SUPPORTING INFORMATION

From canonical to unique: extension of a lipophilicity scale of amino acids to non-standard residues

Antonio Viayna[1*], Paulina Matamoros[2], David Blázquez-Ruano[3], William J. Zamora[2,4,5*]

[1]Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació, Institut de Biomedicina (IBUB), Universitat de Barcelona (UB), 08921 Santa Coloma de Gramenet, Spain

[2]CBio[3] Laboratory, School of Chemistry, University of Costa Rica, 11501-2060 San José, Costa Rica

[3]CIC bioGUNE, Basque Research and Technology Alliance, 48160 Derio, Spain

[4]Laboratory of Computational Toxicology and Biological Testing Laboratory (LEBi), University of Costa Rica, 11501-2060 San José, Costa Rica

[5]Advanced Computing Lab (CNCA), National High Technology Center (CeNAT), 10109 San José, Costa Rica

*Correspondence: Antonio Viayna, Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació, Institut de Biomedicina (IBUB), Universitat de Barcelona (UB), 08921 Santa Coloma de Gramenet, Spain toniviayna@ub.edu | William J. Zamora, CBio[3] Laboratory, School of Chemistry, University of Costa Rica, 11501-2060 San José, Costa Rica william.zamoraramirez@ucr.ac.cr.

Table S1. Dataset of methionine amino acids examined in this article derived from the research conducted by Kubyshkin. Chemical scaffolds based on Figure 1 from main text.

Amino acid code	Amino acid name	Side chain (R_1)
Met	Methionine	$-CH_2-CH_2-S-CH_3$
Nle	Norleucine	$-CH_2-CH_2-CH_2-CH_3$
Tfnle	Trifluoronorleucine	$-CH_2-CH_2-CH_2-CF_3$
Eth	Ethionine	$-CH_2-CH_2-S-CH_2-CH_3$
Nva	Norvaline	$-CH_2-CH_2-CH_3$
Aha	Azidohomoalanine	$-CH_2-CH_2-N_3$
Mox	Methoxinine	$-CH_2-CH_2-O-CH_3$
MetO	Methionine sulfoxide	$-CH_2-CH_2-SO-CH_3$

Table S2. Dataset of lysine amino acids examined in this article derived from the research conducted by Kubyshkin. Chemical scaffolds based on Figure 1 from main text.

Amino acid code	Amino acid name	Side chain (R_1)
(Boc)Lys	tert-butoxycarbonyllysine	$-CH_2-CH_2-CH_2-CH_2-NH-CO-CH_2-C-(CH_3)_3$
Sac	S-alylcysteine	$-CH_2-S-CH_2-CH=CH_2$
(Pro)Lys	N-propargyloxycarbonyllysine	$-CH_2-CH_2-CH_2-CH_2-NH-CO-CH_2-CH_2-C\equiv CH$
(Ac)Lys	N-acetyllysine	$-CH_2-CH_2-CH_2-CH_2-NH-CO-CH_3$

Table S3. Dataset of tyrosine amino acids examined in this article derived from the research conducted by Kubyshkin. Chemical scaffolds based on Figure 1 from main text.

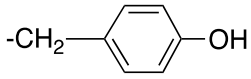
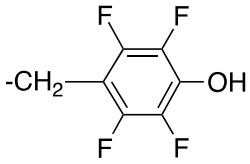
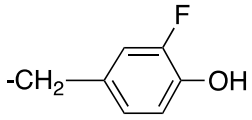
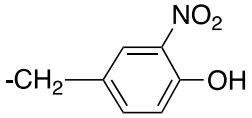
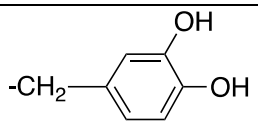
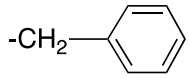
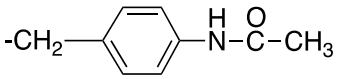
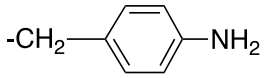
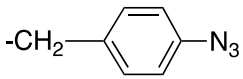
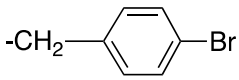
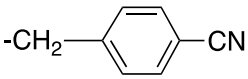
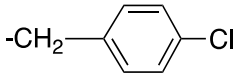
Amino acid code	Amino acid name	Side chain (R_1)
Tyr	Tyrosine	
2,3,5,6-tetraFTyr	2,3,5,6-tetrafluorotyrosine	
3-F-Tyr	3-fluorotyrosine	
3-NO2-Tyr	3-nitrotyrosine	
Dopa	Levodopa	

Table S4. Dataset of phenylalanine amino acids examined in this article derived from the research conducted by Kubyshkin. Chemical scaffolds based on Figure 1 from main text.

