| A | -8.928 | -1.339 | 7.589 | 5.134 | 3.794 | -0.086 | 10.155 | -3.488 | -3.129 | 95.315 |
| **(Phe-H)+** | G | -10.070 | -5.094 | 4.976 | 7.582 | 2.488 | -0.623 | 9.215 | -4.706 | -3.393 | 129.649 | **36** |
| A | -7.186 | -1.257 | 5.929 | 4.221 | 2.964 | 0.044 | 12.271 | -4.426 | -2.818 | 124.915 |
| **(Glu-H)+** | G | -11.094 | -5.333 | 5.761 | 8.214 | 2.880 | -0.648 | 10.100 | -2.986 | -4.138 | 103.948 | **38** |
| A | -8.270 | -1.177 | 7.092 | 4.723 | 3.546 | -0.034 | 12.750 | -3.165 | -4.243 | 91.212 |
| **(Gln-H)+** | G | -10.348 | -5.210 | 5.138 | 7.779 | 2.569 | -0.642 | 8.210 | -3.321 | -4.277 | 103.878 | **40** |
| A | -7.602 | -1.222 | 6.380 | 4.412 | 3.190 | 0.011 | 10.720 | -3.006 | -3.707 | 97.692 |
| **(Asn-H)+** | G | -12.082 | -5.031 | 7.051 | 8.557 | 3.525 | -0.578 | 2.120 | -2.910 | -3.911 | 93.700 | **45** |
| A | -8.270 | -1.166 | 7.104 | 4.718 | 3.552 | -0.033 | 3.385 | -3.006 | -3.995 | 103.223 |
| **(Arg-2H)2+** | G | -14.650 | -7.484 | 7.166 | 11.067 | 3.583 | -0.919 | 5.185 | -5.012 | -2.963 | 110.265 | **46** |
| A | -8.116 | -1.277 | 6.839 | 4.697 | 3.420 | -0.032 | 6.442 | -3.947 | -3.656 | 133.608 |
| **(The-H)+** | G | -12.544 | -5.136 | 7.409 | 8.840 | 3.704 | -0.588 | 3.559 | -2.729 | -3.409 | 87.822 | **50** |
| A | -8.597 | -1.202 | 7.395 | 4.899 | 3.697 | -0.057 | 5.334 | -2.287 | -2.772 | 83.834 |
| **(Lys-2H)2+** | G | -15.197 | -7.576 | 7.622 | 11.387 | 3.811 | -0.906 | 10.814 | -4.159 | -4.016 | 98.876 | **56** |
| A | -8.951 | -1.253 | 7.697 | 5.102 | 3.849 | -0.081 | 12.205 | -4.167 | -4.027 | 113.251 |
| **(Cys-H)+** | G | -10.576 | -5.359 | 5.217 | 7.967 | 2.609 | -0.668 | 7.173 | -2.018 | -2.539 | 85.137 | **58** |
| A | -7.227 | -1.362 | 5.864 | 4.295 | 2.932 | 0.032 | 9.926 | -1.757 | -1.907 | 69.930 |
| **(Try-H)+** | G | -8.720 | -4.917 | 3.803 | 6.818 | 1.901 | -0.615 | 9.230 | -5.858 | -3.977 | 167.472 | **67** |
| A | -5.996 | -1.215 | 4.781 | 3.606 | 2.390 | 0.183 | 12.642 | -3.604 | -2.109 | 128.723 |
| **(Met-H)+** | G | -9.164 | -5.225 | 3.940 | 7.195 | 1.970 | -0.689 | 8.565 | -3.426 | -3.614 | 127.025 | **80** |
| A | -6.414 | -1.241 | 5.173 | 3.828 | 2.586 | 0.126 | 10.782 | -3.036 | -2.967 | 102.357 |
| **(Tyr-H)+** | G | -9.101 | -5.503 | 3.598 | 7.302 | 1.799 | -0.784 | 13.405 | -5.012 | -3.998 | 122.796 | **82** |
| A | -6.521 | -1.251 | 5.270 | 3.886 | 2.635 | 0.113 | 17.300 | -3.716 | -2.732 | 154.870 |

