Journal of the Serbian Chemical Society
JSCS@tmf.bg.ac.rs • www.shd.org.rs/JSCS
*J. Serb. Chem. Soc. 80* (3) S74–S102 (2015) *Supplementary material*

SUPPLEMENTARY MATERIAL TO

# A quantitative structure–activity relationship (QSAR) study of peptide drugs based on a new descriptor of amino acids

JIAN-BO TONG*, JIA CHANG, SHU-LING LIU and MIN BAI

*Shaanxi University of Science and Technology, Xi'an 710021, China*

*J. Serb. Chem. Soc.* 80 (3) (2015) 343–353

TABLE S-I. The meanings of 99 WHIM descriptors

| ID | Symbol | Meaning |
|---|---|---|
| 1 | L1u | 1st component size directional WHIM index/unweighted |
| 2 | L2u | 2nd component size directional WHIM index/unweighted |
| 3 | L3u | 3rd component size directional WHIM index/unweighted |
| 4 | P1u | 1st component shape directional WHIM index/unweighted |
| 5 | P2u | 2nd component shape directional WHIM index/unweighted |
| 6 | G1u | 1st component symmetry directional WHIM index/unweighted |
| 7 | G2u | 2nd component symmetry directional WHIM index/unweighted |
| 8 | G3u | 3rd component symmetry directional WHIM index/unweighted |
| 9 | E1u | 1st component accessibility directional WHIM index/unweighted |
| 10 | E2u | 2nd component accessibility directional WHIM index/unweighted |
| 11 | E3u | 3rd component accessibility directional WHIM index/unweighted |
| 12 | L1m | 1st component size directional WHIM index/weighted by atomic masses |
| 13 | L2m | 2nd component size directional WHIM index/weighted by atomic masses |
| 14 | L3m | 3rd component size directional WHIM index/weighted by atomic masses |
| 15 | P1m | 1st component shape directional WHIM index/weighted by atomic masses |
| 16 | P2m | 2nd component shape directional WHIM index/weighted by atomic masses |
| 17 | G1m | 1st component symmetry directional WHIM index/weighted by atomic masses |
| 18 | G2m | 2nd component symmetry directional WHIM index/weighted by atomic masses |
| 19 | G3m | 3rd component symmetry directional WHIM index/weighted by atomic masses |
| 20 | E1m | 1st component accessibility directional WHIM index/weighted by atomic masses |
| 21 | E2m | 2nd component accessibility directional WHIM index/weighted by atomic masses |
| 22 | E3m | 3rd component accessibility directional WHIM index/weighted by atomic masses |

* Corresponding author. E-mail: jianbotong@aliyun.com

TABLE S-I. Continued

| ID | Symbol | Meaning |
|---|---|---|
| 23 | L1v | 1st component size directional WHIM index/weighted by atomic van der Waals volumes |
| 24 | L2v | 2nd component size directional WHIM index/weighted by atomic van der Waals volumes |
| 25 | L3v | 3rd component size directional WHIM index/weighted by atomic van der Waals volumes |
| 26 | P1v | 1st component shape directional WHIM index/weighted by atomic van der Waals volumes |
| 27 | P2v | 2nd component shape directional WHIM index/weighted by atomic van der Waals volumes |
| 28 | G1v | 1st component symmetry directional WHIM index/weighted by atomic van der Waals volumes |
| 29 | G2v | 2nd component symmetry directional WHIM index/weighted by atomic van der Waals volumes |
| 30 | G3v | 3rd component symmetry directional WHIM index/weighted by atomic van der Waals volumes |
| 31 | E1v | 1st component accessibility directional WHIM index/weighted by atomic van der Waals volumes |
| 32 | E2v | 2nd component accessibility directional WHIM index/weighted by atomic van der Waals volumes |
| 33 | E3v | 3rd component accessibility directional WHIM index/weighted by atomic van der Waals volumes |
| 34 | L1e | 1st component size directional WHIM index/weighted by atomic Sanderson electronegativities |
| 35 | L2e | 2nd component size directional WHIM index/weighted by atomic Sanderson electronegativities |
| 36 | L3e | 3rd component size directional WHIM index/weighted by atomic Sanderson electronegativities |
| 37 | P1e | 1st component shape directional WHIM index/weighted by atomic Sanderson electronegativities |
| 38 | P2e | 2nd component shape directional WHIM index/weighted by atomic Sanderson electronegativities |
| 39 | G1e | 1st component symmetry directional WHIM index/weighted by atomic Sanderson electronegativities |
| 40 | G2e | 2nd component symmetry directional WHIM index/weighted by atomic Sanderson electronegativities |
| 41 | G3e | 3rd component symmetry directional WHIM index/weighted by atomic Sanderson electronegativities |
| 42 | E1e | 1st component accessibility directional WHIM index/weighted by atomic Sanderson electronegativities |
| 43 | E2e | 2nd component accessibility directional WHIM index/weighted by atomic Sanderson electronegativities |
| 44 | E3e | 3rd component accessibility directional WHIM index/weighted by atomic Sanderson electronegativities |

TABLE S-I. Continued

| ID | Symbol | Meaning |
|---|---|---|
| 45 | L1p | 1st component size directional WHIM index/weighted by atomic polarizabilities |
| 46 | L2p | 2nd component size directional WHIM index/weighted by atomic polarizabilities |
| 47 | L3p | 3rd component size directional WHIM index/weighted by atomic polarizabilities |
| 48 | P1p | 1st component shape directional WHIM index/weighted by atomic polarizabilities |
| 49 | P2p | 2nd component shape directional WHIM index/weighted by atomic polarizabilities |
| 50 | G1p | 1st component symmetry directional WHIM index/weighted by atomic polarizabilities |
| 51 | G2p | 2nd component symmetry directional WHIM index/weighted by atomic polarizabilities |
| 52 | G3p | 3rd component symmetry directional WHIM index/weighted by atomic polarizabilities |
| 53 | E1p | 1st component accessibility directional WHIM index/weighted by atomic polarizabilities |
| 54 | E2p | 2nd component accessibility directional WHIM index/weighted by atomic polarizabilities |
| 55 | E3p | 3rd component accessibility directional WHIM index/weighted by atomic polarizabilities |
| 56 | L1s | 1st component size directional WHIM index/weighted by atomic electrotopological states |
| 57 | L2s | 2nd component size directional WHIM index/weighted by atomic electrotopological states |
| 58 | L3s | 3rd component size directional WHIM index/weighted by atomic electrotopological states |
| 59 | P1s | 1st component shape directional WHIM index/weighted by atomic electrotopological states |
| 60 | P2s | 2nd component shape directional WHIM index/weighted by atomic electrotopological states |
| 61 | G1s | 1st component symmetry directional WHIM index/weighted by atomic electrotopological states |
| 62 | G2s | 2nd component symmetry directional WHIM index/weighted by atomic electrotopological states |
| 63 | G3s | 3rd component symmetry directional WHIM index/weighted by atomic electrotopological states |
| 64 | E1s | 1st component accessibility directional WHIM index/weighted by atomic electrotopological states |
| 65 | E2s | 2nd component accessibility directional WHIM index/weighted by atomic electrotopological states |
| 66 | E3s | 3rd component accessibility directional WHIM index/weighted by atomic electrotopological states |
| 67 | Tu | T total size index/unweighted |

TABLE S-I. Continued

| ID | Symbol | Meaning |
|---|---|---|
| 68 | Tm | T total size index/weighted by atomic masses |
| 69 | Tv | T total size index/weighted by atomic van der Waals volumes |
| 70 | Te | T total size index/weighted by atomic Sanderson electronegativities |
| 71 | Tp | T total size index/weighted by atomic polarizabilities |
| 72 | Ts | T total size index/weighted by atomic electrotopological states |
| 73 | Au | A total size index/unweighted |
| 74 | Am | A total size index/weighted by atomic masses |
| 75 | Av | A total size index/weighted by atomic van der Waals volumes |
| 76 | Ae | A total size index/weighted by atomic Sanderson electronegativities |
| 77 | Ap | A total size index/weighted by atomic polarizabilities |
| 78 | As | A total size index/weighted by atomic electrotopological states |
| 79 | Gu | G total symmetry index/unweighted |
| 80 | Gm | G total symmetry index/weighted by atomic masses |
| 81 | Gs | G total symmetry index/weighted by atomic Sanderson electronegativities |
| 82 | Ku | K global shape index/unweighted |
| 83 | Km | K global shape index/weighted by atomic masses |
| 84 | Kv | K global shape index/weighted by atomic van der Waals volumes |
| 85 | Ke | K global shape index/weighted by atomic Sanderson electronegativities |
| 86 | Kp | K global shape index/weighted by atomic polarizabilities |
| 87 | Ks | K global shape index/weighted by atomic electrotopological states |
| 88 | Du | D total accessibility index/unweighted |
| 89 | Dm | D total accessibility index/weighted by atomic masses |
| 90 | Dv | D total accessibility index/weighted by atomic van der Waals volumes |
| 91 | De | D total accessibility index/weighted by atomic Sanderson electronegativities |
| 92 | Dp | D total accessibility index/weighted by atomic polarizabilities |
| 93 | Ds | D total accessibility index/weighted by atomic electrotopological states |
| 94 | Vu | V total size index/unweighted |
| 95 | Vm | V total size index/weighted by atomic masses |
| 96 | Vv | V total size index/weighted by atomic van der Waals volumes |
| 97 | Ve | V total size index/weighted by atomic Sanderson electronegativities |
| 98 | Vp | V total size index/weighted by atomic polarizabilities |
| 99 | Vs | V total size index/weighted by atomic electrotopological states |

TABLE S-II. The meanings of 74 geometrical descriptors

| ID | Symbol | Meaning |
|---|---|---|
| 1 | W3D | 3D-Wener index |
| 2 | J3D | 3D-Balaban index |
| 3 | H3D | 3D-Harary index |
| 4 | AGDD | Average geometric distance degree |
| 5 | DDI | D/D index |
| 6 | ADDD | Average distance/ distance degree |
| 7 | G1 | Gravitational index G1 |
| 8 | G2 | Gravitational index G2(bond-restricted) |
| 9 | RGyr | Radius of gyration (mass weighted) |
| 10 | SPAN | Span R |

TABLE S-II. Continued

| ID | Symbol | Meaning |
|---|---|---|
| 11 | SPAM | Average span R |
| 12 | MEcc | Average eccentricity |
| 13 | SPH | Spherosity |
| 14 | ASP | Asphericity |
| 15 | FDI | Folding degree index |
| 16 | PJI3 | 3D petitjean shape index |
| 17 | L/Bw | Length-to-breadth ratio by WHIM |
| 18 | SEig | Absolute eigenvalue sum on geometry matrix |
| 19 | HOMA | Harmonic Oscillator Model of Aromaticity index |
| 20 | RCI | Jug RG index |
| 21 | AROM | Aromaticity index |
| 22 | HOMT | HOMA total |
| 23 | DISPm | d COMMA2 value /weighted by atomic masses |
| 24 | QXXm | Qxx COMMA 2 value /weighted by atomic masses |
| 25 | QYYm | Qyy COMMA 2 value /weighted by atomic masses |
| 26 | QZZm | Qzz COMMA 2 value /weighted by atomic masses |
| 27 | DISPv | d COMMA2 value /weighted by atomic van der waals volumes |
| 28 | QXXv | Qxx COMMA 2 value /weighted by atomic van der Waals volumes |
| 29 | QYYv | Qyy COMMA 2 value /weighted by atomic van der Waals volumes |
| 30 | QZZv | Qzz COMMA 2 value /weighted by atomic van der Waals volumes |
| 31 | DISPe | d COMMA2 value /weighted by atomic Sanderson electronegativities |
| 32 | QXXe | Qxx COMMA 2 value /weighted by atomic Sanderson electronegativities |
| 33 | QYYe | Qyy COMMA 2 value /weighted by atomic Sanderson electronegativities |
| 34 | QZZe | Qzz COMMA 2 value /weighted by atomic Sanderson electronegativities |
| 35 | DISPp | d COMMA2 value /weighted by atomic polarizabilities |
| 36 | QXXp | Qxx COMMA 2 value /weighted by atomic polarizabilities |
| 37 | QYYp | Qyy COMMA 2 value /weighted by atomic polarizabilities |
| 38 | QZZp | Qzz COMMA 2 value /weighted by atomic polarizabilities |
| 39 | G(N..N) | Sum of geometrical distance between N..N |
| 40 | G(N..O) | Sum of geometrical distance between N..O |
| 41 | G(N..S) | Sum of geometrical distance between N..S |
| 42 | G(N..P) | Sum of geometrical distance between N..P |
| 43 | G(N..F) | Sum of geometrical distance between N..F |
| 44 | G(N..Cl) | Sum of geometrical distance between N..Cl |
| 45 | G(N..Br) | Sum of geometrical distance between N..Br |
| 46 | G(N..I) | Sum of geometrical distance between N..I |
| 47 | G(O..O) | Sum of geometrical distance between O..O |
| 48 | G(O..S) | Sum of geometrical distance between O..S |
| 49 | G(O..P) | Sum of geometrical distance between O..P |
| 50 | G(O..F) | Sum of geometrical distance between O..F |
| 51 | G(O..Cl) | Sum of geometrical distance between O..Cl |
| 52 | G(O..Br) | Sum of geometrical distance between O..Br |
| 53 | G(O..I) | Sum of geometrical distance between O..I |
| 54 | G(S..S) | Sum of geometrical distance between S..S |
| 55 | G(S..P) | Sum of geometrical distance between S..P |

TABLE S-II. Continued

| ID | Symbol | Meaning |
|---|---|---|
| 56 | G(S..F) | Sum of geometrical distance between S..F |
| 57 | G(S..Cl) | Sum of geometrical distance between S..Cl |
| 58 | G(S..Br) | Sum of geometrical distance between S..Br |
| 59 | G(S..I) | Sum of geometrical distance between S..I |
| 60 | G(P..P) | Sum of geometrical distance between P..P |
| 61 | G(P..F) | Sum of geometrical distance between P..F |
| 62 | G(P..Cl) | Sum of geometrical distance between P..Cl |
| 63 | G(P..Br) | Sum of geometrical distance between P..Br |
| 64 | G(P..I) | Sum of geometrical distance between P..I |
| 65 | G(F..F) | Sum of geometrical distance between F..F |
| 66 | G(F..Cl) | Sum of geometrical distance between F..Cl |
| 67 | G(F..Br) | Sum of geometrical distance between F..Br |
| 68 | G(F..I) | Sum of geometrical distance between F..I |
| 69 | G(Cl..Cl) | Sum of geometrical distance between Cl..Cl |
| 70 | G(Cl..Br) | Sum of geometrical distance between Cl..Br |
| 71 | G(Cl..I) | Sum of geometrical distance between Cl..I |
| 72 | G(Br..Br) | Sum of geometrical distance between Br..Br |
| 73 | G(Br..I) | Sum of geometrical distance between Br..I |
| 74 | G(I..I) | Sum of geometrical distance between I..I |

TABLE S-III. The meanings of 160 3D-MORSE descriptors

| ID | Symbol | Meaning |
|---|---|---|
| 1 | Mor01u | 3D-MoRSE -signal 01/unweighted |
| 2 | Mor02u | 3D-MoRSE -signal 02/unweighted |
| 3 | Mor03u | 3D-MoRSE -signal 03/unweighted |
| 4 | Mor04u | 3D-MoRSE -signal 04/unweighted |
| 5 | Mor05u | 3D-MoRSE -signal 05/unweighted |
| 6 | Mor06u | 3D-MoRSE -signal 06/unweighted |
| 7 | Mor07u | 3D-MoRSE -signal 07/unweighted |
| 8 | Mor08u | 3D-MoRSE -signal 08/unweighted |
| 9 | Mor09u | 3D-MoRSE -signal 09/unweighted |
| 10 | Mor10u | 3D-MoRSE -signal 10/unweighted |
| 11 | Mor11u | 3D-MoRSE -signal 11/unweighted |
| 12 | Mor12u | 3D-MoRSE -signal 12/unweighted |
| 13 | Mor13u | 3D-MoRSE -signal 13/unweighted |
| 14 | Mor14u | 3D-MoRSE -signal 14/unweighted |
| 15 | Mor15u | 3D-MoRSE -signal 15/unweighted |
| 16 | Mor16u | 3D-MoRSE -signal 16/unweighted |
| 17 | Mor17u | 3D-MoRSE -signal 17/unweighted |
| 18 | Mor18u | 3D-MoRSE -signal 18/unweighted |
| 19 | Mor19u | 3D-MoRSE -signal 19/unweighted |
| 20 | Mor20u | 3D-MoRSE -signal 20/unweighted |
| 21 | Mor21u | 3D-MoRSE -signal 21/unweighted |
| 22 | Mor22u | 3D-MoRSE -signal 22/unweighted |
| 23 | Mor23u | 3D-MoRSE -signal 23/unweighted |