Amino acid code	Amino acid name	Side chain (R_1)
Phe	Phenylalanine	
4-AcNH-Phe	4-acetylamidophenylalanine	
4-NH2-Phe	4-aminophenylalanine	
4-N3-Phe	4-azidophenylalanine	
4-Br-Phe	4-bromophenylalanine	
4-CN-Phe	4-cyanophenylalanine	
4-Cl-Phe	4-chlorophenylalanine	

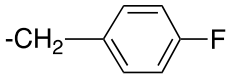
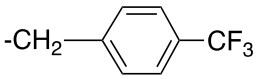
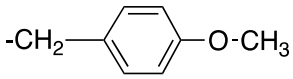
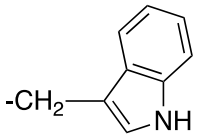
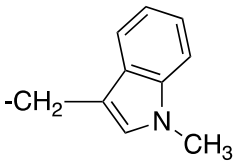
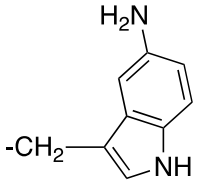
4-F-Phe	4-fluorophenylalanine	
4-CF3-Phe	4-trifluoromethylphenylalanine	
(Me)Tyr	Methyltyrosine	

Table S5. Dataset of tryptophan amino acids examined in this article derived from the research conducted by Kubyshkin. Chemical scaffolds based on Figure 1 from main text.

Amino acid code	**Amino acid name**	**Side chain (R_1)**
Trp	Tryptophan	
1-CH3-Trp	1-methyltryptophan	
5-NH2-Trp	5-aminotryptophan	