**Table A.3.** The linear coefficient of determinations (R2) between calculated descriptors for neutral and protonated amino acids (in gas and aqueous phases) with the experimental IE.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Forms** | **Phase** | **EHOMO** | **ELUMO** | **∆E** | **χ** | **η** | **ΔN** | **μ** | **TNCMP** | **TNCNBO** | **MV** | **Eads** |
| Neutral | G | 0.30 | 0.08 | 0.32 | 0.22 | 0.32 | 0.27 | 0.00 | 0.18 | 0.15 | 0.40 | 0.31 |
| A | 0.37 | 0.09 | 0.40 | 0.24 | 0.40 | 0.31 | 0.02 | 0.19 | 0.15 | 0.15 | 0.26 |
| Protonated | G | 0.26 | 0.02 | 0.49 | 0.12 | 0.49 | 0.15 | 0.18 | 0.18 | 0.16 | 0.27 | 0.23 |
| A | 0.51 | 0.09 | 0.53 | 0.49 | 0.53 | 0.48 | 0.19 | 0.12 | 0.03 | 0.28 | 0.09 |

**Table A.4.** Obtained adsorption energies (Eads, in Kcal mol-1) of studied amino acids at neutral and protonated states on Cu(111) surface with and without 40 molecules of water.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Amino acid** | **Phase** | **State of protonation** | | **IE (%)** |
| **Neutral** | **Protonated** |
| **Val** | G | -41.307 | -36.327 | **-20** |
| A | -73.941 | -161.388 |
| **Gly** | G | -24.424 | -24.636 | **-1** |
| A | -57.662 | -148.368 |
| **Ala** | G | -27.960 | -28.660 | **25** |
| A | -61.297 | -152.460 |
| **Asp** | G | -38.189 | -40.205 | **28** |
| A | -93.461 | -175.900 |
| **Pro** | G | -36.360 | -55.948 | **30** |
| A | -64.342 | -161.416 |
| **Leu** | G | -39.513 | -40.208 | **34** |
| A | -73.848 | -155.816 |
| **Phe** | G | -48.142 | -46.164 | **36** |
| A | -83.794 | -163.558 |
| **Glu** | G | -45.040 | -69.282 | **38** |
| A | -94.601 | -200.15 |
| **Gln** | G | -45.610 | -61.111 | **40** |
| A | -91.632 | -187.194 |
| **Asn** | G | -44.526 | -40.332 | **45** |
| A | -90.466 | -161.211 |
| **Arg** | G | -60.024 | -59.620 | **46** |
| A | -112.987 | -194.911 |
| **The** | G | -33.565 | -33.566 | **50** |
| A | -69.648 | -151.596 |
| **Lys** | G | -49.211 | -54.547 | **56** |
| A | -88.352 | -289.624 |
| **Cys** | G | -41.677 | -36.175 | **58** |
| A | -77.725 | -158.186 |
| **Try** | G | -53.960 | -55.403 | **67** |
| A | -93.101 | -179.467 |
| **Met** | G | -46.165 | -58.484 | **80** |
| A | -87.445 | -177.890 |
| **Tyr** | G | -51.153 | -50.220 | **82** |
| A | -95.241 | -180.416 |