TABLE S-III. Continued

| ID | Symbol | Meaning |
|---|---|---|
| 24 | Mor24u | 3D-MoRSE -signal 24/unweighted |
| 25 | Mor25u | 3D-MoRSE -signal 25/unweighted |
| 26 | Mor26u | 3D-MoRSE -signal 26/unweighted |
| 27 | Mor27u | 3D-MoRSE -signal 27/unweighted |
| 28 | Mor28u | 3D-MoRSE -signal 28/unweighted |
| 29 | Mor29u | 3D-MoRSE -signal 29/unweighted |
| 30 | Mor30u | 3D-MoRSE -signal 30/unweighted |
| 31 | Mor31u | 3D-MoRSE -signal 31/unweighted |
| 32 | Mor32u | 3D-MoRSE -signal 32/unweighted |
| 33 | Mor01m | 3D-MoRSE -signal 01/weighted by atomic masses |
| 34 | Mor02m | 3D-MoRSE -signal 02/weighted by atomic masses |
| 35 | Mor03m | 3D-MoRSE -signal 03/weighted by atomic masses |
| 36 | Mor04m | 3D-MoRSE -signal 04/weighted by atomic masses |
| 37 | Mor05m | 3D-MoRSE -signal 05/weighted by atomic masses |
| 38 | Mor06m | 3D-MoRSE -signal 06/weighted by atomic masses |
| 39 | Mor07m | 3D-MoRSE -signal 07/weighted by atomic masses |
| 40 | Mor08m | 3D-MoRSE -signal 08/weighted by atomic masses |
| 41 | Mor09m | 3D-MoRSE -signal 09/weighted by atomic masses |
| 42 | Mor10m | 3D-MoRSE -signal 10/weighted by atomic masses |
| 43 | Mor11m | 3D-MoRSE -signal 11/weighted by atomic masses |
| 44 | Mor12m | 3D-MoRSE -signal 12/weighted by atomic masses |
| 45 | Mor13m | 3D-MoRSE -signal 13/weighted by atomic masses |
| 46 | Mor14m | 3D-MoRSE -signal 14/weighted by atomic masses |
| 47 | Mor15m | 3D-MoRSE -signal 15/weighted by atomic masses |
| 48 | Mor16m | 3D-MoRSE -signal 16/weighted by atomic masses |
| 49 | Mor17m | 3D-MoRSE -signal 17/weighted by atomic masses |
| 50 | Mor18m | 3D-MoRSE -signal 18/weighted by atomic masses |
| 51 | Mor19m | 3D-MoRSE -signal 19/weighted by atomic masses |
| 52 | Mor20m | 3D-MoRSE -signal 20/weighted by atomic masses |
| 53 | Mor21m | 3D-MoRSE -signal 21/weighted by atomic masses |
| 54 | Mor22m | 3D-MoRSE -signal 22/weighted by atomic masses |
| 55 | Mor23m | 3D-MoRSE -signal 23/weighted by atomic masses |
| 56 | Mor24m | 3D-MoRSE -signal 24/weighted by atomic masses |
| 57 | Mor25m | 3D-MoRSE -signal 25/weighted by atomic masses |
| 58 | Mor26m | 3D-MoRSE -signal 26/weighted by atomic masses |
| 59 | Mor27m | 3D-MoRSE -signal 27/weighted by atomic masses |
| 60 | Mor28m | 3D-MoRSE -signal 28/weighted by atomic masses |
| 61 | Mor29m | 3D-MoRSE -signal 29/weighted by atomic masses |
| 62 | Mor30m | 3D-MoRSE -signal 30/weighted by atomic masses |
| 63 | Mor31m | 3D-MoRSE -signal 31/weighted by atomic masses |
| 64 | Mor32m | 3D-MoRSE -signal 32/weighted by atomic masses |
| 65 | Mor01v | 3D-MoRSE -signal 01/weighted by atomic van der Waals volumes |
| 66 | Mor02v | 3D-MoRSE -signal 02/weighted by atomic van der Waals volumes |
| 67 | Mor03v | 3D-MoRSE -signal 03/weighted by atomic van der Waals volumes |
| 68 | Mor04v | 3D-MoRSE -signal 04/weighted by atomic van der Waals volumes |

TABLE S-III. Continued

| ID | Symbol | Meaning |
|---|---|---|
| 69 | Mor05v | 3D-MoRSE -signal 05/weighted by atomic van der Waals volumes |
| 70 | Mor06v | 3D-MoRSE -signal 06/weighted by atomic van der Waals volumes |
| 71 | Mor07v | 3D-MoRSE -signal 07/weighted by atomic van der Waals volumes |
| 72 | Mor08v | 3D-MoRSE -signal 08/weighted by atomic van der Waals volumes |
| 73 | Mor09v | 3D-MoRSE -signal 09/weighted by atomic van der Waals volumes |
| 74 | Mor10v | 3D-MoRSE -signal 10/weighted by atomic van der Waals volumes |
| 75 | Mor11v | 3D-MoRSE -signal 11/weighted by atomic van der Waals volumes |
| 76 | Mor12v | 3D-MoRSE -signal 12/weighted by atomic van der Waals volumes |
| 77 | Mor13v | 3D-MoRSE -signal 13/weighted by atomic van der Waals volumes |
| 78 | Mor14v | 3D-MoRSE -signal 14/weighted by atomic van der Waals volumes |
| 79 | Mor15v | 3D-MoRSE -signal 15/weighted by atomic van der Waals volumes |
| 80 | Mor16v | 3D-MoRSE -signal 16/weighted by atomic van der Waals volumes |
| 81 | Mor17v | 3D-MoRSE -signal 17/weighted by atomic van der Waals volumes |
| 82 | Mor18v | 3D-MoRSE -signal 18/weighted by atomic van der Waals volumes |
| 83 | Mor19v | 3D-MoRSE -signal 19/weighted by atomic van der Waals volumes |
| 84 | Mor20v | 3D-MoRSE -signal 20/weighted by atomic van der Waals volumes |
| 85 | Mor21v | 3D-MoRSE -signal 21/weighted by atomic van der Waals volumes |
| 86 | Mor22v | 3D-MoRSE -signal 22/weighted by atomic van der Waals volumes |
| 87 | Mor23v | 3D-MoRSE -signal 23/weighted by atomic van der Waals volumes |
| 88 | Mor24v | 3D-MoRSE -signal 24/weighted by atomic van der Waals volumes |
| 89 | Mor25v | 3D-MoRSE -signal 25/weighted by atomic van der Waals volumes |
| 90 | Mor26v | 3D-MoRSE -signal 26/weighted by atomic van der Waals volumes |
| 91 | Mor27v | 3D-MoRSE -signal 27/weighted by atomic van der Waals volumes |
| 92 | Mor28v | 3D-MoRSE -signal 28/weighted by atomic van der Waals volumes |
| 93 | Mor29v | 3D-MoRSE -signal 29/weighted by atomic van der Waals volumes |
| 94 | Mor30v | 3D-MoRSE -signal 30/weighted by atomic van der Waals volumes |
| 95 | Mor31v | 3D-MoRSE -signal 31/weighted by atomic van der Waals volumes |
| 96 | Mor32v | 3D-MoRSE -signal 32/weighted by atomic van der Waals volumes |
| 97 | Mor01e | 3D-MoRSE -signal 01/weighted by atomic Sanderson electronegativities |
| 98 | Mor02e | 3D-MoRSE -signal 02/weighted by atomic Sanderson electronegativities |
| 99 | Mor03e | 3D-MoRSE -signal 03/weighted by atomic Sanderson electronegativities |
| 100 | Mor04e | 3D-MoRSE -signal 04/weighted by atomic Sanderson electronegativities |
| 101 | Mor05e | 3D-MoRSE -signal 05/weighted by atomic Sanderson electronegativities |
| 102 | Mor06e | 3D-MoRSE -signal 06/weighted by atomic Sanderson electronegativities |
| 103 | Mor07e | 3D-MoRSE -signal 07/weighted by atomic Sanderson electronegativities |
| 104 | Mor08e | 3D-MoRSE -signal 08/weighted by atomic Sanderson electronegativities |
| 105 | Mor09e | 3D-MoRSE -signal 09/weighted by atomic Sanderson electronegativities |
| 106 | Mor10e | 3D-MoRSE -signal 10/weighted by atomic Sanderson electronegativities |
| 107 | Mor11e | 3D-MoRSE -signal 11/weighted by atomic Sanderson electronegativities |
| 108 | Mor12e | 3D-MoRSE -signal 12/weighted by atomic Sanderson electronegativities |
| 109 | Mor13e | 3D-MoRSE -signal 13/weighted by atomic Sanderson electronegativities |
| 110 | Mor14e | 3D-MoRSE -signal 14/weighted by atomic Sanderson electronegativities |
| 111 | Mor15e | 3D-MoRSE -signal 15/weighted by atomic Sanderson electronegativities |
| 112 | Mor16e | 3D-MoRSE -signal 16/weighted by atomic Sanderson electronegativities |
| 113 | Mor17e | 3D-MoRSE -signal 17/weighted by atomic Sanderson electronegativities |

TABLE S-III. Continued

| ID | Symbol | Meaning |
|---|---|---|
| 114 | Mor18e | 3D-MoRSE -signal 18/weighted by atomic Sanderson electronegativities |
| 115 | Mor19e | 3D-MoRSE -signal 19/weighted by atomic Sanderson electronegativities |
| 116 | Mor20e | 3D-MoRSE -signal 20/weighted by atomic Sanderson electronegativities |
| 117 | Mor21e | 3D-MoRSE -signal 21/weighted by atomic Sanderson electronegativities |
| 118 | Mor22e | 3D-MoRSE -signal 22/weighted by atomic Sanderson electronegativities |
| 119 | Mor23e | 3D-MoRSE -signal 23/weighted by atomic Sanderson electronegativities |
| 120 | Mor24e | 3D-MoRSE -signal 24/weighted by atomic Sanderson electronegativities |
| 121 | Mor25e | 3D-MoRSE -signal 25/weighted by atomic Sanderson electronegativities |
| 122 | Mor26e | 3D-MoRSE -signal 26/weighted by atomic Sanderson electronegativities |
| 123 | Mor27e | 3D-MoRSE -signal 27/weighted by atomic Sanderson electronegativities |
| 124 | Mor28e | 3D-MoRSE -signal 28/weighted by atomic Sanderson electronegativities |
| 125 | Mor29e | 3D-MoRSE -signal 29/weighted by atomic Sanderson electronegativities |
| 126 | Mor30e | 3D-MoRSE -signal 30/weighted by atomic Sanderson electronegativities |
| 127 | Mor31e | 3D-MoRSE -signal 31/weighted by atomic Sanderson electronegativities |
| 128 | Mor32e | 3D-MoRSE -signal 32/weighted by atomic Sanderson electronegativities |
| 129 | Mor01p | 3D-MoRSE -signal 01/weighted by atomic polarizabilities |
| 130 | Mor02p | 3D-MoRSE -signal 02/weighted by atomic polarizabilities |
| 131 | Mor03p | 3D-MoRSE -signal 03/weighted by atomic polarizabilities |
| 132 | Mor04p | 3D-MoRSE -signal 04/weighted by atomic polarizabilities |
| 133 | Mor05p | 3D-MoRSE -signal 05/weighted by atomic polarizabilities |
| 134 | Mor06p | 3D-MoRSE -signal 06/weighted by atomic polarizabilities |
| 135 | Mor07p | 3D-MoRSE -signal 07/weighted by atomic polarizabilities |
| 136 | Mor08p | 3D-MoRSE -signal 08/weighted by atomic polarizabilities |
| 137 | Mor09p | 3D-MoRSE -signal 09/weighted by atomic polarizabilities |
| 138 | Mor10p | 3D-MoRSE -signal 10/weighted by atomic polarizabilities |
| 139 | Mor11p | 3D-MoRSE -signal 11/weighted by atomic polarizabilities |
| 140 | Mor12p | 3D-MoRSE -signal 12/weighted by atomic polarizabilities |
| 141 | Mor13p | 3D-MoRSE -signal 13/weighted by atomic polarizabilities |
| 142 | Mor14p | 3D-MoRSE -signal 14/weighted by atomic polarizabilities |
| 143 | Mor15p | 3D-MoRSE -signal 15/weighted by atomic polarizabilities |
| 144 | Mor16p | 3D-MoRSE -signal 16/weighted by atomic polarizabilities |
| 145 | Mor17p | 3D-MoRSE -signal 17/weighted by atomic polarizabilities |
| 146 | Mor18p | 3D-MoRSE -signal 18/weighted by atomic polarizabilities |
| 147 | Mor19p | 3D-MoRSE -signal 19/weighted by atomic polarizabilities |
| 148 | Mor20p | 3D-MoRSE -signal 20/weighted by atomic polarizabilities |
| 149 | Mor21p | 3D-MoRSE -signal 21/weighted by atomic polarizabilities |
| 150 | Mor22p | 3D-MoRSE -signal 22/weighted by atomic polarizabilities |
| 151 | Mor23p | 3D-MoRSE -signal 23/weighted by atomic polarizabilities |
| 152 | Mor24p | 3D-MoRSE -signal 24/weighted by atomic polarizabilities |
| 153 | Mor25p | 3D-MoRSE -signal 25/weighted by atomic polarizabilities |
| 154 | Mor26p | 3D-MoRSE -signal 26/weighted by atomic polarizabilities |
| 155 | Mor27p | 3D-MoRSE -signal 27/weighted by atomic polarizabilities |
| 156 | Mor28p | 3D-MoRSE -signal 28/weighted by atomic polarizabilities |
| 157 | Mor29p | 3D-MoRSE -signal 29/weighted by atomic polarizabilities |
| 158 | Mor30p | 3D-MoRSE -signal 30/weighted by atomic polarizabilities |
| 159 | Mor31p | 3D-MoRSE -signal 31/weighted by atomic polarizabilities |
| 160 | Mor32p | 3D-MoRSE -signal 32/weighted by atomic polarizabilities |

TABLE S-IV. The values of 99 WHIM descriptors of 20 amino acid

| Amino acid | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ala A | 2.076 | 1.214 | 0.482 | 0.55 | 0.322 | 0.213 | 0.213 | 0.213 | 0.427 | 0.51 |
| Arg R | 7.806 | 1.54 | 0.675 | 0.779 | 0.154 | 0.175 | 0.175 | 0.175 | 0.543 | 0.387 |
| Asn N | 2.115 | 1.541 | 0.908 | 0.463 | 0.338 | 0.197 | 0.236 | 0.197 | 0.579 | 0.503 |
| Asp D | 2.579 | 1.498 | 0.863 | 0.522 | 0.303 | 0.2 | 0.2 | 0.2 | 0.447 | 0.395 |
| Cys C | 2.627 | 1.089 | 0.638 | 0.603 | 0.25 | 0.208 | 0.208 | 0.208 | 0.43 | 0.424 |
| Gln Q | 4.74 | 1.209 | 0.66 | 0.717 | 0.183 | 0.188 | 0.219 | 0.188 | 0.509 | 0.448 |
| Glu E | 4.059 | 1.324 | 0.86 | 0.65 | 0.212 | 0.191 | 0.191 | 0.191 | 0.465 | 0.37 |
| Gly G | 2.36 | 0.727 | 0.168 | 0.725 | 0.223 | 0.231 | 0.231 | 0.231 | 0.482 | 0.446 |
| His H | 4.126 | 1.475 | 0.964 | 0.628 | 0.225 | 0.188 | 0.188 | 0.188 | 0.502 | 0.35 |
| Ile I | 3.163 | 1.692 | 1.022 | 0.538 | 0.288 | 0.183 | 0.183 | 0.183 | 0.49 | 0.461 |
| Leu L | 3.25 | 1.692 | 0.959 | 0.551 | 0.287 | 0.183 | 0.183 | 0.211 | 0.594 | 0.548 |
| Lys K | 6.389 | 1.219 | 0.862 | 0.754 | 0.144 | 0.179 | 0.179 | 0.179 | 0.55 | 0.35 |
| Met M | 5.175 | 1.355 | 0.866 | 0.7 | 0.183 | 0.188 | 0.188 | 0.188 | 0.491 | 0.367 |
| Phe F | 6.581 | 1.404 | 0.702 | 0.758 | 0.162 | 0.181 | 0.207 | 0.181 | 0.497 | 0.464 |
| Pro P | 2.17 | 1.347 | 0.775 | 0.506 | 0.314 | 0.197 | 0.197 | 0.197 | 0.497 | 0.553 |
| Ser S | 2.076 | 1.272 | 0.741 | 0.508 | 0.311 | 0.208 | 0.208 | 0.208 | 0.452 | 0.456 |
| Thr T | 2.884 | 1.397 | 0.624 | 0.588 | 0.285 | 0.197 | 0.197 | 0.197 | 0.471 | 0.36 |
| Trp W | 7.091 | 2.364 | 0.711 | 0.698 | 0.232 | 0.174 | 0.174 | 0.195 | 0.569 | 0.487 |
| Tyr Y | 6.392 | 1.859 | 0.716 | 0.713 | 0.207 | 0.179 | 0.179 | 0.179 | 0.523 | 0.448 |
| Val V | 3.033 | 1.424 | 0.784 | 0.579 | 0.272 | 0.191 | 0.191 | 0.191 | 0.433 | 0.527 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Ala A | 0.306 | 1.835 | 0.891 | 0.292 | 0.608 | 0.295 | 0.213 | 0.213 | 0.213 | 0.352 |
| Arg R | 0.423 | 7.543 | 1.347 | 0.419 | 0.81 | 0.145 | 0.175 | 0.175 | 0.175 | 0.518 |
| Asn N | 0.455 | 2.38 | 1.404 | 0.428 | 0.565 | 0.333 | 0.197 | 0.197 | 0.197 | 0.791 |
| Asp D | 0.489 | 2.684 | 0.95 | 0.664 | 0.624 | 0.221 | 0.2 | 0.2 | 0.2 | 0.492 |
| Cys C | 0.428 | 3.449 | 0.708 | 0.323 | 0.77 | 0.158 | 0.208 | 0.208 | 0.208 | 0.767 |
| Gln Q | 0.408 | 4.594 | 1.12 | 0.304 | 0.777 | 0.176 | 0.188 | 0.188 | 0.188 | 0.637 |
| Glu E | 0.515 | 4.685 | 0.899 | 0.643 | 0.752 | 0.144 | 0.191 | 0.191 | 0.191 | 0.639 |
| Gly G | 0.202 | 1.814 | 0.661 | 0.045 | 0.72 | 0.262 | 0.231 | 0.231 | 0.231 | 0.29 |
| His H | 0.418 | 3.985 | 1.136 | 0.541 | 0.704 | 0.201 | 0.188 | 0.188 | 0.188 | 0.516 |
| Ile I | 0.549 | 3.146 | 1.111 | 0.712 | 0.633 | 0.224 | 0.183 | 0.211 | 0.183 | 0.502 |
| Leu L | 0.544 | 2.924 | 1.278 | 0.6 | 0.609 | 0.266 | 0.183 | 0.183 | 0.183 | 0.524 |
| Lys K | 0.508 | 6.63 | 0.917 | 0.556 | 0.818 | 0.113 | 0.179 | 0.179 | 0.179 | 0.598 |
| Met M | 0.428 | 4.902 | 0.849 | 0.646 | 0.766 | 0.133 | 0.188 | 0.188 | 0.188 | 0.489 |
| Phe F | 0.465 | 6.635 | 0.825 | 0.409 | 0.843 | 0.105 | 0.181 | 0.181 | 0.181 | 0.504 |
| Pro P | 0.451 | 2.747 | 0.862 | 0.325 | 0.698 | 0.219 | 0.197 | 0.197 | 0.197 | 0.774 |
| Ser S | 0.467 | 1.854 | 0.94 | 0.466 | 0.569 | 0.288 | 0.208 | 0.208 | 0.208 | 0.41 |
| Thr T | 0.428 | 2.703 | 1.067 | 0.418 | 0.646 | 0.255 | 0.197 | 0.197 | 0.197 | 0.421 |
| Trp W | 0.273 | 6.05 | 1.497 | 0.667 | 0.737 | 0.182 | 0.174 | 0.174 | 0.174 | 0.424 |
| Tyr Y | 0.349 | 6.215 | 1.161 | 0.577 | 0.782 | 0.146 | 0.179 | 0.179 | 0.204 | 0.53 |
| Val V | 0.526 | 2.932 | 0.983 | 0.402 | 0.679 | 0.228 | 0.191 | 0.191 | 0.191 | 0.405 |
| | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| Ala A | 0.272 | 0.102 | 1.672 | 0.982 | 0.347 | 0.557 | 0.327 | 0.213 | 0.213 | 0.213 |
| Arg R | 0.342 | 0.172 | 7.174 | 1.273 | 0.472 | 0.804 | 0.143 | 0.175 | 0.175 | 0.175 |
| Asn N | 0.357 | 0.08 | 1.985 | 1.204 | 0.68 | 0.513 | 0.311 | 0.197 | 0.197 | 0.217 |