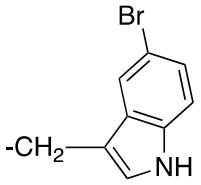
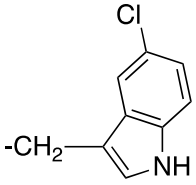
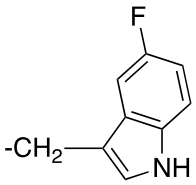
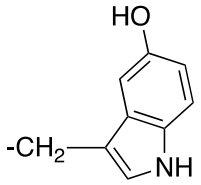
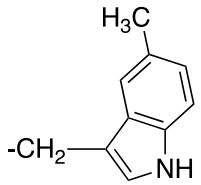
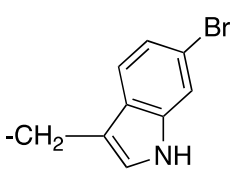
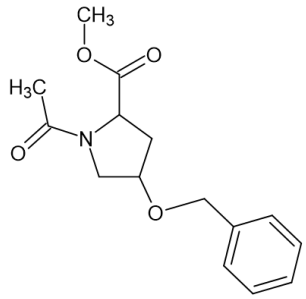
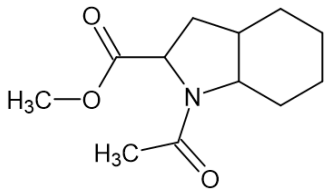
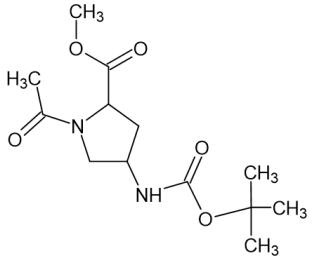
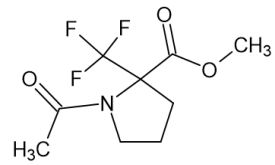
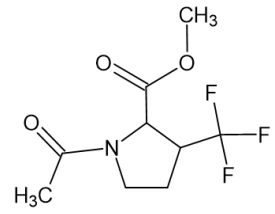
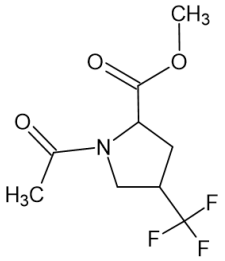
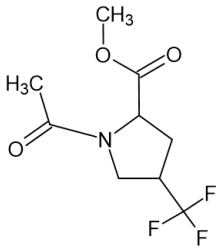
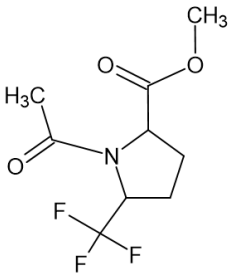
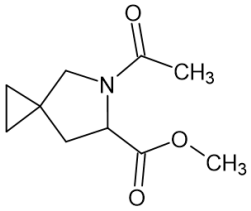
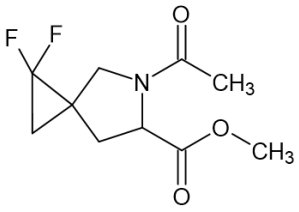
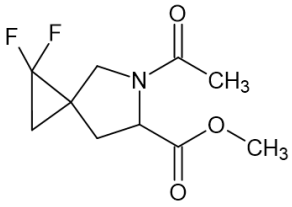
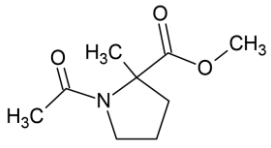
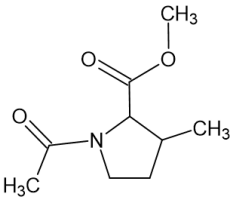
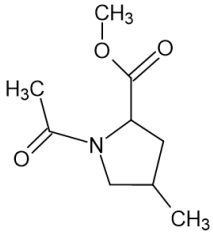
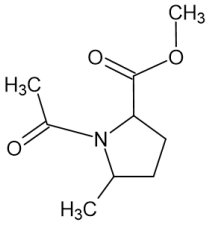
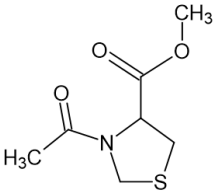
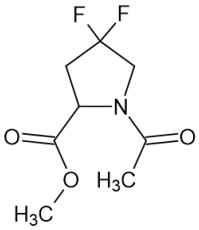
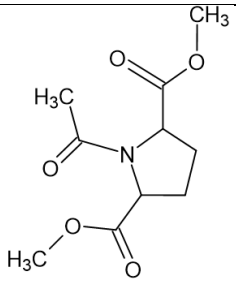
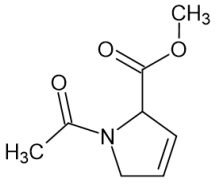
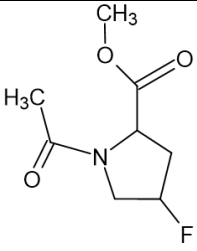
5-Br-Trp	5-bromotryptophan	
5-Cl-Trp	5-chlorotryptophan	
5-F-Trp	5-fluorotryptophan	
5-OH-Trp	5-hydroxytryptophan	
5-CH3-Trp	5-methyltryptophan	
6-Br-Trp	6-bromotryptophan	

Table S6. Dataset of proline amino acids examined in this article derived from the research conducted by Kubyshkin. Chemical scaffolds based on Figure S1.

Amino acid code	**Amino acid name**	**R1**
(Bn)Hyp	*O*-benzyl-hydroxyproline	
Oic	Octahydroindole-2-carboxylic acid	
(Boc)Amp	4-N-Bocproline	
2TfmPro	2-trifluoromethylproline	
3TfmPro	3-trifluoromethylproline	
4TfmProA	4-trifluoromethylproline A	

4TfmProB	4-trifluoromethylproline B	
5TfmPro	5-trifluoromethylproline	
Ash	5-spiroproline	
cF2Ash	1,1-cis-difluoro-5-spiroproline	
tF2Ash	1,1-trans-difluoro-5-spiroproline	
2Mep	2-methylproline	
3Mep	3-methylproline	

Mep	4-methylproline	
5Mep	5-methylproline	
Cys[ΨPro]	4-thioproline	
Dfp	4,4-difluoroproline	
(Me)rPrc	5-S-acetylproline	
Dhp	Dehydroproline	
Flp	4-fluoroproline	