**Table A.5.** Correlation (Pearson) matrix between different obtained descriptors in aqueous phase (N and P subscripts denote the neutral and protonated molecules forms, respectively). Values in bold are different from 0 with a significance level alpha=0.05.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | **EHOMO(N)** | **ELUMO(N)** | **ΔE(N)** | **η(N)** | **χ(N)** | **ΔN(N)** | **μ(N)** | **TNCMP(N)** | **TNCNBO(N)** | **MV(N)** | **EAds(N)** | **EHOMO(P)** | **ELUMO(P)** | **ΔE(P)** | **η(P)** | **χ(P)** | **ΔN(P)** | **μ(P)** | **TNCMP(P)** | **TNCNBO(P)** | **MV(P)** | **EAds(P)** |
| **EHOMO(N)** | **1** | -0.388 | **-0.954** | **-0.954** | **-0.920** | **0.982** | 0.296 | **-0.693** | -0.339 | 0.392 | -0.413 | **0.668** | -0.025 | **-0.659** | **-0.659** | **-0.673** | **0.715** | 0.348 | -0.436 | 0.144 | **0.653** | -0.254 |
| **ELUMO(N)** | -0.388 | **1** | **0.645** | **0.645** | -0.003 | -0.215 | -0.226 | 0.402 | 0.201 | -0.243 | 0.475 | **-0.721** | 0.412 | **0.744** | **0.744** | **0.693** | **-0.691** | -0.463 | 0.164 | -0.106 | -0.401 | -0.160 |
| **ΔE(N)** | **-0.954** | **0.645** | **1** | **1.000** | **0.762** | **-0.884** | -0.318 | **0.705** | 0.346 | -0.403 | **0.496** | **-0.788** | 0.154 | **0.787** | **0.787** | **0.783** | **-0.816** | -0.439 | 0.415 | -0.153 | **-0.671** | 0.158 |
| **η(N)** | **-0.954** | **0.645** | **1.000** | **1** | **0.762** | **-0.884** | -0.318 | **0.705** | 0.346 | -0.403 | **0.496** | **-0.788** | 0.154 | **0.787** | **0.787** | **0.783** | **-0.816** | -0.439 | 0.415 | -0.153 | **-0.671** | 0.158 |
| **χ(N)** | **-0.920** | -0.003 | **0.762** | **0.762** | **1** | **-0.974** | -0.225 | **0.581** | 0.283 | -0.322 | 0.247 | -0.419 | -0.148 | 0.399 | 0.399 | 0.437 | **-0.483** | -0.182 | 0.404 | -0.111 | **-0.539** | 0.343 |
| **ΔN(N)** | **0.982** | -0.215 | **-0.884** | **-0.884** | **-0.974** | **1** | 0.304 | **-0.660** | -0.322 | 0.361 | -0.337 | **0.566** | 0.077 | **-0.550** | **-0.550** | **-0.579** | **0.627** | 0.270 | -0.421 | 0.140 | **0.606** | -0.291 |
| **μ(N)** | 0.296 | -0.226 | -0.318 | -0.318 | -0.225 | 0.304 | **1** | -0.234 | -0.282 | -0.015 | -0.156 | 0.174 | 0.311 | -0.144 | -0.144 | -0.203 | 0.240 | -0.348 | -0.019 | -0.012 | 0.132 | 0.083 |
| **TNCMP(N)** | **-0.693** | 0.402 | **0.705** | **0.705** | **0.581** | **-0.660** | -0.234 | **1** | **0.652** | **-0.856** | **0.589** | **-0.583** | 0.182 | **0.588** | **0.588** | **0.573** | **-0.594** | -0.452 | **0.834** | 0.188 | **-0.924** | 0.252 |
| **TNCNBO(N)** | -0.339 | 0.201 | 0.346 | 0.346 | 0.283 | -0.322 | -0.282 | **0.652** | **1** | **-0.608** | **0.906** | -0.254 | 0.274 | 0.273 | 0.273 | 0.233 | -0.206 | -0.181 | **0.761** | **0.782** | **-0.765** | **0.525** |
| **MV(N)** | 0.392 | -0.243 | -0.403 | -0.403 | -0.322 | 0.361 | -0.015 | **-0.856** | **-0.608** | **1** | **-0.499** | 0.445 | -0.258 | -0.459 | -0.459 | -0.427 | 0.412 | **0.537** | **-0.838** | -0.343 | **0.818** | -0.208 |