TABLE S-IV. Continued

| Amino acid | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|
| Asp D | 0.166 | 0.296 | 2.274 | 1.094 | 0.656 | 0.565 | 0.272 | 0.2 | 0.2 | 0.2 |
| Cys C | 0.188 | 0.103 | 2.586 | 0.843 | 0.434 | 0.669 | 0.218 | 0.208 | 0.208 | 0.208 |
| Gln Q | 0.243 | 0.088 | 4.28 | 1.008 | 0.464 | 0.744 | 0.175 | 0.188 | 0.188 | 0.188 |
| Glu E | 0.224 | 0.27 | 3.762 | 0.967 | 0.639 | 0.701 | 0.18 | 0.191 | 0.191 | 0.224 |
| Gly G | 0.376 | 0.014 | 1.852 | 0.573 | 0.106 | 0.732 | 0.226 | 0.231 | 0.231 | 0.313 |
| His H | 0.221 | 0.135 | 3.912 | 1.084 | 0.691 | 0.688 | 0.191 | 0.188 | 0.188 | 0.188 |
| Ile I | 0.206 | 0.236 | 2.862 | 1.306 | 0.784 | 0.578 | 0.264 | 0.183 | 0.183 | 0.183 |
| Leu L | 0.284 | 0.211 | 2.86 | 1.362 | 0.703 | 0.581 | 0.277 | 0.183 | 0.183 | 0.183 |
| Lys K | 0.207 | 0.208 | 5.952 | 0.909 | 0.635 | 0.794 | 0.121 | 0.179 | 0.179 | 0.179 |
| Met M | 0.15 | 0.194 | 4.82 | 0.971 | 0.701 | 0.743 | 0.15 | 0.188 | 0.188 | 0.188 |
| Phe F | 0.178 | 0.143 | 5.965 | 1.035 | 0.486 | 0.797 | 0.138 | 0.181 | 0.181 | 0.181 |
| Pro P | 0.224 | 0.063 | 2.055 | 1.018 | 0.525 | 0.571 | 0.283 | 0.197 | 0.197 | 0.197 |
| Ser S | 0.266 | 0.181 | 1.726 | 1.014 | 0.539 | 0.526 | 0.309 | 0.208 | 0.208 | 0.208 |
| Thr T | 0.205 | 0.166 | 2.574 | 1.059 | 0.449 | 0.631 | 0.259 | 0.197 | 0.197 | 0.197 |
| Trp W | 0.198 | 0.262 | 6.203 | 1.762 | 0.562 | 0.727 | 0.207 | 0.174 | 0.174 | 0.174 |
| Tyr Y | 0.216 | 0.221 | 5.536 | 1.341 | 0.561 | 0.744 | 0.18 | 0.179 | 0.179 | 0.179 |
| Val V | 0.236 | 0.079 | 2.601 | 1.1 | 0.596 | 0.605 | 0.256 | 0.191 | 0.191 | 0.191 |
| | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
| Ala A | 0.289 | 0.326 | 0.159 | 2.111 | 1.192 | 0.476 | 0.559 | 0.315 | 0.213 | 0.213 |
| Arg R | 0.459 | 0.267 | 0.213 | 7.869 | 1.562 | 0.669 | 0.779 | 0.155 | 0.175 | 0.175 |
| Asn N | 0.519 | 0.279 | 0.242 | 2.183 | 1.582 | 0.868 | 0.471 | 0.341 | 0.197 | 0.197 |
| Asp D | 0.348 | 0.208 | 0.286 | 2.673 | 1.445 | 0.861 | 0.537 | 0.29 | 0.2 | 0.2 |
| Cys C | 0.426 | 0.246 | 0.197 | 2.705 | 1.06 | 0.63 | 0.616 | 0.241 | 0.208 | 0.208 |
| Gln Q | 0.423 | 0.303 | 0.204 | 4.822 | 1.291 | 0.626 | 0.715 | 0.192 | 0.188 | 0.188 |
| Glu E | 0.402 | 0.209 | 0.294 | 4.264 | 1.297 | 0.867 | 0.663 | 0.202 | 0.191 | 0.191 |
| Gly G | 0.299 | 0.282 | 0.081 | 2.34 | 0.746 | 0.153 | 0.722 | 0.23 | 0.231 | 0.231 |
| His H | 0.467 | 0.193 | 0.223 | 4.12 | 1.501 | 0.932 | 0.629 | 0.229 | 0.188 | 0.188 |
| Ile I | 0.41 | 0.285 | 0.315 | 3.217 | 1.666 | 1.024 | 0.545 | 0.282 | 0.183 | 0.183 |
| Leu L | 0.465 | 0.367 | 0.291 | 3.255 | 1.701 | 0.949 | 0.551 | 0.288 | 0.183 | 0.183 |
| Lys K | 0.478 | 0.201 | 0.275 | 6.51 | 1.232 | 0.858 | 0.757 | 0.143 | 0.179 | 0.179 |
| Met M | 0.431 | 0.2 | 0.292 | 5.147 | 1.365 | 0.876 | 0.697 | 0.185 | 0.188 | 0.188 |
| Phe F | 0.408 | 0.25 | 0.221 | 6.751 | 1.36 | 0.695 | 0.767 | 0.154 | 0.181 | 0.181 |
| Pro P | 0.44 | 0.315 | 0.205 | 2.314 | 1.318 | 0.745 | 0.529 | 0.301 | 0.197 | 0.197 |
| Ser S | 0.319 | 0.291 | 0.249 | 2.089 | 1.251 | 0.73 | 0.513 | 0.307 | 0.208 | 0.208 |
| Thr T | 0.381 | 0.207 | 0.218 | 2.908 | 1.392 | 0.626 | 0.59 | 0.283 | 0.197 | 0.197 |
| Trp W | 0.436 | 0.27 | 0.172 | 7.065 | 2.311 | 0.749 | 0.698 | 0.228 | 0.174 | 0.174 |
| Tyr Y | 0.4 | 0.247 | 0.225 | 6.537 | 1.826 | 0.737 | 0.718 | 0.201 | 0.179 | 0.179 |
| Val V | 0.317 | 0.31 | 0.303 | 3.103 | 1.397 | 0.77 | 0.589 | 0.265 | 0.191 | 0.191 |
| | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
| Ala A | 0.213 | 0.442 | 0.489 | 0.294 | 1.716 | 1.025 | 0.371 | 0.551 | 0.33 | 0.213 |
| Arg R | 0.175 | 0.554 | 0.411 | 0.419 | 7.174 | 1.305 | 0.507 | 0.798 | 0.145 | 0.175 |
| Asn N | 0.197 | 0.625 | 0.512 | 0.408 | 1.952 | 1.226 | 0.744 | 0.498 | 0.313 | 0.197 |
| Asp D | 0.2 | 0.482 | 0.371 | 0.488 | 2.265 | 1.164 | 0.686 | 0.551 | 0.283 | 0.2 |
| Cys C | 0.208 | 0.457 | 0.406 | 0.415 | 2.865 | 0.841 | 0.438 | 0.691 | 0.203 | 0.208 |
| Gln Q | 0.188 | 0.531 | 0.511 | 0.37 | 4.26 | 1.003 | 0.512 | 0.738 | 0.174 | 0.188 |

TABLE S-IV. Continued

| Amino acid | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|
| Glu E | 0.191 | 0.515 | 0.367 | 0.535 | 3.69 | 1.02 | 0.668 | 0.686 | 0.19 | 0.191 |
| Gly G | 0.231 | 0.474 | 0.469 | 0.167 | 1.929 | 0.587 | 0.123 | 0.731 | 0.222 | 0.231 |
| His H | 0.188 | 0.505 | 0.364 | 0.389 | 3.907 | 1.131 | 0.751 | 0.675 | 0.195 | 0.188 |
| Ile I | 0.183 | 0.505 | 0.447 | 0.555 | 2.885 | 1.376 | 0.818 | 0.568 | 0.271 | 0.183 |
| Leu L | 0.183 | 0.604 | 0.563 | 0.531 | 2.91 | 1.408 | 0.752 | 0.574 | 0.278 | 0.183 |
| Lys K | 0.179 | 0.571 | 0.364 | 0.505 | 5.917 | 0.948 | 0.676 | 0.785 | 0.126 | 0.179 |
| Met M | 0.188 | 0.487 | 0.384 | 0.445 | 4.922 | 0.967 | 0.74 | 0.742 | 0.146 | 0.188 |
| Phe F | 0.181 | 0.523 | 0.44 | 0.468 | 5.979 | 1.107 | 0.518 | 0.786 | 0.146 | 0.181 |
| Pro P | 0.197 | 0.557 | 0.528 | 0.397 | 2.005 | 1.081 | 0.577 | 0.547 | 0.295 | 0.197 |
| Ser S | 0.208 | 0.462 | 0.445 | 0.457 | 1.775 | 1.05 | 0.578 | 0.522 | 0.308 | 0.208 |
| Thr T | 0.197 | 0.478 | 0.353 | 0.431 | 2.614 | 1.1 | 0.477 | 0.624 | 0.263 | 0.197 |
| Trp W | 0.174 | 0.566 | 0.467 | 0.314 | 6.333 | 1.859 | 0.566 | 0.723 | 0.212 | 0.174 |
| Tyr Y | 0.179 | 0.551 | 0.446 | 0.385 | 5.555 | 1.433 | 0.573 | 0.735 | 0.189 | 0.179 |
| Val V | 0.191 | 0.45 | 0.499 | 0.49 | 2.638 | 1.167 | 0.63 | 0.595 | 0.263 | 0.191 |
| | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
| Ala A | 0.213 | 0.213 | 0.302 | 0.358 | 0.182 | 2.032 | 0.955 | 0.268 | 0.624 | 0.293 |
| Arg R | 0.175 | 0.175 | 0.459 | 0.277 | 0.243 | 7.922 | 1.51 | 0.456 | 0.801 | 0.153 |
| Asn N | 0.197 | 0.197 | 0.496 | 0.296 | 0.3 | 2.73 | 1.628 | 0.336 | 0.582 | 0.347 |
| Asp D | 0.2 | 0.2 | 0.344 | 0.235 | 0.311 | 2.993 | 0.967 | 0.763 | 0.634 | 0.205 |
| Cys C | 0.208 | 0.208 | 0.531 | 0.241 | 0.201 | 3.181 | 0.818 | 0.361 | 0.73 | 0.188 |
| Gln Q | 0.188 | 0.188 | 0.415 | 0.309 | 0.247 | 5.576 | 1.22 | 0.243 | 0.792 | 0.173 |
| Glu E | 0.191 | 0.191 | 0.385 | 0.226 | 0.317 | 5.339 | 0.946 | 0.708 | 0.763 | 0.135 |
| Gly G | 0.231 | 0.231 | 0.324 | 0.296 | 0.11 | 1.891 | 0.779 | 0.027 | 0.701 | 0.289 |
| His H | 0.188 | 0.188 | 0.461 | 0.208 | 0.262 | 4.151 | 1.242 | 0.523 | 0.702 | 0.21 |
| Ile I | 0.183 | 0.183 | 0.415 | 0.313 | 0.339 | 3.509 | 1.111 | 0.688 | 0.661 | 0.209 |
| Leu L | 0.183 | 0.183 | 0.478 | 0.389 | 0.332 | 3.183 | 1.267 | 0.605 | 0.63 | 0.251 |
| Lys K | 0.179 | 0.179 | 0.472 | 0.216 | 0.31 | 7.233 | 0.975 | 0.565 | 0.824 | 0.111 |
| Met M | 0.188 | 0.188 | 0.451 | 0.198 | 0.325 | 4.939 | 0.995 | 0.68 | 0.747 | 0.15 |
| Phe F | 0.181 | 0.181 | 0.41 | 0.284 | 0.245 | 7.237 | 0.828 | 0.365 | 0.858 | 0.098 |
| Pro P | 0.197 | 0.197 | 0.423 | 0.356 | 0.254 | 3.391 | 0.846 | 0.222 | 0.761 | 0.19 |
| Ser S | 0.208 | 0.208 | 0.334 | 0.309 | 0.286 | 2.074 | 0.985 | 0.465 | 0.589 | 0.28 |
| Thr T | 0.197 | 0.197 | 0.392 | 0.225 | 0.25 | 2.992 | 1.177 | 0.438 | 0.649 | 0.255 |
| Trp W | 0.174 | 0.174 | 0.454 | 0.3 | 0.171 | 6.281 | 1.403 | 0.769 | 0.743 | 0.166 |
| Tyr Y | 0.179 | 0.179 | 0.4 | 0.277 | 0.229 | 6.809 | 1.199 | 0.593 | 0.792 | 0.139 |
| Val V | 0.191 | 0.191 | 0.327 | 0.353 | 0.34 | 3.29 | 1.063 | 0.286 | 0.709 | 0.229 |
| | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 |
| Ala A | 0.279 | 0.279 | 0.279 | 0.445 | 0.323 | 0.081 | 3.773 | 3.017 | 3.001 | 3.778 |
| Arg R | 0.218 | 0.218 | 0.218 | 0.576 | 0.457 | 0.204 | 10.02 | 9.309 | 8.918 | 10.1 |
| Asn N | 0.24 | 0.24 | 0.24 | 1.063 | 0.483 | 0.046 | 4.564 | 4.212 | 3.869 | 4.633 |
| Asp D | 0.24 | 0.24 | 0.24 | 0.619 | 0.185 | 0.388 | 4.94 | 4.299 | 4.024 | 4.979 |
| Cys C | 0.263 | 0.263 | 0.263 | 0.639 | 0.299 | 0.122 | 4.354 | 4.48 | 3.863 | 4.395 |
| Gln Q | 0.231 | 0.231 | 0.231 | 0.872 | 0.23 | 0.054 | 6.61 | 6.378 | 5.751 | 6.739 |
| Glu E | 0.231 | 0.231 | 0.231 | 0.85 | 0.28 | 0.277 | 6.242 | 6.226 | 5.369 | 6.428 |
| Gly G | 0.301 | 0.301 | 0.218 | 0.6 | 0.196 | 0.082 | 8.686 | 7.868 | 7.487 | 8.806 |
| His H | 0.224 | 0.224 | 0.224 | 0.585 | 0.268 | 0.128 | 6.565 | 5.662 | 5.687 | 6.553 |

TABLE S-IV. Continued

| Amino acid | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ile I | 0.24 | 0.24 | 0.24 | 0.646 | 0.207 | 0.183 | 5.877 | 4.969 | 4.952 | 5.907 |
| Leu L | 0.24 | 0.24 | 0.24 | 0.638 | 0.239 | 0.215 | 5.901 | 4.802 | 4.925 | 5.904 |
| Lys K | 0.231 | 0.231 | 0.231 | 0.723 | 0.225 | 0.197 | 8.47 | 8.103 | 7.496 | 8.599 |
| Met M | 0.24 | 0.24 | 0.24 | 0.508 | 0.196 | 0.214 | 7.396 | 6.398 | 6.492 | 7.389 |
| Phe F | 0.218 | 0.251 | 0.218 | 0.6 | 0.196 | 0.082 | 8.686 | 7.868 | 7.487 | 8.806 |
| Pro P | 0.25 | 0.25 | 0.25 | 1.18 | 0.215 | 0.026 | 4.292 | 3.935 | 3.598 | 4.377 |
| Ser S | 0.263 | 0.263 | 0.263 | 0.559 | 0.298 | 0.173 | 4.088 | 3.261 | 3.279 | 4.07 |
| Thr T | 0.25 | 0.25 | 0.25 | 0.552 | 0.251 | 0.169 | 4.904 | 4.188 | 4.082 | 4.925 |
| Trp W | 0.204 | 0.204 | 0.204 | 0.469 | 0.177 | 0.334 | 10.17 | 8.214 | 8.527 | 10.13 |
| Tyr Y | 0.213 | 0.213 | 0.213 | 0.654 | 0.264 | 0.169 | 8.968 | 7.953 | 7.438 | 9.101 |
| Val V | 0.25 | 0.25 | 0.25 | 0.522 | 0.273 | 0.033 | 5.241 | 4.316 | 4.296 | 5.269 |
| | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 |
| Ala A | 3.112 | 3.255 | 4.108 | 2.429 | 2.563 | 4.085 | 2.776 | 2.741 | 0.213 | 0.213 |
| Arg R | 8.986 | 9.888 | 18.34 | 13.89 | 13.12 | 18.61 | 13.66 | 16.27 | 0.175 | 0.175 |
| Asn N | 3.922 | 4.694 | 6.578 | 4.961 | 4.56 | 6.721 | 4.757 | 5.909 | 0.209 | 0.197 |
| Asp D | 4.115 | 4.723 | 7.381 | 4.966 | 4.697 | 7.409 | 4.988 | 5.917 | 0.2 | 0.2 |
| Cys C | 4.144 | 4.36 | 5.231 | 3.784 | 3.668 | 5.238 | 4.032 | 4.047 | 0.208 | 0.208 |
| Gln Q | 5.774 | 7.039 | 9.662 | 7.393 | 6.764 | 10.05 | 6.965 | 8.455 | 0.198 | 0.188 |
| Glu E | 5.378 | 6.993 | 10 | 7.798 | 6.664 | 10.35 | 6.912 | 9.504 | 0.191 | 0.191 |
| Gly G | 7.604 | 8.431 | 14.84 | 1.309 | 1.318 | 2.218 | 1.443 | 1.545 | 0.231 | 0.231 |
| His H | 5.788 | 5.917 | 11.48 | 7.296 | 7.693 | 11.42 | 8.199 | 7.979 | 0.188 | 0.188 |
| Ile I | 5.079 | 5.309 | 10.31 | 6.528 | 7.005 | 10.36 | 7.454 | 7.079 | 0.183 | 0.192 |
| Leu L | 5.069 | 5.054 | 10.24 | 6.258 | 6.864 | 10.24 | 7.342 | 6.723 | 0.192 | 0.183 |
| Lys K | 7.541 | 8.774 | 14.35 | 10.27 | 9.77 | 14.66 | 10.25 | 11.7 | 0.179 | 0.179 |
| Met M | 6.63 | 6.615 | 12.67 | 7.88 | 8.736 | 12.73 | 9.12 | 8.952 | 0.188 | 0.188 |
| Phe F | 7.604 | 8.431 | 14.84 | 8.522 | 9.581 | 14.82 | 10.29 | 8.94 | 0.189 | 0.181 |
| Pro P | 3.664 | 4.459 | 5.647 | 3.544 | 3.706 | 5.756 | 3.949 | 3.808 | 0.197 | 0.197 |
| Ser S | 3.403 | 3.524 | 5.12 | 3.046 | 3.227 | 5.052 | 3.497 | 3.465 | 0.208 | 0.208 |
| Thr T | 4.191 | 4.607 | 6.698 | 4.459 | 4.356 | 6.737 | 4.648 | 5.348 | 0.197 | 0.197 |
| Trp W | 8.758 | 8.453 | 23.48 | 14.09 | 15.41 | 23.36 | 16.41 | 14.72 | 0.181 | 0.174 |
| Tyr Y | 7.561 | 8.6 | 17.8 | 11.47 | 11.28 | 18.1 | 11.97 | 12.91 | 0.179 | 0.187 |
| Val V | 4.434 | 4.639 | 7.813 | 4.453 | 5.066 | 7.796 | 5.473 | 4.742 | 0.191 | 0.191 |
| | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 |
| Ala A | 0.279 | 0.326 | 0.412 | 0.336 | 0.338 | 0.327 | 0.436 | 0.414 | 0.242 | 0.258 |
| Arg R | 0.218 | 0.668 | 0.715 | 0.707 | 0.669 | 0.698 | 0.702 | 0.451 | 0.344 | 0.313 |
| Asn N | 0.24 | 0.201 | 0.348 | 0.269 | 0.219 | 0.247 | 0.393 | 0.512 | 0.41 | 0.347 |
| Asp D | 0.24 | 0.283 | 0.437 | 0.348 | 0.305 | 0.326 | 0.45 | 0.444 | 0.318 | 0.28 |
| Cys C | 0.263 | 0.405 | 0.655 | 0.504 | 0.423 | 0.537 | 0.594 | 0.427 | 0.353 | 0.29 |
| Gln Q | 0.231 | 0.576 | 0.665 | 0.616 | 0.573 | 0.607 | 0.688 | 0.455 | 0.322 | 0.31 |
| Glu E | 0.231 | 0.475 | 0.629 | 0.551 | 0.495 | 0.529 | 0.645 | 0.45 | 0.378 | 0.301 |
| Gly G | 0.301 | 0.588 | 0.58 | 0.598 | 0.584 | 0.596 | 0.552 | 0.377 | 0.227 | 0.221 |
| His H | 0.224 | 0.443 | 0.556 | 0.532 | 0.443 | 0.512 | 0.552 | 0.423 | 0.291 | 0.294 |
| Ile I | 0.24 | 0.307 | 0.45 | 0.367 | 0.317 | 0.352 | 0.492 | 0.5 | 0.315 | 0.337 |
| Leu L | 0.24 | 0.326 | 0.413 | 0.371 | 0.327 | 0.361 | 0.445 | 0.562 | 0.34 | 0.375 |
| Lys K | 0.231 | 0.631 | 0.727 | 0.691 | 0.636 | 0.677 | 0.737 | 0.469 | 0.337 | 0.318 |

TABLE S-IV. Continued

| Amino acid | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 |
|---|---|---|---|---|---|---|---|---|---|---|
| Met M | 0.24 | 0.55 | 0.649 | 0.614 | 0.545 | 0.614 | 0.62 | 0.429 | 0.277 | 0.308 |
| Phe F | 0.228 | 0.636 | 0.765 | 0.695 | 0.65 | 0.679 | 0.788 | 0.475 | 0.275 | 0.293 |
| Pro P | 0.25 | 0.259 | 0.547 | 0.357 | 0.293 | 0.321 | 0.641 | 0.5 | 0.354 | 0.32 |
| Ser S | 0.263 | 0.262 | 0.353 | 0.29 | 0.27 | 0.283 | 0.383 | 0.459 | 0.286 | 0.286 |
| Thr T | 0.25 | 0.382 | 0.468 | 0.446 | 0.386 | 0.435 | 0.474 | 0.42 | 0.264 | 0.269 |
| Trp W | 0.204 | 0.546 | 0.605 | 0.591 | 0.547 | 0.585 | 0.615 | 0.443 | 0.295 | 0.292 |
| Tyr Y | 0.213 | 0.569 | 0.672 | 0.616 | 0.577 | 0.602 | 0.688 | 0.44 | 0.322 | 0.291 |
| Val V | 0.25 | 0.368 | 0.519 | 0.408 | 0.383 | 0.392 | 0.564 | 0.495 | 0.24 | 0.31 |
| | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | |
| Ala A | 0.408 | 0.281 | 0.283 | 9.096 | 5.923 | 6.134 | 9.058 | 6.541 | 6.516 | |
| Arg R | 0.461 | 0.326 | 0.412 | 36.48 | 27.45 | 26.35 | 36.93 | 27.4 | 31.61 | |
| Asn N | 0.515 | 0.364 | 0.531 | 14.1 | 10.6 | 10.06 | 14.35 | 10.46 | 12.1 | |
| Asp D | 0.447 | 0.297 | 0.397 | 15.66 | 10.96 | 10.35 | 15.71 | 10.91 | 12.85 | |
| Cys C | 0.426 | 0.325 | 0.353 | 11.41 | 9.053 | 8.477 | 11.44 | 9.23 | 9.347 | |
| Gln Q | 0.47 | 0.323 | 0.385 | 20.06 | 15.46 | 14.51 | 20.69 | 14.93 | 17.15 | |
| Glu E | 0.472 | 0.309 | 0.469 | 20.86 | 16.73 | 14.36 | 21.57 | 14.81 | 20.08 | |
| Gly G | 0.37 | 0.243 | 0.284 | 5.777 | 3.882 | 3.961 | 5.725 | 4.222 | 4.281 | |
| His H | 0.419 | 0.31 | 0.327 | 23.92 | 15.41 | 16.31 | 23.74 | 17.3 | 16.59 | |
| Ile I | 0.503 | 0.356 | 0.345 | 21.66 | 13.99 | 14.89 | 21.76 | 15.78 | 15.07 | |
| Leu L | 0.566 | 0.4 | 0.364 | 21.41 | 13.3 | 14.53 | 21.39 | 15.49 | 14.22 | |
| Lys K | 0.48 | 0.333 | 0.382 | 29.54 | 21.75 | 20.7 | 30.14 | 21.59 | 24.46 | |
| Met M | 0.438 | 0.325 | 0.306 | 26.14 | 16.97 | 18.51 | 26.28 | 19.27 | 18.91 | |
| Phe F | 0.477 | 0.313 | 0.293 | 30.01 | 18.63 | 20.07 | 30.01 | 21.32 | 19.56 | |
| Pro P | 0.494 | 0.344 | 0.474 | 12.2 | 8.25 | 8.402 | 12.41 | 8.864 | 8.903 | |
| Ser S | 0.455 | 0.31 | 0.344 | 11.16 | 7.12 | 7.448 | 11.03 | 7.977 | 7.938 | |
| Thr T | 0.421 | 0.289 | 0.314 | 14.12 | 9.851 | 9.661 | 14.2 | 10.21 | 11.5 | |
| Trp W | 0.449 | 0.308 | 0.327 | 45.57 | 28.35 | 30.08 | 45.72 | 31.83 | 29.95 | |
| Tyr Y | 0.461 | 0.302 | 0.363 | 35.28 | 23.58 | 22.88 | 36.01 | 24.09 | 26.35 | |
| Val V | 0.48 | 0.34 | 0.276 | 16.44 | 9.926 | 11.07 | 16.4 | 11.85 | 10.38 | |

TABLE S-V. The values of 74 geometrical descriptors of 20 amino acid

| Amino acid | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ala A | 209.85 | 4.831 | 34.08 | 32.285 | 77.895 | 11.984 | 6.024 | 3.62 | 1.808 | 2.969 |
| Arg R | 1340.1 | 6.439 | 103.07 | 103.08 | 313.6 | 24.123 | 14.186 | 7.503 | 4.49 | 5.056 |
| Asn N | 400.83 | 5.79 | 54.002 | 47.156 | 129.08 | 15.186 | 10.862 | 5.677 | 2.442 | 2.965 |
| Asp D | 365.24 | 5.311 | 47.078 | 45.655 | 118.97 | 14.871 | 10.904 | 5.738 | 2.393 | 3.453 |
| Cys C | 261.67 | 4.892 | 37.263 | 38.381 | 93.191 | 13.313 | 8.069 | 4.449 | 2.285 | 3.051 |
| Gln Q | 651.02 | 5.934 | 68.797 | 65.102 | 185.19 | 18.519 | 11.626 | 6.239 | 3.259 | 4.285 |
| Glu E | 573 | 5.786 | 62.691 | 60.316 | 168.64 | 17.751 | 11.777 | 1.297 | 3.139 | 4.257 |
| Gly G | 112.57 | 3.913 | 21.436 | 22.514 | 45.863 | 9.173 | 4.721 | 3.077 | 1.448 | 2.741 |
| His H | 656.44 | 3.238 | 67.416 | 65.644 | 189.66 | 18.966 | 13.618 | 7.28 | 3.07 | 4.027 |
| Ile I | 760.03 | 6.831 | 83.806 | 69.094 | 222.15 | 20.196 | 10.101 | 5.299 | 3.018 | 3.586 |
| Leu L | 762.04 | 6.797 | 83.553 | 69.278 | 220.61 | 20.055 | 10.026 | 5.28 | 2.967 | 3.356 |
| Lys K | 1054.4 | 6.418 | 42.492 | 87.866 | 269.59 | 22.466 | 10.965 | 5.971 | 4.025 | 4.804 |
| Met M | 686.07 | 5.621 | 66.375 | 68.607 | 191.27 | 19.127 | 10.824 | 5.91 | 3.264 | 4.536 |

TABLE S-V. Continued

| Amino acid | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Phe F | 981.29 | 3.314 | 83.241 | 85.3 | 256.27 | 22.284 | 13.931 | 7.395 | 3.883 | 4.898 |
| Pro P | 388.62 | 3.401 | 55.181 | 45.72 | 136.69 | 16.081 | 9.051 | 5.088 | 2.361 | 2.981 |
| Ser S | 254.55 | 5.031 | 38.337 | 36.357 | 90.225 | 12.889 | 7.715 | 4.33 | 1.949 | 3.066 |
| Thr T | 411.16 | 5.71 | 53.554 | 48.372 | 133.74 | 15.734 | 9.08 | 4.865 | 2.434 | 3.35 |
| Trp W | 1476.1 | 2.543 | 107.02 | 109.34 | 356.38 | 26.399 | 19.432 | 9.707 | 4.299 | 4.982 |
| Tyr Y | 1092.5 | 3.374 | 88.813 | 91.041 | 276.38 | 23.031 | 15.917 | 8.158 | 3.987 | 4.85 |
| Val V | 532.76 | 6.231 | 65.34 | 56.08 | 167.45 | 17.627 | 8.74 | 4.725 | 2.612 | 3.685 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Ala A | 0.478 | 0.987 | 0.617 | 0.199 | 0.93 | 0.598 | 2.1 | 24.192 | 0 | 0 |
| Arg R | 0.441 | 0.998 | 0.798 | 0.519 | 0.937 | 0.967 | 5.6 | 50.44 | 0 | 0 |
| Asn N | 0.418 | 0.984 | 0.403 | 0.161 | 0.905 | 0.349 | 1.7 | 32.509 | 0 | 0 |
| Asp D | 0.465 | 0.969 | 0.476 | 0.194 | 0.938 | 0.499 | 2.8 | 30.58 | 0 | 0 |
| Cys C | 0.467 | 0.996 | 0.56 | 0.434 | 0.963 | 0.77 | 4.9 | 26.962 | 0 | 0 |
| Gln Q | 0.463 | 0.998 | 0.7 | 0.455 | 0.936 | 0.824 | 4.4 | 38.085 | 0 | 0 |
| Glu E | 0.473 | 0.991 | 0.587 | 0.396 | 0.943 | 0.818 | 5.2 | 36.537 | 0 | 0 |
| Gly G | 0.524 | 1 | 0.845 | 0.381 | 0.925 | 0.601 | 2.7 | 18.495 | 0 | 0 |
| His H | 0.449 | 0.991 | 0.56 | 0.317 | 0.956 | 0.73 | 3.5 | 39.695 | 0 | 0 |
| Ile I | 0.404 | 0.974 | 0.478 | 0.207 | 0.927 | 0.656 | 2.8 | 42.26 | 0 | 0 |
| Leu L | 0.391 | 0.979 | 0.512 | 0.186 | 0.921 | 0.488 | 2.3 | 42.613 | 0 | 0 |
| Lys K | 0.447 | 0.996 | 0.695 | 0.531 | 0.945 | 0.807 | 7.2 | 16.121 | 0 | 0 |
| Met M | 0.476 | 0.991 | 0.649 | 0.422 | 0.968 | 0.924 | 5.8 | 39.621 | 0 | 0 |
| Phe F | 0.461 | 0.998 | 0.758 | 0.587 | 0.977 | 0.766 | 8 | 46.298 | 0.969 | 1.4 |
| Pro P | 0.419 | 0.993 | 0.458 | 0.313 | 0.957 | 0.566 | 3.2 | 33.063 | 0 | 0 |
| Ser S | 0.463 | 0.968 | 0.456 | 0.14 | 0.929 | 0.591 | 2 | 26.024 | 0 | 0 |
| Thr T | 0.444 | 0.988 | 0.618 | 0.237 | 0.935 | 0.681 | 2.5 | 31.864 | 0 | 0 |
| Trp W | 0.43 | 0.994 | 0.79 | 0.373 | 0.984 | 0.689 | 4 | 55.378 | 0.908 | 1.42 |
| Tyr Y | 0.45 | 0.996 | 0.76 | 0.456 | 0.967 | 0.933 | 5.4 | 48.317 | 0.969 | 1.4 |
| Val V | 0.44 | 0.991 | 0.551 | 0.283 | 0.936 | 0.863 | 3 | 36.127 | 0 | 0 |
| | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| Ala A | 0 | 0 | 9.52 | 8.94 | 15.818 | 20.016 | 3.203 | 9.126 | 13.683 | 18.066 |
| Arg R | 0 | 0 | 5.07 | 25.892 | 115.41 | 128.82 | 0.889 | 24.416 | 106.95 | 118.11 |
| Asn N | 0 | 0 | 6.65 | 25.521 | 30.948 | 40.219 | 3.075 | 17.774 | 24.768 | 29.55 |
| Asp D | 0 | 0 | 5.398 | 17.96 | 37.146 | 40.194 | 1.729 | 15.481 | 25.878 | 29.735 |
| Cys C | 0 | 0 | 4.95 | 11.063 | 37.658 | 41.691 | 6.063 | 10.297 | 23.655 | 27.065 |
| Gln Q | 0 | 0 | 3.755 | 21.315 | 60.137 | 73.784 | 2.057 | 16.368 | 51.478 | 57.674 |
| Glu E | 0 | 0 | 3.762 | 19.35 | 65.428 | 67.79 | 0.806 | 16.788 | 45.949 | 49.278 |
| Gly G | 0 | 0 | 7.527 | 4.455 | 11.594 | 15.499 | 2.694 | 3.55 | 10.198 | 12.634 |
| His H | 0 | 0 | 5.286 | 22.59 | 57.811 | 65.904 | 6.272 | 21.096 | 54.274 | 58.816 |
| Ile I | 0 | 0 | 11.813 | 20.65 | 42.088 | 45.836 | 3.066 | 24.279 | 42.299 | 48.339 |
| Leu L | 0 | 0 | 8.735 | 21.601 | 37.431 | 45.877 | 1.356 | 23.985 | 41.33 | 48.979 |
| Lys K | 0 | 0 | 10.49 | 18.551 | 87.172 | 91.572 | 4.062 | 19.465 | 82.997 | 86.389 |
| Met M | 0 | 0 | 4.673 | 21.268 | 68.445 | 69.282 | 3.065 | 18.711 | 61.299 | 64.021 |
| Phe F | 0.992 | 5.813 | 5.748 | 17.032 | 97.253 | 102.19 | 5.382 | 21.317 | 90.361 | 98.039 |
| Pro P | 0 | 0 | 9.057 | 11.857 | 29.494 | 34.106 | 3.304 | 14.529 | 24.259 | 28.91 |
| Ser S | 0 | 0 | 6.939 | 12.891 | 20.124 | 24.059 | 4.034 | 11.395 | 16.587 | 20.034 |

TABLE S-V. Continued

| Amino acid | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|
| Thr T | 0 | 0 | 10.611 | 15.099 | 31.106 | 36.883 | 4.392 | 13.513 | 26.964 | 32.336 |
| Trp W | 0.977 | 5.447 | 3.784 | 37.626 | 114.21 | 127.55 | 6.332 | 39.537 | 115 | 135.38 |
| Tyr Y | 0.992 | 5.813 | 4.54 | 27.607 | 102.19 | 110.19 | 3.83 | 27.763 | 88.396 | 99.797 |
| Val V | 0 | 0 | 8.755 | 13.959 | 33.239 | 37.012 | 3.116 | 16.986 | 31.98 | 37.008 |
| | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
| Ala A | 0.37 | 22.353 | 34.687 | 44.253 | 0.069 | 10.115 | 14.97 | 19.721 | 0 | 6.33 |
| Arg R | 0.245 | 59.118 | 226.07 | 249.74 | 0.061 | 26.72 | 113.22 | 124.96 | 21.46 | 43.73 |
| Asn N | 0.322 | 43.754 | 54.465 | 67.051 | 0.192 | 19.139 | 25.986 | 30.633 | 3.03 | 19.05 |
| Asp D | 0.302 | 39.354 | 60.293 | 70.259 | 0.085 | 16.859 | 26.877 | 31.219 | 0 | 13.84 |
| Cys C | 0.35 | 24.471 | 48.318 | 54.515 | 0.957 | 11.792 | 28.741 | 32.782 | 0 | 6.22 |
| Gln Q | 0.148 | 39.923 | 112.7 | 126.65 | 0.131 | 17.421 | 54.397 | 60.074 | 5.64 | 22.29 |
| Glu E | 0.206 | 43.192 | 102.33 | 110.89 | 0.087 | 18.362 | 47.383 | 51.172 | 0 | 16.7 |
| Gly G | 0.336 | 9.463 | 26.238 | 32.496 | 0.087 | 3.87 | 11.154 | 13.68 | 0 | 6.37 |
| His H | 0.243 | 50.194 | 104.09 | 115.84 | 0.385 | 23.064 | 56.819 | 61.426 | 10.69 | 24.35 |
| Ile I | 0.414 | 59.397 | 93.552 | 107.65 | 0.034 | 27.41 | 46.235 | 53.148 | 0 | 6.31 |
| Leu L | 0.357 | 58.499 | 92.662 | 109.3 | 0.106 | 26.961 | 45.72 | 53.881 | 0 | 6.39 |
| Lys K | 0.363 | 50.531 | 177.96 | 187.03 | 0.162 | 21.915 | 88.958 | 92.572 | 6.86 | 20.57 |
| Met M | 0.405 | 45.462 | 121.93 | 131.86 | 0.634 | 21.372 | 70.101 | 72.539 | 0 | 6.33 |
| Phe F | 0.439 | 47.633 | 172.58 | 187.93 | 0.479 | 23.922 | 95.659 | 104.3 | 0 | 6.22 |
| Pro P | 0.385 | 35.72 | 52.9 | 62.738 | 0.144 | 16.524 | 25.718 | 30.738 | 0 | 6.22 |
| Ser S | 0.329 | 29.217 | 41.531 | 49.199 | 0.243 | 12.462 | 18.014 | 21.615 | 0 | 9.15 |
| Thr T | 0.411 | 35.555 | 62.266 | 75.733 | 0.152 | 14.887 | 29.099 | 34.93 | 0 | 9.91 |
| Trp W | 0.365 | 83.527 | 213.16 | 255.71 | 0.479 | 42.999 | 122.31 | 145.21 | 5.95 | 15.99 |
| Tyr Y | 0.27 | 62.87 | 178.21 | 204.88 | 0.344 | 30.514 | 92.946 | 106.03 | 0 | 14.09 |
| Val V | 0.311 | 41.562 | 74.292 | 86.206 | 0.149 | 19.263 | 35.049 | 40.779 | 0 | 6.24 |
| | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
| Ala A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Arg R | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Asn N | 0 | 0 | 0 | 0 | 0 | 0 | 7.32 | 0 | 0 | 0 |
| Asp D | 0 | 0 | 0 | 0 | 0 | 0 | 16.5 | 0 | 0 | 0 |
| Cys C | 3.33 | 0 | 0 | 0 | 0 | 0 | 0 | 9.71 | 0 | 0 |
| Gln Q | 0 | 0 | 0 | 0 | 0 | 0 | 11.79 | 0 | 0 | 0 |
| Glu E | 0 | 0 | 0 | 0 | 0 | 0 | 22.57 | 0 | 0 | 0 |
| Gly G | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| His H | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ile I | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Leu L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Lys K | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Met M | 4.8 | 0 | 0 | 0 | 0 | 0 | 0 | 10.42 | 0 | 0 |
| Phe F | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pro P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ser S | 0 | 0 | 0 | 0 | 0 | 0 | 6.91 | 0 | 0 | 0 |
| Thr T | 0 | 0 | 0 | 0 | 0 | 0 | 6.66 | 0 | 0 | 0 |
| Trp W | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tyr Y | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