| **EAds(N)** | -0.413 | 0.475 | **0.496** | **0.496** | 0.247 | -0.337 | -0.156 | **0.589** | **0.906** | **-0.499** | **1** | -0.479 | 0.392 | **0.504** | **0.504** | 0.450 | -0.407 | -0.305 | **0.668** | **0.669** | **-0.755** | 0.467 |
| **EHOMO(P)** | **0.668** | **-0.721** | **-0.788** | **-0.788** | -0.419 | **0.566** | 0.174 | **-0.583** | -0.254 | 0.445 | -0.479 | **1** | -0.166 | **-0.996** | **-0.996** | **-0.996** | **0.986** | **0.511** | -0.363 | 0.103 | **0.572** | 0.026 |
| **ELUMO(P)** | -0.025 | 0.412 | 0.154 | 0.154 | -0.148 | 0.077 | 0.311 | 0.182 | 0.274 | -0.258 | 0.392 | -0.166 | **1** | 0.249 | 0.249 | 0.080 | -0.054 | -0.349 | 0.210 | 0.168 | -0.227 | 0.122 |
| **ΔE(P)** | **-0.659** | **0.744** | **0.787** | **0.787** | 0.399 | **-0.550** | -0.144 | **0.588** | 0.273 | -0.459 | **0.504** | **-0.996** | 0.249 | **1** | **1.000** | **0.985** | **-0.973** | **-0.532** | 0.375 | -0.087 | **-0.581** | -0.015 |
| **η(P)** | **-0.659** | **0.744** | **0.787** | **0.787** | 0.399 | **-0.550** | -0.144 | **0.588** | 0.273 | -0.459 | **0.504** | **-0.996** | 0.249 | **1.000** | **1** | **0.985** | **-0.973** | **-0.532** | 0.375 | -0.087 | **-0.581** | -0.015 |
| **χ(P)** | **-0.673** | **0.693** | **0.783** | **0.783** | 0.437 | **-0.579** | -0.203 | **0.573** | 0.233 | -0.427 | 0.450 | **-0.996** | 0.080 | **0.985** | **0.985** | **1** | **-0.992** | **-0.486** | 0.349 | -0.119 | **-0.558** | -0.037 |
| **ΔN(P)** | **0.715** | **-0.691** | **-0.816** | **-0.816** | **-0.483** | **0.627** | 0.240 | **-0.594** | -0.206 | 0.412 | -0.407 | **0.986** | -0.054 | **-0.973** | **-0.973** | **-0.992** | **1** | **0.498** | -0.337 | 0.169 | **0.550** | 0.028 |
| **μ(P)** | 0.348 | -0.463 | -0.439 | -0.439 | -0.182 | 0.270 | -0.348 | -0.452 | -0.181 | **0.537** | -0.305 | **0.511** | -0.349 | **-0.532** | **-0.532** | **-0.486** | **0.498** | **1** | -0.358 | -0.013 | 0.428 | -0.350 |
| **TNCMP(P)** | -0.436 | 0.164 | 0.415 | 0.415 | 0.404 | -0.421 | -0.019 | **0.834** | **0.761** | **-0.838** | **0.668** | -0.363 | 0.210 | 0.375 | 0.375 | 0.349 | -0.337 | -0.358 | **1** | **0.586** | **-0.876** | **0.520** |
| **TNCNBO(P)** | 0.144 | -0.106 | -0.153 | -0.153 | -0.111 | 0.140 | -0.012 | 0.188 | **0.782** | -0.343 | **0.669** | 0.103 | 0.168 | -0.087 | -0.087 | -0.119 | 0.169 | -0.013 | **0.586** | **1** | -0.420 | **0.581** |