TABLE S-V. Continued

| Amino acid | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|
| Val V | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
| Ala A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Arg R | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Asn N | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Asp D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cys C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Gln Q | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Glu E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Gly G | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| His H | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ile I | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Leu L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Lys K | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Met M | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Phe F | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pro P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ser S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Thr T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Trp W | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tyr Y | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Val V | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 |
| Ala A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Arg R | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Asn N | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Asp D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cys C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Gln Q | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Glu E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Gly G | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| His H | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ile I | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Leu L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Lys K | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Met M | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Phe F | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pro P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ser S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Thr T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Trp W | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tyr Y | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Val V | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 71 | 72 | 73 | 74 | | | | | | |
| Ala A | 0 | 0 | 0 | 0 | | | | | | |

TABLE S-V. Continued

| Amino acid | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 |
|---|---|---|---|---|---|---|---|---|---|---|
| Asn N | 0 | 0 | 0 | 0 | | | | | | |
| Asp D | 0 | 0 | 0 | 0 | | | | | | |
| Cys C | 0 | 0 | 0 | 0 | | | | | | |
| Gln Q | 0 | 0 | 0 | 0 | | | | | | |
| Glu E | 0 | 0 | 0 | 0 | | | | | | |
| Gly G | 0 | 0 | 0 | 0 | | | | | | |
| His H | 0 | 0 | 0 | 0 | | | | | | |
| Ile I | 0 | 0 | 0 | 0 | | | | | | |
| Leu L | 0 | 0 | 0 | 0 | | | | | | |
| Lys K | 0 | 0 | 0 | 0 | | | | | | |
| Met M | 0 | 0 | 0 | 0 | | | | | | |
| Phe F | 0 | 0 | 0 | 0 | | | | | | |
| Pro P | 0 | 0 | 0 | 0 | | | | | | |
| Ser S | 0 | 0 | 0 | 0 | | | | | | |
| Thr T | 0 | 0 | 0 | 0 | | | | | | |
| Trp W | 0 | 0 | 0 | 0 | | | | | | |
| Tyr Y | 0 | 0 | 0 | 0 | | | | | | |
| Val V | 0 | 0 | 0 | 0 | | | | | | |

TABLE S-VI. The values of 160 3D-MORSE descriptors of 20 amino acid

| Amino acid | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ala A | 78 | 16.945 | –0.943 | –0.031 | –1.478 | –1.004 | 0.556 | 0.387 | 0.569 | 0.135 |
| Arg R | 325 | 27.716 | –3.756 | 3.192 | –4.551 | –1.631 | 1.132 | 0.944 | 0.563 | –0.323 |
| Asn N | 136 | 18.987 | –0.236 | 0.051 | –1.883 | –1.672 | 1.042 | 0.623 | 0.238 | 0.419 |
| Asp D | 120 | 16.214 | –0.831 | 0.233 | –1.719 | –1.436 | 1.27 | 0.206 | 0.373 | 0.133 |
| Cys C | 91 | 15.743 | –1.442 | 0.12 | –1.55 | –1.034 | 0.743 | 0.473 | 0.406 | 0.064 |
| Gln Q | 190 | 21.665 | –2.455 | 0.811 | –3.009 | –1.294 | 0.416 | 1.492 | 0.12 | 0.456 |
| Glu E | 171 | 19.606 | –2.639 | 0.744 | –2.419 | –1.696 | 1.199 | 0.408 | 0.266 | 0.343 |
| Gly G | 45 | 13.001 | –0.719 | –0.047 | –0.908 | –1.159 | 0.752 | 0.228 | 0.255 | 0.179 |
| His H | 190 | 15.678 | –0.496 | 0.655 | –2.74 | –1.134 | 1.222 | 1.82 | –0.219 | 0.035 |
| Ile I | 231 | 20.727 | –2.519 | 1.631 | –3.19 | –1.295 | 0.711 | 0.579 | 1.166 | 0.12 |
| Leu L | 231 | 20.07 | –1.863 | 0.765 | –2.925 | –1.92 | 0.88 | 0.972 | 1.144 | –0.069 |
| Lys K | 276 | 27.074 | –5.673 | 2.853 | –3.553 | –1.577 | 1.072 | 0.899 | 0.675 | 0.229 |
| Met M | 190 | 19.917 | –3.055 | 0.82 | –2.259 | –1.028 | 0.705 | 1.109 | 0.908 | 0.099 |
| Phe F | 253 | 20.049 | –2.17 | 0.998 | –3.235 | –2.403 | 2.762 | 0.172 | 0.235 | –0.258 |
| Pro P | 136 | 22.565 | –2.379 | 0.157 | –1.578 | –1.901 | 1.621 | 0.287 | 0.318 | 0.311 |
| Ser S | 91 | 16.798 | –0.475 | –0.093 | –1.333 | –1.4 | 0.777 | 0.337 | 0.527 | –0.029 |
| Thr T | 136 | 19.421 | –1.768 | 0.69 | –2.436 | –1.11 | 0.444 | 0.762 | 0.742 | –0.145 |
| Trp W | 351 | 20.721 | –1.789 | 0.169 | –2.74 | –2.385 | 3.609 | 0.628 | –0.417 | 0.027 |
| Tyr Y | 276 | 19.38 | –1.418 | 1.528 | –2.606 | –2.614 | 2.676 | 0.186 | 0.532 | –0.611 |
| Val V | 171 | 20.914 | –1.088 | 0.131 | –2.712 | –0.97 | 0.228 | 0.852 | 0.964 | 0.136 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Ala A | –0.648 | –0.414 | 0.161 | 0.308 | 0.316 | –0.181 | –0.036 | –0.513 | 0.338 | 0.172 |
| Arg R | –0.338 | –0.857 | –0.429 | 1.284 | 0.218 | 0.002 | –0.243 | –0.905 | 0.301 | 0.585 |
| Asn N | –0.685 | –0.413 | –0.266 | 0.663 | 0.321 | –0.112 | –0.189 | –0.246 | –0.073 | 0.198 |

TABLE S-VI. Continued

| Amino acid | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Asp D | –0.281 | –0.573 | –0.163 | 0.666 | –0.062 | 0.143 | –0.081 | –0.321 | 0.114 | 0.056 |
| Cys C | –0.449 | –0.228 | –0.089 | 0.265 | 0.349 | –0.136 | –0.051 | –0.345 | 0.042 | 0.33 |
| Gln Q | –0.26 | –1.281 | 0.312 | 0.931 | 0.097 | –0.186 | –0.165 | –0.633 | 0.478 | 0.131 |
| Glu E | –0.334 | –0.934 | 0.147 | 0.661 | –0.068 | –0.022 | –0.143 | –0.529 | 0.4 | 0.189 |
| Gly G | –0.353 | –0.299 | –0.044 | 0.255 | 0.207 | –0.037 | –0.008 | –0.323 | –0.035 | 0.187 |
| His H | 0.155 | –1.109 | 0.309 | 0.076 | 0.535 | 0.279 | –0.107 | –0.922 | 0.345 | 0.433 |
| Ile I | –0.986 | –1.047 | 0.756 | 0.847 | 0.054 | 0.044 | –0.391 | –0.958 | 0.985 | 0.213 |
| Leu L | –0.849 | –1.531 | 0.89 | 1.038 | –0.148 | 0.216 | –0.842 | –0.704 | 0.98 | 0.428 |
| Lys K | –0.816 | –1.444 | 0.291 | 1.297 | –0.252 | 0.237 | –0.395 | –0.937 | 0.621 | 0.501 |
| Met M | –1.089 | –0.576 | 0.399 | 0.661 | 0.275 | –0.153 | –0.302 | –0.616 | 0.649 | 0.571 |
| Phe F | 0.165 | –1.165 | 0.249 | 0.817 | 0.097 | 0.327 | –0.619 | –0.826 | 0.325 | 1.094 |
| Pro P | –0.685 | –0.959 | 0.731 | 0.359 | 0.219 | 0.034 | –0.756 | –0.146 | 0.466 | 0.443 |
| Ser S | –0.107 | –0.572 | –0.255 | 0.553 | 0.228 | –0.008 | –0.084 | –0.348 | –0.02 | 0.287 |
| Thr T | –0.221 | –0.849 | 0.11 | 0.541 | 0.253 | 0.079 | –0.364 | –0.391 | 0.295 | 0.227 |
| Trp W | 0.455 | –1.198 | –0.248 | 0.726 | 0.263 | 0.634 | –0.698 | –1.053 | 0.82 | 0.645 |
| Tyr Y | 0.275 | –1.007 | –0.058 | 0.564 | 0.375 | 0.093 | –0.364 | –0.77 | 0.221 | 0.617 |
| Val V | –0.802 | –0.934 | 0.47 | 0.768 | 0.108 | –0.057 | –0.47 | –0.562 | 0.601 | 0.445 |
| | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| Ala A | –0.106 | 0.158 | –0.18 | –0.056 | –0.032 | 0.087 | 0.17 | –0.14 | –0.056 | 0.019 |
| Arg R | –0.041 | 0.191 | –0.3 | –0.006 | –0.018 | 0.106 | 0.441 | –0.078 | –0.061 | –0.251 |
| Asn N | 0.167 | 0.122 | –0.16 | –0.126 | –0.171 | 0.158 | 0.157 | 0.07 | –0.14 | –0.052 |
| Asp D | 0.169 | 0.051 | 0.037 | –0.212 | –0.024 | 0.079 | 0.009 | 0.089 | –0.032 | –0.015 |
| Cys C | –0.105 | 0.208 | –0.199 | –0.128 | 0 | 0.104 | 0.045 | –0.028 | –0.026 | 0.055 |
| Gln Q | –0.079 | 0.153 | –0.053 | –0.28 | –0.091 | 0.215 | 0.137 | 0.012 | 0.133 | –0.006 |
| Glu E | –0.278 | 0.309 | –0.242 | –0.152 | –0.04 | 0.06 | 0.097 | 0.017 | –0.143 | –0.069 |
| Gly G | 0.01 | 0.009 | 0.052 | –0.177 | 0.01 | 0.017 | 0.058 | 0.019 | 0.063 | –0.131 |
| His H | –0.101 | 0.24 | –0.316 | –0.171 | 0.202 | 0.229 | –0.086 | –0.003 | –0.11 | –0.053 |
| Ile I | –0.065 | –0.177 | –0.132 | –0.312 | 0.281 | 0.191 | 0.193 | –0.225 | –0.357 | 0.188 |
| Leu L | –0.29 | –0.072 | –0.041 | –0.204 | 0.231 | 0.245 | –0.002 | –0.118 | –0.465 | 0.313 |
| Lys K | –0.476 | 0.488 | –0.665 | 0.136 | –0.034 | 0.41 | 0.127 | –0.031 | –0.326 | –0.205 |
| Met M | –0.282 | 0.198 | –0.225 | –0.215 | 0.21 | 0.065 | 0.186 | –0.281 | –0.165 | 0.098 |
| Phe F | –0.803 | 0.468 | –0.769 | –0.053 | 0.631 | –0.108 | –0.057 | –0.214 | 0.052 | –0.062 |
| Pro P | –0.579 | 0.403 | –0.356 | –0.14 | 0.193 | 0.138 | 0.152 | –0.155 | –0.381 | 0.185 |
| Ser S | –0.036 | 0.019 | 0.151 | –0.32 | 0.058 | 0.012 | 0.107 | –0.071 | 0.093 | –0.042 |
| Thr T | –0.159 | 0.085 | –0.151 | 0.008 | –0.082 | –0.203 | –0.005 | 0.008 | –0.101 | –0.033 |
| Trp W | 0.069 | –0.088 | –0.639 | –0.148 | 0.437 | –0.003 | –0.029 | –0.097 | –0.004 | –0.151 |
| Tyr Y | –0.25 | 0.369 | –0.529 | –0.211 | 0.462 | 0.109 | –0.302 | –0.086 | 0.136 | –0.131 |
| Val V | –0.374 | 0.098 | –0.191 | –0.066 | –0.131 | 0.379 | 0.067 | –0.153 | –0.277 | 0.107 |
| | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
| Ala A | 0.017 | 0.042 | 23.545 | 7.14 | –0.808 | –0.249 | –0.826 | –0.015 | 0.883 | –0.365 |
| Arg R | 0.378 | –0.057 | 97.692 | 11.139 | –1.326 | –0.784 | –1.636 | 0.389 | 1.813 | –0.635 |
| Asn N | –0.015 | –0.061 | 54.475 | 9.434 | –0.585 | –0.181 | –1.776 | 0.293 | 1.137 | –0.209 |
| Asp D | –0.085 | 0.05 | 55.173 | 9.337 | –0.665 | –0.398 | –1.395 | –0.026 | 1.495 | –0.465 |
| Cys C | –0.115 | 0.003 | 43.358 | 6.797 | –1.352 | –0.042 | –0.568 | 0.035 | 0.507 | –0.348 |
| Gln Q | –0.085 | 0.107 | 67.503 | 8.786 | –1.697 | –0.176 | –1.57 | 0.264 | 1.427 | –0.42 |

TABLE S-VI. Continued

| Amino acid | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
|---|---|---|---|---|---|---|---|---|---|---|
| Glu E | 0.163 | –0.035 | 68.296 | 8.636 | –2.156 | –0.486 | –1.398 | 0.147 | 1.438 | –0.559 |
| Gly G | 0.003 | –0.013 | 16.067 | 5.861 | 0.466 | –0.285 | –0.684 | –0.102 | 0.764 | –0.14 |
| His H | 0.106 | –0.075 | 76.626 | 9.474 | –0.185 | –1.376 | –1.243 | –0.281 | 1.841 | –0.292 |
| Ile I | 0.284 | –0.01 | 54.168 | 8.141 | –0.794 | –0.413 | –1.316 | 0.473 | 0.979 | 0.708 |
| Leu L | 0.397 | –0.193 | 54.168 | 7.653 | –0.22 | –0.778 | –0.991 | 0.194 | 1.066 | –0.736 |
| Lys K | 0.445 | –0.107 | 67.928 | 9.252 | –1.038 | –0.535 | –1.109 | 0.488 | 1.015 | –0.681 |
| Met M | 0.244 | –0.068 | 68.647 | 7.665 | –2.776 | –0.355 | –0.545 | 0.599 | 0.292 | –0.245 |
| Phe F | 0.267 | –0.054 | 87.62 | 11.38 | –1.573 | –0.056 | –2.002 | 0.54 | 1.834 | 0.974 |
| Pro P | 0.268 | –0.099 | 40.976 | 9.56 | –1.014 | –0.807 | –0.729 | –0.057 | 1.377 | –0.833 |
| Ser S | –0.08 | 0.036 | 33.43 | 8.436 | –0.884 | –0.406 | –0.965 | 0.056 | 1.061 | –0.484 |
| Thr T | 0.158 | –0.117 | 43.829 | 8.793 | –1.035 | –0.315 | –1.103 | 0.12 | 1.157 | –0.674 |
| Trp W | –0.02 | –0.015 | 135.935 | 11.872 | –1.073 | –1.115 | –1.896 | –0.197 | 2.571 | –1.007 |
| Tyr Y | 0.159 | –0.173 | 105.945 | 11.075 | –1.371 | –0.531 | –2.208 | 0.448 | 2.385 | –0.976 |
| Val V | 0.276 | –0.051 | 42.596 | 8.325 | –1.061 | –0.131 | –1.324 | 0.47 | 0.776 | –0.548 |
| | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
| Ala A | –0.137 | 0.103 | –0.051 | 0.077 | –0.093 | 0.013 | 0.144 | –0.151 | 0.028 | –0.98 |
| Arg R | –0.445 | 0.177 | 0.171 | 0.117 | –0.517 | 0.261 | 0.377 | –0.089 | –0.097 | –0.179 |
| Asn N | –0.479 | 0.185 | –0.14 | 0.323 | –0.284 | 0.102 | 0.108 | –0.198 | 0.037 | 0.135 |
| Asp D | –0.533 | 0.318 | –0.098 | 0.168 | –0.104 | –0.026 | 0.065 | –0.068 | 0.026 | 0.029 |
| Cys C | 0.283 | –0.145 | –0.16 | 0.198 | –0.014 | –0.17 | 0.249 | –0.124 | –0.053 | 0.01 |
| Gln Q | –0.231 | 0.132 | –0.031 | 0.072 | –0.202 | 0.283 | 0.057 | –0.293 | 0.113 | 0.046 |
| Glu E | –0.412 | 0.303 | 0.047 | 0.046 | –0.055 | 0.056 | 0.043 | –0.206 | –0.025 | 0.117 |
| Gly G | –0.28 | 0.115 | –0.065 | 0.13 | –0.047 | –0.109 | 0.167 | –0.099 | 0.027 | –0.022 |
| His H | –1.083 | 0.179 | 0.445 | –0.01 | –0.056 | –0.984 | 0.247 | 0.296 | –0.083 | –0.254 |
| Ile I | –1.083 | 0.179 | 0.445 | –0.01 | –0.056 | –0.984 | 0.247 | 0.296 | –0.083 | –0.254 |
| Leu L | 0.063 | 0.134 | 0.07 | –0.179 | –0.218 | 0.434 | –0.055 | –0.083 | –0.176 | 0.039 |
| Lys K | –0.163 | 0.19 | 0.177 | –0.365 | –0.299 | 0.425 | 0.181 | –0.056 | –0.203 | 0.075 |
| Met M | 0.296 | –0.141 | –0.161 | 0.355 | 0.025 | –0.366 | 0.188 | 0.017 | –0.166 | 0.102 |
| Phe F | –0.471 | 0.148 | 0.463 | 0.143 | –0.505 | –0.096 | 0.398 | –0.021 | –0.063 | –0.284 |
| Pro P | –0.18 | 0.389 | –0.167 | 0.112 | –0.31 | 0.234 | 0.232 | –0.333 | –0.035 | 0.108 |
| Ser S | –0.22 | 0.164 | –0.034 | 0.214 | –0.356 | 0.155 | 0.071 | –0.017 | –0.048 | –0.053 |
| Thr T | –0.076 | 0.097 | 0.128 | –0.089 | –0.109 | 0.012 | 0.239 | –0.171 | –0.021 | –0.018 |
| Trp W | –1.183 | 0.272 | 0.479 | 0.438 | –0.867 | –0.327 | 0.577 | 0.315 | –0.36 | –0.337 |
| Tyr Y | –0.587 | 0.157 | 0.348 | 0.347 | –0.563 | –0.316 | 0.636 | –0.251 | 0.087 | –0.378 |
| Val V | –0.04 | 0.206 | –0.046 | –0.031 | –0.236 | 0.34 | 0.005 | –0.137 | –0.076 | 0.015 |
| | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
| Ala A | 0.082 | –0.017 | –0.111 | 0.097 | –0.009 | 0.029 | –0.059 | –0.011 | 0.039 | 0.018 |
| Arg R | 0.172 | –0.052 | –0.059 | 0.033 | 0.083 | 0.063 | –0.054 | –0.1 | 0.162 | –0.08 |
| Asn N | –0.054 | –0.014 | –0.179 | 0.257 | –0.062 | 0.022 | –0.11 | –0.008 | 0.104 | 0 |
| Asp D | 0.052 | –0.098 | –0.051 | 0.072 | 0.057 | 0.024 | –0.124 | 0.024 | –0.01 | 0.067 |
| Cys C | 0.125 | –0.038 | –0.152 | 0.151 | –0.041 | 0.038 | –0.064 | 0.027 | –0.038 | 0.045 |
| Gln Q | 0.303 | –0.28 | –0.047 | 0.11 | 0.094 | –0.064 | –0.098 | 0.041 | 0.097 | 0.003 |
| Glu E | 0.11 | –0.036 | –0.172 | 0.117 | –0.052 | 0.077 | –0.167 | 0.038 | 0.014 | 0.088 |
| Gly G | 0.018 | 0.048 | –0.098 | 0.04 | 0.016 | 0.03 | –0.024 | –0.059 | 0.029 | 0.038 |
| His H | 0.043 | 0.119 | –0.021 | 0.05 | –0.096 | 0.084 | 0.002 | 0.04 | –0.053 | 0.023 |

TABLE S-VI. Continued

| Amino acid | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ile I | 0.043 | 0.119 | –0.021 | 0.05 | –0.096 | 0.084 | 0.002 | 0.04 | –0.053 | 0.023 |
| Leu L | 0.324 | –0.227 | –0.092 | 0.092 | 0.093 | –0.051 | –0.099 | 0.032 | 0.119 | –0.061 |
| Lys K | 0.193 | –0.123 | –0.222 | 0.206 | 0.016 | 0.085 | –0.163 | 0.087 | 0.074 | 0.013 |
| Met M | 0.182 | 0.013 | –0.214 | 0.314 | –0.116 | –0.048 | –0.082 | 0.174 | –0.057 | –0.048 |
| Phe F | 0.152 | 0.4 | –0.315 | 0.057 | –0.142 | 0.141 | 0.139 | –0.084 | –0.099 | 0.009 |
| Pro P | 0.117 | –0.02 | –0.268 | 0.192 | 0.114 | –0.15 | –0.024 | 0.02 | 0.073 | 0.04 |
| Ser S | 0.103 | –0.012 | –0.045 | –0.04 | 0.114 | –0.06 | 0.06 | –0.139 | 0.094 | –0.014 |
| Thr T | 0.079 | 0.069 | –0.2 | 0.108 | –0.01 | 0.072 | –0.098 | 0.012 | –0.002 | 0.053 |
| Trp W | 0.19 | 0.248 | 0.032 | –0.185 | –0.286 | 0.254 | 0.048 | –0.147 | –0.157 | –0.006 |
| Tyr Y | 0.003 | 0.325 | –0.077 | 0.068 | –0.207 | 0.069 | 0.161 | –0.023 | –0.169 | 0.047 |
| Val V | 0.222 | –0.119 | –0.151 | 0.144 | 0.009 | 0.045 | –0.164 | 0.065 | 0.089 | –0.047 |
| | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 |
| Ala A | –0.048 | 0.047 | –0.021 | –0.024 | 20.873 | 6.303 | –0.446 | –0.239 | –0.614 | 0.113 |
| Arg R | 0.101 | –0.041 | 0.057 | 0.045 | 92.957 | 10.834 | –1.227 | –0.069 | –1.718 | 0.034 |
| Asn N | –0.07 | 0.073 | –0.048 | –0.048 | 40.157 | 8.201 | –0.456 | –0.39 | –0.877 | –0.074 |
| Asp D | –0.022 | 0.041 | –0.049 | –0.077 | 35.939 | 7.424 | –0.517 | –0.335 | –0.855 | 0.009 |
| Cys C | –0.02 | 0.127 | –0.208 | 0.065 | 28.284 | 6.41 | –0.77 | –0.203 | –0.615 | 0.028 |
| Gln Q | –0.03 | 0.11 | –0.09 | –0.036 | 55.727 | 8.7 | –1.05 | –0.22 | –1.179 | 0.042 |
| Glu E | –0.082 | 0.113 | –0.03 | –0.071 | 50.739 | 8.289 | –1.044 | –0.217 | –1.097 | 0.015 |
| Gly G | –0.002 | –0.007 | –0.014 | –0.043 | 11.858 | 4.46 | –0.148 | –0.289 | –0.368 | –0.183 |
| His H | –0.057 | 0.151 | –0.048 | –0.11 | 65.203 | 8.596 | –0.204 | –0.779 | –1.067 | –0.237 |
| Ile I | –0.107 | 0.06 | 0.097 | –0.123 | 63.23 | 9.217 | –0.779 | –0.034 | –1.449 | 0.154 |
| Leu L | –0.109 | 0.091 | 0.061 | –0.062 | 63.23 | 9.026 | –0.547 | –0.291 | –1.189 | –0.099 |
| Lys K | –0.088 | –0.004 | 0.114 | –0.069 | 74.967 | 10.245 | –1.451 | –0.049 | –1.314 | 0.196 |
| Met M | 0.049 | 0.075 | –0.004 | –0.09 | 57.446 | 8.017 | –1.531 | –0.099 | –1.017 | 0.161 |
| Phe F | 0.058 | 0.103 | –0.107 | –0.137 | 92.58 | 11.443 | –1.185 | –0.197 | –1.824 | –0.087 |
| Pro P | –0.202 | 0.13 | 0.042 | –0.058 | 40.847 | 9.749 | –1.02 | –0.566 | –0.554 | –0.387 |
| Ser S | 0.015 | 0.037 | –0.08 | 0.012 | 24.358 | 6.569 | –0.448 | –0.322 | –0.604 | –0.072 |
| Thr T | –0.061 | 0.077 | –0.052 | –0.039 | 36.743 | 7.754 | –0.713 | 0.176 | –0.94 | 0.069 |
| Trp W | 0.134 | 0.065 | –0.333 | –0.075 | 137.702 | 12.289 | –1.027 | –1.072 | –1.798 | –0.389 |
| Tyr Y | 0.059 | 0.153 | –0.116 | 0.21 | 99.748 | 10.75 | –1.066 | –0.356 | –1.706 | –0.061 |
| Val V | 0.073 | 0.088 | –0.038 | 0.037 | 46.559 | 8.842 | –0.718 | –0.168 | –1.212 | 0.141 |
| | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 |
| Ala A | 0.432 | –0.093 | 0.056 | 0.061 | –0.063 | –0.148 | –0.043 | 0.176 | 0.119 | 0.119 |
| Arg R | 0.864 | –0.079 | –0.087 | 0.019 | 0.247 | –0.327 | –0.32 | 0.499 | 0.192 | –0.106 |
| Asn N | 0.607 | –0.091 | –0.097 | 0.158 | –0.07 | –0.076 | –0.228 | 0.315 | 0.073 | –0.113 |
| Asp D | 0.61 | –0.168 | –0.072 | 0.112 | 0.007 | –0.141 | –0.146 | 0.27 | 0.025 | –0.068 |
| Cys C | 0.316 | –0.081 | 0.106 | 0.021 | –0.079 | –0.052 | –0.087 | –0.087 | 0.179 | –0.129 |
| Gln Q | 0.562 | 0.009 | –0.078 | 0.158 | 0.096 | –0.382 | –0.133 | 0.451 | 0.041 | –0.209 |
| Glu E | 0.695 | –0.203 | –0.111 | 0.181 | 0.075 | –0.328 | –0.102 | 0.363 | –0.001 | –0.134 |
| Gly G | 0.369 | –0.008 | –0.048 | 0.044 | –0.021 | –0.054 | –0.068 | 0.073 | 0.114 | 0.077 |
| His H | 1.045 | 0.019 | –0.562 | 0.094 | 0.338 | –0.305 | –0.057 | 0.005 | 0.186 | 0.153 |
| Ile I | 0.546 | –0.18 | 0.175 | 0.06 | 0.039 | –0.522 | 0.014 | 0.574 | –0.026 | –0.138 |
| Leu L | 0.679 | –0.239 | 0.283 | 0.027 | 0.035 | –0.615 | 0.065 | 0.555 | 0.008 | –0.147 |
| Lys K | 0.652 | –0.108 | –0.016 | 0.152 | 0.128 | –0.688 | –0.087 | 0.597 | 0.062 | –0.079 |

TABLE S-VI. Continued

| Amino acid | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 |
|---|---|---|---|---|---|---|---|---|---|---|
| Met M | 0.315 | 0.154 | 0.238 | 0.036 | –0.293 | –0.119 | 0.009 | 0.204 | 0.142 | –0.126 |
| Phe F | 1.65 | –0.562 | –0.294 | –0.073 | 0.527 | –0.279 | –0.322 | 0.158 | 0.296 | 0.094 |
| Pro P | 1.006 | –0.456 | 0.022 | 0.309 | –0.163 | –0.306 | 0.003 | 0.298 | 0.188 | –0.248 |
| Ser S | 0.447 | –0.131 | 0.031 | 0.038 | 0.041 | –0.128 | –0.173 | 0.225 | 0.108 | –0.099 |
| Thr T | 0.425 | –0.107 | 0.063 | 0.048 | 0.069 | –0.31 | –0.058 | 0.263 | 0.124 | –0.125 |
| Trp W | 2.291 | –0.523 | –0.856 | 0.148 | 0.572 | –0.092 | –0.563 | –0.001 | 0.47 | 0.293 |
| Tyr Y | 1.737 | –0.592 | –0.298 | –0.138 | 0.48 | –0.123 | –0.379 | 0.034 | 0.459 | –0.042 |
| Val V | 0.403 | –0.1 | 0.16 | 0.096 | –0.042 | –0.368 | –0.072 | 0.501 | –0.008 | –0.127 |
| | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 |
| Ala A | –0.042 | –0.078 | 0.141 | 0.02 | –0.089 | 0.068 | –0.027 | –0.025 | –0.007 | 0.022 |
| Arg R | –0.177 | –0.248 | 0.269 | 0.041 | –0.085 | 0.109 | –0.051 | 0.005 | –0.05 | 0.04 |
| Asn N | –0.074 | 0.018 | 0.018 | 0.038 | –0.083 | 0.119 | –0.011 | –0.065 | –0.065 | 0.062 |
| Asp D | –0.065 | –0.03 | 0.125 | –0.05 | –0.051 | 0.044 | 0.059 | –0.057 | –0.065 | 0.072 |
| Cys C | –0.042 | –0.06 | 0.103 | 0.037 | –0.107 | 0.113 | –0.044 | –0.047 | –0.01 | 0.051 |
| Gln Q | –0.058 | –0.084 | 0.279 | –0.108 | 0.101 | 0.107 | 0.022 | –0.109 | 0.066 | 0.122 |
| Glu E | –0.1 | –0.049 | 0.211 | –0.081 | –0.162 | 0.13 | –0.029 | –0.031 | –0.059 | 0.058 |
| Gly G | 0.016 | –0.065 | 0.047 | 0.065 | –0.066 | 0.007 | 0.034 | –0.032 | 0 | –0.014 |
| His H | –0.118 | –0.303 | 0.177 | 0.134 | –0.067 | 0.068 | –0.124 | 0 | 0.051 | 0.085 |
| Ile I | –0.208 | –0.121 | 0.42 | –0.133 | –0.026 | 0.004 | 0.003 | –0.132 | 0.045 | 0.085 |
| Leu L | –0.273 | –0.1 | 0.462 | –0.102 | –0.109 | 0.048 | 0.026 | –0.129 | 0.035 | 0.12 |
| Lys K | –0.236 | –0.103 | 0.342 | –0.019 | –0.239 | 0.201 | –0.093 | 0.007 | –0.049 | 0.134 |
| Met M | –0.206 | –0.054 | 0.263 | 0.105 | –0.148 | 0.149 | –0.088 | –0.09 | 0.028 | 0.104 |
| Phe F | –0.228 | –0.394 | 0.196 | 0.514 | –0.351 | 0.107 | –0.274 | 0.068 | 0.214 | –0.011 |
| Pro P | –0.178 | 0.017 | 0.231 | 0.039 | –0.251 | 0.174 | –0.002 | –0.149 | 0.053 | 0.075 |
| Ser S | –0.045 | –0.068 | 0.087 | 0.045 | –0.071 | 0.028 | 0.044 | –0.069 | 0.016 | –0.022 |
| Thr T | –0.114 | –0.065 | 0.186 | 0.011 | –0.116 | 0.069 | –0.024 | –0.006 | –0.045 | 0.071 |
| Trp W | –0.413 | –0.434 | 0.34 | 0.347 | –0.002 | –0.145 | –0.297 | 0.102 | 0.174 | –0.054 |
| Tyr Y | –0.086 | –0.434 | 0.152 | 0.356 | –0.126 | 0.025 | –0.203 | –0.018 | 0.205 | 0.044 |
| Val V | –0.188 | –0.082 | –0.313 | –0.008 | –0.168 | 0.125 | –0.073 | 0.008 | –0.101 | 0.154 |
| | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 |
| Ala A | 0.055 | –0.035 | –0.056 | 0.027 | 0.031 | 0.005 | 82.837 | 17.763 | –0.833 | 0.048 |
| Arg R | 0.162 | –0.071 | –0.036 | –0.046 | 0.164 | 0.001 | 336.929 | 28.849 | –3.378 | 3.075 |
| Asn N | 0.072 | –0.003 | –0.103 | 0.05 | 0.017 | –0.029 | 149.531 | 20.016 | –0.021 | 0.35 |
| Asp D | –0.002 | 0.022 | –0.048 | 0.002 | 0.043 | –0.027 | 136.252 | 17.582 | –0.698 | 0.462 |
| Cys C | 0.02 | –0.022 | –0.041 | 0.058 | –0.039 | 0.027 | 97.269 | 16.618 | –1.469 | 0.24 |
| Gln Q | 0.048 | –0.036 | –0.071 | 0.031 | 0.032 | –0.006 | 203.734 | 22.159 | –2.48 | 1.033 |
| Glu E | 0.033 | 0.013 | –0.102 | 0.011 | 0.102 | –0.031 | 188.222 | 20.43 | –3.045 | 0.878 |
| Gly G | 0.023 | 0.019 | –0.012 | –0.023 | 0.008 | 0.006 | 49.721 | 14.26 | –0.768 | 0.091 |
| His H | –0.035 | –0.012 | –0.059 | 0.023 | 0.086 | –0.073 | 201.631 | 16.419 | –0.163 | 0.525 |
| Ile I | –0.035 | –0.012 | –0.059 | 0.023 | 0.086 | –0.073 | 201.631 | 16.419 | –0.163 | 0.525 |
| Leu L | 0.054 | –0.096 | –0.181 | 0.1 | 0.192 | –0.067 | 232.078 | 19.713 | –1.482 | 0.796 |
| Lys K | 0.058 | –0.029 | –0.157 | –0.019 | 0.223 | –0.061 | 279.532 | 27.373 | –5.107 | 2.638 |
| Met M | 0.022 | –0.078 | –0.066 | 0.05 | 0.106 | –0.056 | 194.655 | 20.285 | –2.995 | 0.836 |
| Phe F | –0.117 | –0.044 | 0.05 | 0.019 | 0.077 | –0.155 | 256.714 | 20.944 | –2.253 | 1.086 |
| Pro P | 0.052 | –0.018 | –0.222 | 0.118 | 0.127 | –0.068 | 140.56 | 22.745 | –2.214 | 0.1 |