| **MV(P)** | **0.653** | -0.401 | **-0.671** | **-0.671** | **-0.539** | **0.606** | 0.132 | **-0.924** | **-0.765** | **0.818** | **-0.755** | **0.572** | -0.227 | **-0.581** | **-0.581** | **-0.558** | **0.550** | 0.428 | **-0.876** | -0.420 | **1** | -0.377 |
| **EAds(P)** | -0.254 | -0.160 | 0.158 | 0.158 | 0.343 | -0.291 | 0.083 | 0.252 | **0.525** | -0.208 | 0.467 | 0.026 | 0.122 | -0.015 | -0.015 | -0.037 | 0.028 | -0.350 | **0.520** | **0.581** | -0.377 | **1** |

**Table A.6.** Pearson correlation matrix between screened descriptors (N and P subscripts denote the neutral and protonated molecules forms, respectively).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | **ELUMO(N)** | **ΔE(N)** | **χ(N)** | **μ(N)** | **TNCMP(N)** | **TNCNBO(N)** | **MV(N)** | **ELUMO(P)** | **ΔN(P)** | **μ(P)** | **TNCMP(P)** | **TNCNBO(P)** | **EAds(P)** |
| **ELUMO(N)** | 1 | 0.645 | -0.003 | -0.226 | 0.402 | 0.201 | -0.243 | 0.412 | -0.691 | -0.463 | 0.164 | -0.106 | -0.160 |
| **ΔE(N)** | 0.645 | 1 | 0.762 | -0.318 | 0.705 | 0.346 | -0.403 | 0.154 | -0.816 | -0.439 | 0.415 | -0.153 | 0.158 |
| **χ(N)** | -0.003 | 0.762 | 1 | -0.225 | 0.581 | 0.283 | -0.322 | -0.148 | -0.483 | -0.182 | 0.404 | -0.111 | 0.343 |
| **μ(N)** | -0.226 | -0.318 | -0.225 | 1 | -0.234 | -0.282 | -0.015 | 0.311 | 0.240 | -0.348 | -0.019 | -0.012 | 0.083 |
| **TNCMP(N)** | 0.402 | 0.705 | 0.581 | -0.234 | 1 | 0.652 | -0.856 | 0.182 | -0.594 | -0.452 | 0.834 | 0.188 | 0.252 |
| **TNCNBO(N)** | 0.201 | 0.346 | 0.283 | -0.282 | 0.652 | 1 | -0.608 | 0.274 | -0.206 | -0.181 | 0.761 | 0.782 | 0.525 |
| **MV(N)** | -0.243 | -0.403 | -0.322 | -0.015 | -0.856 | -0.608 | 1 | -0.258 | 0.412 | 0.537 | -0.838 | -0.343 | -0.208 |
| **ELUMO(P)** | 0.412 | 0.154 | -0.148 | 0.311 | 0.182 | 0.274 | -0.258 | 1 | -0.054 | -0.349 | 0.210 | 0.168 | 0.122 |
| **ΔN(P)** | -0.691 | -0.816 | -0.483 | 0.240 | -0.594 | -0.206 | 0.412 | -0.054 | 1 | 0.498 | -0.337 | 0.169 | 0.028 |
| **μ(P)** | -0.463 | -0.439 | -0.182 | -0.348 | -0.452 | -0.181 | 0.537 | -0.349 | 0.498 | 1 | -0.358 | -0.013 | -0.350 |
| **TNCMP(P)** | 0.164 | 0.415 | 0.404 | -0.019 | 0.834 | 0.761 | -0.838 | 0.210 | -0.337 | -0.358 | 1 | 0.586 | 0.520 |
| **TNCNBO(P)** | -0.106 | -0.153 | -0.111 | -0.012 | 0.188 | 0.782 | -0.343 | 0.168 | 0.169 | -0.013 | 0.586 | 1 | 0.581 |
| **EAds(P)** | -0.160 | 0.158 | 0.343 | 0.083 | 0.252 | 0.525 | -0.208 | 0.122 | 0.028 | -0.350 | 0.520 | 0.581 | 1 |

**Table A.7.** Constructed ANN model constants.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Hidden layer parameters | | | | | Output layer parameters | |
| B1 | **IW1** | | | | **B2** | **LW2** |
| -2.50843  1.51041  1.12799  0.19200  0.22118  -0.72954  1.63359  -1.80519 | 0.95197  -2.32290  -1.78348  0.03634  0.86310  -2.10223  0.71684  -0.63067 | 0.50119  0.72572  1.58773  -1.27007  -1.12347  0.07105  2.22929  0.55384 | -1.22486  1.20356  0.71697  1.80490  -1.18040  -0.00457  0.39360  -2.95345 | -1.49147  0.41833  0.03116  -1.23056  1.56171  -1.28091  -0.10080  -0.48609 | 0.37479 | -0.28472  -0.13540  -0.47945  -1.05696  -0.69874  0.48510  -0.03913  -0.82627 |

**Table A.8.** The atomic natural charges values of the rest of investigated amino acids (in neutral and protonated forms) obtained by NBO analysis at DFT/B3LYP/6-311+G(d) level in aqueous phase.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Amino acid** | **Natural charge on atomsa** | | | | | | | | | | | | | | |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
| **Val** | -0.681 | -0.646 | 0.822 | -0.113 | -0.852 | -0.230 | -0.578 | -0.570 | - | - | - | - | - | - | - |
| **(Val–H)+** | -0.655 | -0.618 | 0.806 | -0.099 | -0.666 | -0.237 | -0.584 | -0.568 | - | - | - | - | - | - | - |
| **Gly** | -0.689 | -0.641 | 0.810 | -0.284 | -0.852 | - | - | - | - | - | - | - | - | - | - |
| **(Gly–H)+** | -0.660 | -0.612 | 0.799 | -0.273 | -0.657 | - | - | - | - | - | - | - | - | - | - |
| **Ala** | -0.686 | -0.644 | 0.821 | -0.124 | -0.852 | -0.577 | - | - | - | - | - | - | - | - | - |
| **(Ala–H)+** | -0.654 | -0.605 | 0.806 | -0.121 | -0.665 | -0.588 | - | - | - | - | - | - | - | - | - |
| **Asp** | -0.631 | -0.683 | 0.817 | -0.111 | -0.842 | -0.485 | 0.817 | -0.689 | -0.634 | - | - | - | - | - | - |
| **(Asp–H)+** | -0.683 | -0.585 | 0.811 | -0.106 | -0.661 | -0.501 | 0.828 | -0.627 | -0.672 | - | - | - | - | - | - |
| **Pro** | -0.686 | -0.632 | 0.813 | -0.131 | -0.699 | -0.177 | -0.392 | -0.385 | - | - | - | - | - | - | - |
| **(Pro–H)+** | -0.652 | -0.606 | 0.813 | -0.117 | -0.550 | -0.163 | -0.400 | -0.394 | - | - | - | - | - | - | - |
| **Leu** | -0.680 | -0.646 | 0.825 | -0.114 | -0.849 | -0.384 | -0.239 | -0.561 | -0.569 | - | - | - | - | - | - |
| **(Leu–H)+** | -0.651 | -0.601 | 0.803 | -0.109 | -0.661 | -0.395 | -0.240 | -0.561 | -0.573 | - | - | - | - | - | - |
| **Phe** | -0.676 | -0.647 | 0.822 | -0.104 | -0.847 | -0.404 | -0.049 | -0.206 | -0.203 | -0.215 | -0.203 | -0.204 | - | - | - |
| **(Phe–H)+** | -0.650 | -0.616 | 0.803 | -0.093 | -0.662 | -0.405 | -0.057 | -0.196 | -0.200 | -0.204 | -0.199 | -0.199 | - | - | - |