TABLE S-VI. Continued

| Amino acid | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ser S | 0.075 | –0.047 | –0.003 | 0.005 | 0.004 | 0.012 | 100.633 | 18.38 | –0.323 | –0.044 |
| Thr T | 0.026 | –0.024 | –0.066 | 0.009 | 0.089 | –0.04 | 145.892 | 20.31 | –1.579 | 0.816 |
| Trp W | –0.158 | –0.022 | 0.077 | 0.028 | –0.127 | –0.088 | 358.067 | 21.217 | –1.518 | 0.218 |
| Tyr Y | –0.186 | 0.011 | 0.066 | 0.009 | 0.034 | –0.174 | 287.473 | 20.412 | –1.225 | 1.738 |
| Val V | 0.054 | –0.095 | –0.102 | 0.036 | 0.131 | –0.002 | 174.016 | 20.95 | –1.043 | 0.334 |
| | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 |
| Ala A | –1.678 | –1.164 | 0.777 | 0.366 | 0.499 | 0.219 | –0.694 | –0.39 | –0.147 | 0.238 |
| Arg R | –4.895 | –1.662 | 1.434 | 0.889 | 0.524 | –0.24 | –0.379 | –0.823 | –0.556 | 1.245 |
| Asn N | –2.445 | –1.747 | 1.241 | 0.696 | 0.161 | 0.472 | –0.774 | –0.331 | –0.316 | 0.616 |
| Asp D | –2.176 | –1.617 | 1.61 | 0.164 | 0.272 | 0.25 | –0.318 | –0.584 | –0.171 | 0.598 |
| Cys C | –1.698 | –1.21 | 0.932 | 0.444 | 0.413 | 0.089 | –0.506 | –0.26 | –0.096 | 0.221 |
| Gln Q | –3.454 | –1.366 | 0.638 | 1.512 | 0.142 | 0.448 | –0.299 | –1.227 | 0.24 | 0.95 |
| Glu E | –2.76 | –1.855 | 1.496 | 1.376 | 0.178 | 0.487 | –0.395 | –0.888 | 0.159 | 0.526 |
| Gly G | –1.143 | –1.31 | 0.93 | 0.258 | 0.179 | 0.264 | –0.408 | –0.296 | –0.027 | 0.18 |
| His H | –2.887 | –1.356 | 1.528 | 0.809 | –0.377 | 0.118 | 0.174 | –1.086 | 0.328 | –0.086 |
| Ile I | –3.356 | –1.364 | 0.948 | 0.469 | 1.061 | 0.136 | –0.918 | –1.047 | 0.747 | 0.743 |
| Leu L | –3.018 | –2.089 | 1.126 | 0.908 | 1.093 | –0.033 | –0.79 | –1.522 | 0.815 | 0.996 |
| Lys K | –3.641 | –1.706 | 1.28 | 0.888 | 0.607 | 0.33 | –0.788 | –1.476 | 0.26 | 1.194 |
| Met M | –2.399 | –1.17 | 0.944 | 1.046 | 0.825 | 0.178 | –1.115 | –0.547 | 0.363 | 0.541 |
| Phe F | –3.407 | –2.351 | 2.833 | 0.187 | 0.151 | –0.157 | 0.144 | –1.125 | 0.214 | 0.72 |
| Pro P | –1.656 | –1.952 | 1.755 | 0.274 | 0.229 | 0.338 | –0.653 | –0.915 | 0.659 | 0.352 |
| Ser S | –1.62 | –1.571 | 1.001 | 1.301 | 0.518 | 0.016 | –0.09 | –0.597 | –0.335 | 0.575 |
| Thr T | –2.693 | –1.249 | 0.666 | 0.651 | 0.821 | –0.177 | –0.119 | –0.965 | 0.12 | 0.449 |
| Trp W | –2.897 | –2.54 | 3.768 | 0.541 | –0.505 | 0.036 | 0.459 | –1.17 | –0.333 | 0.605 |
| Tyr Y | –3.045 | –2.627 | 2.863 | 0.174 | 0.493 | –0.529 | 0.283 | –1.017 | –0.14 | 0.471 |
| Val V | –2.905 | –1.019 | 0.354 | 0.805 | 0.886 | 0.212 | –0.764 | –0.937 | 0.424 | 0.717 |
| | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 |
| Ala A | 0.397 | –0.195 | –0.011 | –0.52 | 0.3 | 0.16 | –0.086 | 0.179 | –0.173 | –0.058 |
| Arg R | 0.408 | –0.018 | –0.213 | –0.94 | 0.235 | 0.546 | 0.043 | 0.173 | –0.233 | –0.035 |
| Asn N | 0.399 | –0.122 | –0.18 | –0.237 | –0.118 | 0.178 | 0.166 | 0.225 | –0.22 | –0.111 |
| Asp D | –0.026 | 0.233 | –0.108 | –0.317 | 0.056 | –0.116 | 0.24 | 0.081 | 0.036 | –0.187 |
| Cys C | 0.389 | –0.103 | –0.055 | –0.358 | 0.015 | 0.317 | –0.089 | 0.235 | –0.203 | –0.113 |
| Gln Q | 0.165 | –0.213 | –0.123 | –0.631 | 0.475 | 0.054 | –0.01 | 0.178 | –0.042 | –0.297 |
| Glu E | 0.014 | –0.028 | –0.132 | –0.514 | 0.334 | 0.187 | –0.256 | 0.325 | –0.237 | –0.15 |
| Gly G | 0.279 | –0.032 | –0.004 | –0.332 | 0.001 | 0.191 | 0.015 | 0.041 | 0.047 | –0.173 |
| His H | 0.631 | 0.34 | –0.082 | –0.947 | 0.262 | 0.46 | –0.067 | 0.261 | –0.296 | –0.188 |
| Ile I | 0.178 | –0.01 | –0.346 | –0.921 | 0.909 | 0.178 | –0.013 | –0.155 | –0.115 | –0.328 |
| Leu L | –0.085 | 0.211 | –0.833 | –0.665 | 0.925 | 0.37 | –0.254 | –0.032 | –0.024 | –0.208 |
| Lys K | –0.103 | 0.237 | –0.39 | –0.898 | 0.541 | 0.465 | –0.452 | 0.503 | –0.611 | 0.111 |
| Met M | 0.338 | –0.158 | –0.273 | –0.58 | 0.582 | 0.568 | –0.249 | 0.242 | –0.226 | –0.216 |
| Phe F | 0.148 | 0.352 | –0.605 | –0.801 | 0.285 | 1.062 | –0.763 | 0.453 | –0.716 | –0.05 |
| Pro P | 0.242 | 0.027 | –0.73 | –0.138 | 0.443 | 0.411 | –0.575 | 0.439 | –0.339 | –0.16 |
| Ser S | 0.221 | 0.087 | –0.095 | –0.396 | –0.036 | 0.264 | 0.013 | 0.005 | 0.197 | –0.334 |
| Thr T | 0.356 | 0.096 | –0.355 | –0.388 | 0.223 | 0.246 | –0.167 | 0.129 | –0.142 | 0.021 |
| Trp W | 0.343 | 0.698 | –0.717 | –1.059 | 0.747 | 0.61 | 0.102 | –0.095 | –0.637 | –0.11 |

TABLE S-VI. Continued

| | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 |
|---|---|---|---|---|---|---|---|---|---|---|
| Tyr Y | 0.469 | 0.06 | –0.341 | –0.793 | 0.159 | 0.583 | –0.191 | 0.412 | –0.541 | –0.198 |
| Val V | 0.168 | –0.049 | –0.0453 | –0.556 | 0.561 | 0.388 | –0.339 | 0.143 | –0.185 | –0.056 |
| | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 |
| Ala A | –0.059 | 0.073 | 0.168 | –0.097 | –0.055 | 0.02 | –0.015 | 0.026 | 23.503 | 6.852 |
| Arg R | –0.024 | 0.118 | 0.499 | –0.064 | 0.007 | –0.264 | 0.297 | –0.034 | 103.614 | 11.716 |
| Asn N | –0.197 | 0.111 | 0.21 | 0.089 | –0.129 | –0.058 | –0.07 | –0.06 | 43.369 | 8.672 |
| Asp D | –0.098 | 0.071 | 0.005 | 0.138 | –0.012 | –0.008 | –0.145 | 0.024 | 38.361 | 7.734 |
| Cys C | –0.031 | 0.082 | 0.06 | 0.016 | 0.002 | 0.062 | –0.182 | 0.135 | 35.365 | 7.289 |
| Gln Q | –0.134 | 0.18 | 0.178 | 0.047 | –0.098 | –0.005 | –0.166 | 0.113 | 61.289 | 9.482 |
| Glu E | –0.103 | 0.042 | 0.112 | 0.06 | –0.125 | –0.043 | 0.098 | –0.069 | 55.31 | 8.968 |
| Gly G | –0.012 | –0.009 | 0.085 | 0.024 | 0.085 | –0.142 | –0.014 | –0.036 | 13.019 | 4.754 |
| His H | 0.182 | 0.197 | –0.078 | 0.035 | –0.111 | –0.011 | 0.028 | –0.091 | 70.101 | 8.839 |
| Ile I | 0.262 | 0.152 | 0.198 | –0.188 | –0.328 | 0.182 | 0.228 | –0.032 | 73.57 | 10.225 |
| Leu L | 0.197 | 0.183 | 0.043 | –0.106 | –0.435 | 0.31 | 0.321 | –0.167 | 73.57 | 10.049 |
| Lys K | –0.069 | 0.376 | 0.165 | 0.006 | –0.303 | –0.221 | 0.392 | –0.109 | 86.362 | 11.419 |
| Met M | 0.165 | 0.066 | 0.176 | –0.226 | –0.149 | 0.095 | 0.213 | –0.078 | 71.445 | 9.327 |
| Phe F | 0.595 | –0.132 | –0.041 | –0.189 | 0.062 | –0.053 | 0.21 | –0.072 | 102.665 | 12.055 |
| Pro P | 0.181 | 0.111 | 0.164 | –0.124 | –0.371 | 0.176 | 0.235 | –0.101 | 46.05 | 10.611 |
| Ser S | 0.031 | –0.026 | 0.123 | –0.047 | 0.133 | –0.052 | –0.131 | 0.035 | 26.775 | 6.929 |
| Thr T | –0.127 | 0.177 | –0.013 | 0.065 | –0.094 | –0.007 | 0.08 | 0.124 | 41.162 | 8.368 |
| Trp W | 0.37 | –0.028 | –0.01 | –0.068 | 0.021 | –0.175 | –0.099 | –0.021 | 150.163 | 12.814 |
| Tyr Y | 0.429 | 0.082 | –0.297 | –0.054 | 0.157 | –0.094 | 0.097 | –0.213 | 109.356 | 11.268 |
| Val V | –0.168 | 0.344 | 0.092 | –0.105 | –0.277 | 0.113 | 0.2 | –0.044 | 53.778 | 9.748 |
| | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 |
| Ala A | –0.479 | –0.227 | –0.623 | –0.184 | 0.395 | –0.022 | 0.095 | 0.049 | –0.091 | –0.191 |
| Arg R | –1.459 | 0.197 | –1.882 | –0.126 | 0.767 | 0.079 | –0.031 | –0.028 | 0.227 | –0.421 |
| Asn N | –0.516 | –0.387 | –0.795 | –0.211 | 0.576 | –0.022 | –0.062 | 0.16 | –0.084 | –0.142 |
| Asp D | –0.573 | –0.323 | –0.791 | –0.092 | 0.567 | –0.105 | –0.021 | 0.089 | 0.004 | –0.195 |
| Cys C | –0.97 | –0.198 | –0.676 | 0.023 | 0.26 | 0.013 | 0.134 | 0.617 | –0.146 | –0.015 |
| Gln Q | –1.15 | –0.166 | –1.187 | –0.083 | 0.476 | 0.157 | –0.075 | 0.175 | 0.086 | –0.486 |
| Glu E | –1.11 | –0.16 | –1.112 | –0.116 | 0.657 | –0.105 | –0.079 | 0.17 | 0.058 | –0.407 |
| Gly G | –0.175 | –0.287 | –0.34 | –0.246 | 0.347 | 0.026 | –0.013 | 0.033 | –0.03 | –0.082 |
| His H | –0.262 | –0.635 | –1.108 | –0.29 | 0.955 | 0.109 | –0.479 | 0.07 | 0.306 | –0.386 |
| Ile I | –0.901 | 0.103 | –1.564 | 0.001 | 0.517 | –0.033 | 0.226 | 0.072 | –0.054 | –0.589 |
| Leu L | –0.687 | –0.183 | –1.3 | –0.265 | 0.655 | –0.081 | 0.333 | 0.011 | –0.033 | –0.718 |
| Lys K | –1.791 | 0.177 | –1.455 | 0.036 | 0.638 | 0.048 | 0.013 | 0.139 | 0.071 | –0.756 |
| Met M | –2.146 | 0.189 | –1.096 | 0.08 | 0.205 | 0.289 | 0.377 | –0.015 | –0.426 | –0.053 |
| Phe F | –1.22 | –0.157 | –1.856 | –0.332 | 1.713 | –0.446 | –0.252 | –0.128 | 0.526 | –0.399 |
| Pro P | –1.108 | –0.498 | –0.585 | –0.554 | 1.022 | –0.353 | 0.054 | 0.298 | –0.203 | –0.397 |
| Ser S | –0.451 | –0.319 | –0.57 | –0.166 | 0.409 | –0.062 | 0.073 | 0.013 | 0.038 | –0.181 |
| Thr T | –0.78 | –0.133 | –0.974 | –0.024 | 0.367 | 0.015 | 0.095 | 0.026 | 0.045 | –0.364 |
| Trp W | –1.103 | –1.028 | –1.804 | –0.554 | 2.353 | –0.372 | –0.805 | 0.115 | 0.574 | –0.222 |
| Tyr Y | –1.077 | –0.276 | –1.654 | –0.292 | 1.739 | –0.484 | –0.237 | –0.207 | 0.486 | –0.23 |
| Val V | –0.731 | –0.154 | –1.284 | 0.01 | 0.372 | 0.032 | 0.215 | 0.072 | –0.086 | –0.447 |

TABLE S-VI. Continued

| Amino acid | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ala A | –0.012 | 0.203 | 0.118 | –0.141 | –0.055 | –0.102 | 0.164 | 0.032 | –0.09 | 0.067 |
| Arg R | –0.265 | 0.565 | 0.157 | –0.107 | –0.195 | –0.278 | 0.299 | 0.076 | –0.101 | 0.122 |
| Asn N | –0.208 | 0.358 | 0.072 | –0.114 | –0.086 | –0.006 | 0.031 | 0.048 | –0.062 | 0.097 |
| Asp D | –0.133 | 0.315 | 0.013 | –0.068 | –0.073 | –0.048 | 0.14 | –0.045 | –0.043 | 0.038 |
| Cys C | –0.089 | 0.051 | 0.226 | –0.138 | –0.044 | –0.087 | 0.116 | 0.059 | –0.112 | 0.127 |
| Gln Q | –0.072 | 0.494 | 0.029 | –0.2 | –0.087 | –0.121 | 0.297 | –0.073 | –0.112 | 0.103 |
| Glu E | –0.069 | 0.411 | –0.021 | –0.124 | –0.117 | –0.088 | 0.247 | –0.004 | –0.172 | 0.135 |
| Gly G | –0.061 | 0.103 | 0.107 | –0.078 | –0.019 | –0.079 | 0.055 | 0.071 | –0.059 | 0.001 |
| His H | –0.021 | 0.063 | 0.188 | 0.129 | –0.123 | –0.332 | 0.208 | 0.152 | –0.081 | 0.076 |
| Ile I | 0.087 | 0.632 | –0.043 | –0.125 | –0.244 | –0.186 | 0.487 | –0.104 | –0.032 | –0.016 |
| Leu L | 0.161 | 0.607 | –0.007 | –0.133 | –0.329 | –0.146 | 0.522 | –0.059 | –0.124 | 0.032 |
| Lys K | –0.01 | 0.656 | 0.02 | –0.072 | –0.254 | –0.165 | 0.39 | 0.024 | –0.263 | 0.215 |
| Met M | 0.047 | 0.2 | 0.187 | –0.14 | –0.231 | –0.121 | 0.325 | 0.145 | –0.2 | 0.183 |
| Phe F | –0.24 | 0.228 | 0.259 | 0.123 | –0.283 | –0.426 | 0.217 | 0.568 | –0.387 | 0.132 |
| Pro P | 0.099 | 0.31 | 0.177 | –0.223 | –0.232 | –0.005 | 0.267 | 0.065 | –0.262 | 0.18 |
| Ser S | –0.145 | 0.246 | 0.114 | –0.109 | –0.053 | –0.073 | 0.082 | 0.066 | –0.081 | 0.03 |
| Thr T | –0.02 | 0.298 | 0.118 | –0.121 | –0.138 | –0.087 | 0.214 | 0.021 | –0.114 | 0.063 |
| Trp W | –0.486 | 0.085 | 0.429 | 0.302 | –0.445 | –0.471 | 0.397 | 0.38 | –0.01 | –0.138 |
| Tyr Y | –0.318 | 0.105 | 0.425 | –0.012 | –0.127 | –0.452 | 0.181 | 0.382 | –0.149 | 0.038 |
| Val V | –0.001 | 0.537 | –0.01 | –0.128 | –0.222 | –0.118 | 0.352 | 0.032 | –0.183 | 0.113 |
| | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 |
| Ala A | –0.043 | –0.029 | 0.004 | 0.03 | 0.062 | –0.054 | –0.058 | 0.029 | 0.039 | 0.011 |
| Arg R | –0.088 | 0.002 | –0.036 | 0.064 | 0.167 | –0.076 | –0.062 | –0.049 | 0.202 | –0.01 |
| Asn N | –0.013 | –0.072 | –0.064 | 0.079 | 0.066 | –0.004 | –0.11 | 0.046 | 0.029 | –0.03 |
| Asp D | 0.054 | –0.073 | –0.049 | 0.079 | –0.002 | 0.014 | –0.052 | 0.002 | 0.049 | –0.013 |
| Cys C | –0.072 | –0.065 | 0.004 | 0.081 | 0.002 | –0.029 | –0.045 | 0.074 | –0.043 | 0.025 |
| Gln Q | 0.011 | –0.121 | –0.056 | 0.142 | 0.041 | –0.04 | –0.084 | –0.049 | 0.046 | 0.003 |
| Glu E | –0.046 | –0.046 | –0.035 | 0.062 | 0.036 | –0.002 | –0.108 | 0.002 | 0.125 | –0.024 |
| Gly G | 0.033 | –0.042 | 0.005 | –0.004 | 0.021 | 0.011 | –0.01 | –0.026 | 0.011 | 0.011 |
| His H | –0.139 | –0.016 | 0.07 | 0.102 | –0.038 | –0.02 | –0.066 | 0.006 | 0.109 | –0.065 |
| Ile I | –0.018 | –0.15 | 0.088 | 0.154 | 0.025 | –0.082 | –0.2 | 0.082 | 0.21 | –0.068 |
| Leu L | 0.002 | –0.141 | 0.077 | 0.142 | 0.037 | –0.107 | –0.211 | 0.116 | 0.23 | –0.076 |
| Lys K | –0.149 | 0.012 | –0.022 | 0.156 | 0.052 | –0.043 | –0.178 | –0.024 | 0.261 | –0.068 |
| Met M | –0.104 | –0.139 | 0.099 | 0.105 | 0.03 | –0.125 | –0.068 | 0.067 | 0.102 | –0.057 |
| Phe F | –0.325 | 0.051 | 0.251 | –0.005 | –0.119 | –0.067 | 0.046 | 0.005 | 0.119 | –0.151 |
| Pro P | –0.044 | –0.143 | 0.075 | 0.089 | 0.052 | –0.039 | –0.236 | 0.127 | 0.149 | –0.072 |
| Ser S | 0.039 | –0.083 | 0.025 | –0.01 | 0.075 | –0.059 | –0.004 | 0.005 | 0.013 | 0.013 |
| Thr T | –0.042 | –0.009 | –0.033 | 0.089 | 0.022 | –0.036 | –0.074 | 0.004 | 0.114 | –0.044 |
| Trp W | –0.322 | 0.068 | 0.216 | –0.04 | –0.154 | –0.03 | 0.061 | 0.016 | –0.086 | –0.085 |
| Tyr Y | –0.228 | –0.037 | 0.236 | 0.059 | –0.197 | –0.007 | 0.065 | –0.014 | 0.068 | –0.167 |
| Val V | –0.094 | 0 | –0.083 | 0.18 | 0.045 | –0.111 | –0.119 | 0.04 | 0.165 | –0.007 |