| **Glu** | -0.691 | -0.635 | 0.821 | -0.111 | -0.846 | -0.384 | -0.480 | 0.823 | -0.690 | -0.641 | - | - | - | - | - |
| **(Glu–H)+** | -0.684 | -0.595 | 0.809 | -0.097 | -0.662 | -0.397 | -0.487 | 0.825 | -0.636 | -0.684 | - | - | - | - | - |
| **Gln** | -0.645 | -0.680 | 0.821 | -0.113 | -0.848 | -0.386 | -0.466 | 0.681 | -0.705 | -0.776 | - | - | - | - | - |
| **(Gln–H)+** | -0.686 | -0.591 | 0.805 | -0.097 | -0.663 | -0.397 | -0.467 | 0.681 | -0.695 | -0.774 | - | - | - | - | - |
| **Asn** | -0.639 | -0.679 | 0.818 | -0.112 | -0.845 | -0.469 | 0.672 | -0.701 | -0.771 | - | - | - | - | - | - |
| **(Asn–H)+** | -0.678 | -0.597 | 0.814 | -0.104 | -0.684 | -0.481 | 0.685 | -0.706 | -0.744 | - | - | - | - | - | - |
| **Arg** | -0.680 | -0.646 | 0.822 | -0.111 | -0.849 | -0.387 | -0.391 | -0.180 | -0.649 | 0.611 | -0.860 | -0.826 | - | - | - |
| **(Arg–2H)2+** | -0.656 | -0.613 | 0.805 | -0.096 | -0.662 | -0.394 | -0.392 | -0.181 | -0.580 | 0.684 | -0.744 | -0.744 | - | - | - |
| **The** | -0.692 | -0.636 | 0.828 | -0.136 | -0.849 | 0.114 | -0.602 | -0.773 | - | - | - | - | - | - | - |
| **(The–H)+** | -0.658 | -0.616 | 0.818 | -0.121 | -0.677 | 0.109 | -0.603 | -0.774 | - | - | - | - | - | - | - |
| **Lys** | -0.682 | -0.645 | 0.823 | -0.111 | -0.854 | -0.384 | -0.387 | -0.385 | -0.176 | -0.867 | - | - | - | - | - |
| **(Lys–2H)2+** | -0.656 | -0.614 | 0.805 | -0.096 | -0.661 | -0.392 | -0.382 | -0.403 | -0.166 | -0.656 | - | - | - | - | - |
| **Cys** | -0.677 | -0.643 | 0.823 | -0.132 | -0.855 | -0.494 | -0.067 | - | - | - | - | - | - | - | - |
| **(Cys–H)+** | -0.646 | -0.615 | 0.804 | -0.113 | -0.666 | -0.521 | -0.012 | - | - | - | - | - | - | - | - |
| **Try** | -0.691 | -0.629 | 0.810 | -0.123 | -0.846 | -0.405 | -0.169 | 0.019 | -0.531 | 0.150 | -0.232 | -0.217 | -0.232 | -0.200 | -0.101 |
| **(Try–H)+** | -0.650 | -0.620 | 0.804 | -0.093 | -0.663 | -0.401 | -0.142 | 0.031 | -0.530 | 0.151 | -0.232 | -0.218 | -0.232 | -0.203 | -0.096 |
| **Met** | -0.680 | -0.644 | 0.820 | -0.111 | -0.847 | -0.401 | -0.498 | 0.136 | -0.695 | - | - | - | - | - | - |
| **(Met–H)+** | -0.657 | -0.612 | 0.805 | -0.097 | -0.661 | -0.408 | -0.498 | 0.147 | -0.694 | - | - | - | - | - | - |
| **Tyr** | -0.648 | -0.676 | 0.822 | -0.105 | -0.847 | -0.401 | -0.081 | -0.182 | -0.254 | 0.304 | -0.277 | -0.183 | -0.694 | - | - |
| **(Tyr–H)+** | -0.689 | -0.584 | 0.802 | -0.091 | -0.664 | -0.404 | -0.090 | -0.175 | -0.253 | 0.313 | -0.271 | -0.175 | -0.688 | - | - |

a see **Table 1** for numbering non-hydrogen atoms.

**Table A.9.** The calculated used descriptors in QSPR modeling of designed amino acids derivatives in aqueous phase (N and P subscripts denote the neutral and protonated molecules forms, respectively).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Molecule** | **Descriptors** | | | |
| **χ(N)** | **ELUMO(P)** | **ΔN(P)** | **TNCNBO(P)** |
| Para(NH2) | 3.174 | -1.544 | -0.036 | -3.842 |
| Para(SH) | 3.450 | -1.262 | 0.136 | -3.459 |
| Para(SCH3) | 3.307 | -1.259 | 0.169 | -4.186 |
| Tyr-Meta(OH) | 3.314 | -1.169 | 0.140 | -4.730 |
| Tyr-Meta(NH2) | 3.234 | -1.437 | 0.039 | -4.479 |
| Tyr-Meta(SH) | 3.434 | -1.260 | 0.135 | -4.074 |
| Tyr-Meta(SCH3) | 3.277 | -1.237 | 0.166 | -4.806 |
| Tyr-Bi Meta(OH/OH) | 3.277 | -1.155 | 0.135 | -5.415 |
| Tyr-Bi Meta(NH2/NH2) | 3.175 | -2.096 | -0.071 | -5.015 |
| Tyr-Bi Meta(SH/SH) | 3.532 | -1.271 | 0.128 | -4.087 |

**Supplementary Figures**

|  |  |  |
| --- | --- | --- |
|  |  |  |
| (**Val–H**)+ | (**Phe–H**)+ | (**Lys–2H**)2+ |
|  |  |  |
| (**Gly–H**)+ | (**Glu–H**)+ | (**Cys–H**)+ |
|  |  |  |
| (**Ala–H**)+ | (**Gln–H**)+ | (**Try–H**)+ |
|  |  |  |
| (**Asp–H**)+ | (**Asn–H**)+ | (**Met–H**)+ |
|  |  |  |
| (**Pro–H**)+ | (**Arg–2H**)2+ | (**Tyr–H**)+ |
|  |  |  |
| (**Leu–H**)+ | (**The–H**)+ |  |

**Fig. A.1** Molecular structures of investigated full protonated α-amino acids.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Val** | **(Val–H)+** | **Gly** | | **(Gly–H)+** | **Ala** | **(Ala–H)+** | | **Asp** | **(Asp–H)+** |
|  |  |  | |  |  |  | |  |  |
| **Pro** | **(Pro–H)+** | **Leu** | | **(Leu–H)+** | **Phe** | **(Phe–H)+** | | **Glu** | **(Glu–H)+** |
|  |  |  | |  |  |  | |  |  |
| **Gln** | **(Gln–H)+** | **Asn** | | **(Asn–H)+** | **Arg** | **(Arg–2H)2+** | | **The** | **(The–H)+** |
|  |  |  | |  |  |  | |  |  |
|  |  |  | **Lys** | | **(Lys–2H)2+** | |  |  |  |
|  |  |  |  | |  | |  |  |  |

**Fig. A.2** Optimized systems obtained by MM-*Monte Carlo* simulation of neutral and protonated amino acids forms on Cu(111) surface in aqueous phase.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Val** | **(Val–H)+** | **Gly** | | **(Gly–H)+** | **Ala** | **(Ala–H)+** | | **Asp** | **(Asp–H)+** |
|  |  |  | |  |  |  | |  |  |
| **Pro** | **(Pro–H)+** | **Leu** | | **(Leu–H)+** | **Phe** | **(Phe–H)+** | | **Glu** | **(Glu–H)+** |
|  |  |  | |  |  |  | |  |  |
| **Gln** | **(Gln–H)+** | **Asn** | | **(Asn–H)+** | **Arg** | **(Arg–2H)2+** | | **The** | **(The–H)+** |
|  |  |  | |  |  |  | |  |  |
|  |  |  | **Lys** | | **(Lys–2H)2+** | |  |  |  |
|  |  |  |  | |  | |  |  |  |

**Fig. A.3** The contour representation of ESP for neutral and protonated amino acids forms at DFT/B3LYP/6-311+G(d) in aqueous phase. Regions of negative (positive) potential are red (green).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Val** | **(Val–H)+** | **Gly** | **(Gly–H)+** | **Ala** | **(Ala–H)+** | **Asp** | **(Asp–H)+** |
| **LUMO** |  |  |  |  |  |  |  |  |
| **HOMO** |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | **Pro** | **(Pro–H)+** | **Leu** | **(Leu–H)+** | **Phe** | **(Phe–H)+** | **Glu** | **(Glu–H)+** |
| **LUMO** |  |  |  |  |  |  |  |  |
| **HOMO** |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | **Gln** | **(Gln–H)+** | **Asn** | **(Asn–H)+** | **Arg** | **(Arg–2H)2-** | **The** | **(The–H)+** |
| **LUMO** |  |  |  |  |  |  |  |  |
| **HOMO** |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | **Lys** | **(Lys–2H)2+** |  |  |  |  |  |  |
| **LUMO** |  |  |  |  |  |  |  |  |
| **HOMO** |  |  |  |  |  |  |  |  |

**Fig. A.4** The HOMO and LUMO plots of neutral and protonated amino acids forms in aqueous phase at DFT/B3LYP 6-311+G(d) level of theory.