TABLE S-VII. The prominent principal component score and the corresponding eigenvalue of 99 WHIM, 74 geometrical and 160 3D-MORSE descriptors

| No. | WHIM descriptors | | Geometrical descriptors | | | 3D-MORSE descriptors | | |
|---|---|---|---|---|---|---|---|---|
| | Eigenvalue | % | No. | Eigenvalue | % | No. | Eigenvalue | % |
| 1 | 50.53231 | 51.04274 | 1 | 24.0685 | 55.9734 | 1 | 51.4977 | 32.186 |
| 2 | 21.8993 | 22.1205 | 2 | 5.3808 | 12.5134 | 2 | 33.1044 | 20.6902 |
| 3 | 7.79161 | 7.87031 | 3 | 4.2917 | 9.9808 | 3 | 13.9686 | 8.7304 |
| 4 | **6.45442** | **6.51961** | 4 | 2.6995 | 6.2779 | 4 | 12.3945 | 7.7466 |
| 5 | 3.857 | 3.89596 | 5 | **2.1097** | **4.9062** | 5 | 9.3042 | 5.8151 |
| 6 | 2.10476 | 2.12602 | 6 | 1.4465 | 3.364 | 6 | 6.4255 | 4.0159 |
| 7 | 1.90928 | 1.92857 | 7 | 0.7745 | 1.8012 | 7 | 6.1609 | 3.8505 |
| 8 | 1.28317 | 1.29613 | 8 | 0.5868 | 1.3648 | 8 | 5.1864 | 3.2415 |
| 9 | 1.00451 | 1.01465 | 9 | 0.4141 | 0.9629 | 9 | **4.2146** | **2.6341** |
| 10 | 0.49921 | 0.50425 | 10 | 0.3348 | 0.7786 | 10 | 3.7662 | 2.3539 |
| 11 | 0.40984 | 0.41398 | 11 | 0.3145 | 0.7313 | 11 | 3.2758 | 2.0474 |
| 12 | 0.31501 | 0.31819 | 12 | 0.201 | 0.4675 | 12 | 2.7476 | 1.7173 |
| 13 | 0.29745 | 0.30046 | 13 | 0.144 | 0.3349 | 13 | 2.0126 | 1.2579 |
| 14 | 0.20545 | 0.20753 | 14 | 0.1035 | 0.2406 | 14 | 1.673 | 1.0456 |
| 15 | 0.18188 | 0.18372 | 15 | 0.0674 | 0.1567 | 15 | 1.2115 | 0.7572 |
| 16 | 0.12194 | 0.12317 | 16 | 0.0304 | 0.0706 | 16 | 1.0376 | 0.6485 |
| 17 | 0.08147 | 0.0823 | 17 | 0.015 | 0.0349 | 17 | 0.7839 | 0.4899 |
| 18 | 0.03159 | 0.03191 | 18 | 0.0101 | 0.0234 | 18 | 0.696 | 0.435 |
| 19 | 0.0198 | 0.02 | 19 | 0.0073 | 0.017 | 19 | 0.5391 | 0.337 |

TABLE S-VIII. SVWGM descriptors for amino acid

| Amino acid | 4 PCs of 99 WHIM descriptors | | | | 5 PCs of 74 GEOMETRICAL descriptors | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Ala A | –11.6344 | –1.89677 | 1.97792 | –2.60584 | –4.8673 | 4.8567 | 0.2935 | –1.1984 | –1.7882 |
| Arg R | 11.87107 | –2.87032 | 2.74823 | 1.25733 | 2.8074 | –5.7328 | –10.0304 | 0.1795 | –1.8715 |
| Asn N | –5.34951 | 7.68317 | 4.11656 | 4.1737 | –4.2352 | 4.4868 | –1.581 | –0.5684 | 2.9533 |
| Asp D | –4.02676 | 2.99296 | –3.35851 | –3.77013 | –3.675 | 5.6066 | –1.3744 | 2.375 | 2.2126 |
| Cys C | –5.65032 | –2.87927 | –2.99048 | 2.3435 | –4.8303 | 6.3381 | 0.7404 | –3.6652 | –1.4237 |
| Gln Q | 2.17649 | –2.40014 | 0.84542 | 3.57161 | –4.7011 | –0.8588 | –3.5331 | 2.8551 | 0.7795 |
| Glu E | 2.36689 | 0.15236 | –4.04847 | 0.80392 | –2.5167 | –0.6925 | –1.4272 | –0.5017 | 3.8605 |
| Gly G | –11.7823 | –13.6975 | 3.47023 | 0.20083 | –4.6947 | 9.543 | –0.1147 | –0.1988 | –0.4047 |
| His H | 2.33867 | 0.36099 | –1.56522 | –1.07602 | 4.3481 | 3.6123 | 1.8679 | 3.8021 | 0.8011 |
| Ile I | 0.41167 | 6.40422 | –1.24429 | –1.62207 | –2.213 | –5.2 | 3.6825 | 5.6186 | –5.5525 |
| Leu L | 0.26884 | 8.11636 | 2.89694 | 0.98231 | –3.1883 | –9.1722 | 5.8668 | 4.8209 | –0.3621 |
| Lys K | 9.00644 | –2.09657 | –3.35485 | 2.39215 | –0.5855 | –11.934 | –4.8895 | –1.5481 | 3.8189 |
| Met M | 4.36305 | –1.66543 | –3.97675 | –1.02332 | –3.9573 | –4.0717 | 2.3255 | –4.3088 | –5.9397 |
| Phe F | 7.26373 | –4.36604 | –1.09058 | 1.62148 | 13.8796 | –2.8069 | 2.0969 | –6.5485 | –0.7813 |
| Pro P | –5.3069 | 3.18399 | 0.59462 | 4.27735 | –2.7762 | –3.1282 | 7.5571 | –2.5701 | 7.0349 |
| Ser S | –9.15545 | 2.32007 | –0.49944 | –2.26929 | –3.3416 | 6.8606 | –0.9756 | 0.8169 | –1.6716 |
| Thr T | –4.22009 | –0.27234 | –1.39095 | –2.53752 | –2.7966 | 0.8798 | –1.07 | –1.2773 | –0.1541 |
| Trp W | 11.70248 | 0.16164 | 5.62047 | –4.91875 | 19.2063 | 3.7382 | –0.0853 | 6.6477 | 1.0446 |
| Tyr Y | 8.54042 | –1.52583 | 1.74131 | –1.28473 | 13.2974 | 2.2367 | 0.8869 | –4.4618 | –1.3706 |
| Val V | –3.18401 | 2.29446 | –0.49218 | –0.51649 | –5.1598 | –4.5618 | –0.2362 | –0.2687 | –1.1856 |

TABLE S-VIII. Continued

| Amino acid | 9 PCs of 160 3D–MORSE descriptors | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| Ala A | 0.1084 | –0.9456 | 2.594 | –0.4204 | –5.9082 | 0.7197 | –0.4996 | 1.1106 | 1.2405 |
| Arg R | –2.2739 | 1.9078 | 1.1935 | –2.4422 | 7.4386 | –3.1167 | 4.6443 | –0.2724 | 0.8752 |
| Asn N | –1.0546 | 0.4217 | –0.9058 | –0.1832 | –3.2495 | –2.6099 | –1.4085 | –0.0991 | –1.3005 |
| Asp D | 1.3277 | 2.1233 | –0.3749 | 2.3417 | –3.4151 | –1.5233 | –0.4735 | 0.8486 | –2.1593 |
| Cys C | –1.5163 | –2.8869 | –1.281 | –0.17 | –3.4528 | 4.8887 | 0.7554 | –2.7832 | –1.1356 |
| Gln Q | –1.1897 | 0.1375 | –5.7141 | 1.1501 | 0.7412 | –0.9023 | 3.0675 | 1.2869 | –1.4086 |
| Glu E | 1.0923 | –0.405 | –1.4034 | 3.1204 | –0.6074 | –0.935 | 2.76 | 1.4287 | –2.2504 |
| Gly G | 1.5034 | 0.8204 | 3.4931 | 0.5227 | –6.8679 | 2.8835 | 1.4014 | 2.882 | 1.5407 |
| His H | 5.1879 | –5.1697 | –0.707 | –3.4635 | 1.2963 | –0.2267 | 0.2235 | 0.0099 | –1.0526 |
| Ile I | 2.8169 | 1.2975 | –0.3499 | 3.6691 | –0.6107 | –3.1206 | –2.2368 | –1.9152 | 1.3986 |
| Leu L | –0.8706 | 3.2602 | 1.0827 | –2.473 | –0.7406 | –3.4737 | –1.9935 | –1.5138 | 0.4502 |
| Lys K | 2.5185 | –2.3203 | 1.6602 | 1.1263 | 3.019 | –0.349 | 2.4237 | –0.6669 | 3.2261 |
| Met M | –3.4378 | –3.8467 | –2.0643 | 0.6833 | 1.3496 | 3.1447 | 1.4995 | –4.1297 | –0.9596 |
| Phe F | 0.7353 | 0.795 | 2.6529 | 2.9515 | 6.3729 | 3.7134 | –1.7188 | 1.5139 | 1.1937 |
| Pro P | –3.0942 | 0.809 | –0.1377 | –0.9751 | –3.2208 | 0.2478 | –1.5559 | –0.0085 | 0.8157 |
| Ser S | –2.4715 | 3.4412 | 1.2993 | –0.512 | –5.334 | –0.3277 | –1.1852 | 0.4623 | –0.9085 |
| Thr T | 2.2859 | 0.1757 | 1.2979 | –0.8649 | –3.2382 | –0.3181 | –0.8146 | –0.0062 | 1.1008 |
| Trp W | –4.6764 | –2.6452 | 0.9322 | 0.8924 | 11.0178 | 0.4087 | –3.3601 | 0.4752 | –0.9542 |
| Tyr Y | 3.1242 | 4.2243 | –4.419 | –2.4006 | 7.4664 | 1.5601 | –1.4356 | 1.7758 | –0.5768 |
| Val V | –0.1157 | –1.1943 | 1.1514 | –2.5526 | –2.0567 | –0.6638 | –0.0932 | –0.399 | 0.8648 |

TABLE S-IX. Statistics of PLS on variables selected by SMR (OT analogues)

| No. | Variable | $l$[a] | $R^2_{cum}$ | $Q^2_{LOO}$ |
|---|---|---|---|---|
| 1 | $v_{24}$ | 1 | 0.516 | 0.501 |
| 2 | $v_{24}, v_{17}$ | 2 | 0.855 | 0.596 |
| 3 | $v_{24}, v_{17}, v_{36}$ | 2 | 0.909 | 0.777 |
| 4 | $v_{24}, v_{17}, v_{36}, v_{51}$ | 2 | 0.950 | 0.843 |
| 5 | $\boldsymbol{v_{24}, v_{17}, v_{36}, v_{51}, v_{23}}$ | **2** | **0.944** | **0.864** |
| 6 | $v_{24}, v_{17}, v_{36}, v_{51}, v_{23}, v_{14}$ | 3 | 0.955 | 0.647 |
| 7 | $v_{24}, v_{17}, v_{36}, v_{51}, v_{23}, v_{14}, v_{46}$ | 3 | 0.948 | 0.657 |
| 8 | $v_{24}, v_{17}, v_{36}, v_{51}, v_{23}, v_{14}, v_{46}, v_{40}$ | 2 | 0.954 | 0.668 |
| 9 | $v_{24}, v_{17}, v_{36}, v_{51}, v_{23}, v_{14}, v_{46}, v_{40}, v_{27}$ | 2 | 0.953 | 0.601 |
| 10 | $v_{24}, v_{17}, v_{36}, v_{51}, v_{23}, v_{14}, v_{46}, v_{40}, v_{27}, v_{19}$ | 2 | 0.952 | 0.747 |

[a]Principal components

TABLE S-X. Statistics of PLS on variables selected by SMR (BTT)

| No. | Variable | $l$[a] | $R^2_{cum}$ | $Q^2_{LOO}$ |
|---|---|---|---|---|
| 1 | $v_{32}$ | 1 | 0.462 | 0.421 |
| 2 | $v_{32}, v_{1}$ | 1 | 0.692 | 0.646 |
| 3 | $v_{32}, v_{1}, v_{34}$ | 2 | 0.709 | 0.644 |
| 4 | $v_{32}, v_{1}, v_{34}, v_{36}$ | 2 | 0.882 | 0.825 |
| 5 | $v_{32}, v_{1}, v_{34}, v_{36}, v_{25}$ | 2 | 0.887 | 0.799 |
| 6 | $v_{32}, v_{1}, v_{34}, v_{36}, v_{25}, v_{6}$ | 2 | 0.891 | 0.812 |

TABLE S-X. Continued

| No. | Variable | $l^a$ | $R^2_{cum}$ | $Q^2_{LOO}$ |
|---|---|---|---|---|
| 7 | **$v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}$** | **2** | **0.911** | **0.818** |
| 8 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}$ | 2 | 0.913 | 0.812 |
| 9 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}$ | 3 | 0.926 | 0.795 |
| 10 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}, v_{20}$ | 3 | 0.927 | 0.812 |
| 11 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}, v_{20}, v_{23}$ | 3 | 0.92 | 0.814 |
| 12 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}, v_{20}, v_{23}, v_7$ | 3 | 0.926 | 0.772 |
| 13 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}, v_{20}, v_{23}, v_7, v_{18}$ | 3 | 0.921 | 0.659 |
| 14 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}, v_{20}, v_{23}, v_7, v_{18}, v_{28}$ | 3 | 0.920 | 0.672 |
| 15 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}, v_{20}, v_{23}, v_7, v_{18}, v_{28}, v_{22}$ | 3 | 0.921 | 0.641 |
| 16 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}, v_{20}, v_{23}, v_7, v_{18}, v_{28}, v_{22}$ $v_{21}$ | 3 | 0.926 | 0.606 |
| 17 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}, v_{20}, v_{23}, v_7, v_{18}, v_{28}, v_{22}$, $v_{21}, v_{24}$ | 3 | 0.919 | 0.623 |
| 18 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}, v_{20}, v_{23}, v_7, v_{18}, v_{28}, v_{22}$, $v_{21}, v_{24}, v_{35}$ | 3 | 0.930 | 0.589 |
| 19 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}, v_{20}, v_{23}, v_7, v_{18}, v_{28}, v_{22}$, $v_{21}, v_{24}, v_{35}, v_{30}$ | 3 | 0.917 | 0.553 |
| 20 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}, v_{20}, v_{23}, v_7, v_{18}, v_{28}, v_{22}$, $v_{21}, v_{24}, v_{35}, v_{30}, v_5$ | 3 | 0.927 | 0.607 |
| 21 | $v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6, v_{12}, v_{29}, v_{27}, v_{20}, v_{23}, v_7, v_{18}, v_{28}, v_{22}$, $v_{21}, v_{24}, v_{35}, v_{30}, v_5, v_9$ | 3 | 0.931 | 0.564 |

[a]Principal components

TABLE S-XI. Statistics of PLS on variables selected by SMR (ace)

| No. | Variable | $l^a$ | $R^2_{cum}$ | $Q^2_{LOO}$ |
|---|---|---|---|---|
| 1 | *v23* | 1 | 0.637 | 0.633 |
| 2 | *v23, v34* | 1 | 0.703 | 0.691 |
| 3 | *v23, v34, v6* | 1 | 0.756 | 0.734 |
| 4 | *v23, v34, v6, v30* | 1 | 0.786 | 0.768 |
| 5 | *v23, v34, v6, v30, v28* | 2 | 0.835 | 0.795 |
| 6 | *v23, v34, v6, v30, v28, v9* | 2 | 0.852 | 0.786 |
| 7 | ***v23, v34, v6, v30, v28, v9, v21*** | **2** | **0.865** | **0.799** |
| 8 | *v23, v34, v6, v30, v28, v9, v21, v29* | 2 | 0.852 | 0.777 |
| 9 | *v23, v34, v6, v30, v28, v9, v21, v29, v36* | 3 | 0.862 | 0.748 |
| 10 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17* | 3 | 0.858 | 0.757 |
| 11 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4* | 3 | 0.874 | 0.708 |
| 12 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18* | 3 | 0.873 | 0.694 |
| 13 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12* | 3 | 0.882 | 0.719 |
| 14 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31* | 3 | 0.883 | 0.710 |
| 15 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11* | 3 | 0.884 | 0.711 |
| 16 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25* | 3 | 0.883 | 0.709 |
| 17 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26* | 3 | 0.883 | 0.702 |

TABLE S-XI. Continued

| No. | Variable | $l^a$ | $R^2_{cum}$ | $Q^2_{LOO}$ |
|---|---|---|---|---|
| 18 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13* | 3 | 0.876 | 0.686 |
| 19 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20,* | 3 | 0.875 | 0.696 |
| 20 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19* | 3 | 0.876 | 0.692 |
| 21 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19, v2* | 3 | 0.873 | 0.692 |
| 22 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19, v2, v3* | 3 | 0.787 | 0.629 |
| 23 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19, v2, v3, v10* | 3 | 0.87 | 0.685 |
| 24 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19, v2, v3, v10, v8* | 3 | 0.873 | 0.664 |
| 25 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19, v2, v3, v10, v8, v5* | 3 | 0.871 | 0.655 |
| 26 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19, v2, v3, v10, v8, v5, v7* | 3 | 0.871 | 0.642 |
| 27 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19, v2, v3, v10, v8, v5, v7, v14* | 3 | 0.878 | 0.625 |
| 28 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19, v2, v3, v10, v8, v5, v7, v14, v16* | 3 | 0.879 | 0.62 |
| 29 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19, v2, v3, v10, v8, v5, v7, v14, v16, v15* | 3 | 0.883 | 0.619 |
| 30 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19, v2, v3, v10, v8, v5, v7, v14, v16, v15, v33* | 3 | 0.879 | 0.614 |
| 31 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19, v2, v3, v10, v8, v5, v7, v14, v16, v15, v33, v27* | 3 | 0.88 | 0.613 |
| 32 | *v23, v34, v6, v30, v28, v9, v21, v29, v36, v17, v4, v18, v12, v31, v11, v25, v26, v13, v20, v19, v2, v3, v10, v8, v5, v7, v14, v16, v15, v33, v27, v32* | 3 | 0.88 | 0.629 |

[a]Principal components