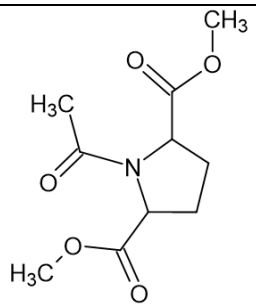
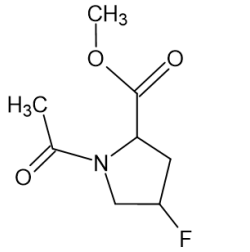
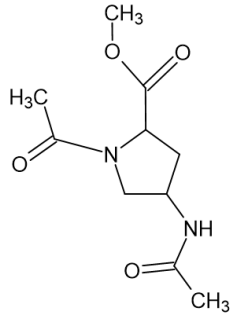
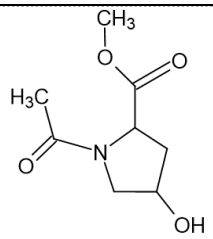
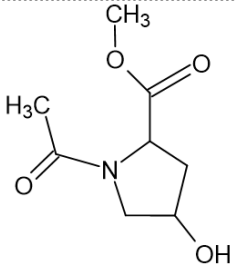
(Me)mPdc	5-R-acetylproline	
flp	4-fluoroproline	
(Ac)Amp	4-acetamidoproline	
Hyp A	4-hydroxyproline A	
Hyp B	4-hydroxyproline B	

Table S7. Summary of molecular descriptors of studied methionine derivatives for Models 1 and 2, estimated with DataWarrior.

Model 1				
Amino acid code	**Molecular Weight (MW) (g/mol)**	**Rotatable Bonds (RB)**	**Hydrogen Bond Acceptors (HBA)**	**Hydrogen Bond Donors (HBD)**
Met	204.29	5	4	2
Nle	186.25	5	4	2
Tfnle	240.22	6	4	2
Eth	218.32	6	4	2
Nva	172.23	4	4	2
Aha	200.22	5	7	3
Mox	188.23	5	5	2
MetO	222.31	5	5	3
Model 2				
Amino acid code	**Molecular Weight (MW) (g/mol)**	**Rotatable Bonds (RB)**	**Hydrogen Bond Acceptors (HBA)**	**Hydrogen Bond Donors (HBD)**
Met	205.28	6	4	1
Nle	187.24	6	4	1
Tfnle	241.21	7	4	1
Eth	219.30	7	4	1
Nva	173.21	5	4	1
Aha	201.21	6	7	2
Mox	189.21	6	5	1
MetO	223.29	6	5	2

Table S8. Summary of molecular descriptors of studied tyrosine derivatives for Models 1 and 2, estimated with DataWarrior.

Model 1				
Amino acid code	**Molecular Weight (MW) (g/mol)**	**Rotatable Bonds (RB)**	**Hydrogen Bond Acceptors (HBA)**	**Hydrogen Bond Donors (HBD)**
Tyr	236.27	4	5	3
2,3,5,6-tetraFTyr	308.23	4	5	3
3-F-Tyr	254.26	4	5	3
3-NO2-Tyr	281.27	5	7	3
Dopa	252.27	4	6	4
Model 2				
Amino acid code	**Molecular Weight (MW) (g/mol)**	**Rotatable Bonds (RB)**	**Hydrogen Bond Acceptors (HBA)**	**Hydrogen Bond Donors (HBD)**
Tyr	237.25	5	5	2
2,3,5,6-tetraFTyr	309.21	5	5	2
3-F-Tyr	255.24	5	5	2
3-NO2-Tyr	282.25	6	7	2
Dopa	253.25	5	6	3

Table S9. Summary of molecular descriptors of studied tyrosine derivatives for Models 1 and 2, estimated with DataWarrior.

Model 1				
Amino acid code	**Molecular Weight (MW) (g/mol)**	**Rotatable Bonds (RB)**	**Hydrogen Bond Acceptors (HBA)**	**Hydrogen Bond Donors (HBD)**
Phe	220.27	4	4	2
4-AcNH-Phe	277.32	5	6	3
4-NH2-Phe	235.29	4	5	3
4-N3-Phe	262.29	5	7	3
4-Br-Phe	299.17	4	4	2
4-CN-Phe	245.28	4	5	2
4-Cl-Phe	254.72	4	4	2
4-F-Phe	238.26	4	4	2
4-CF3-Phe	288.27	5	4	2
(Me)Tyr	250.30	5	5	2
Model 2				
Amino acid code	**Molecular Weight (MW) (g/mol)**	**Rotatable Bonds (RB)**	**Hydrogen Bond Acceptors (HBA)**	**Hydrogen Bond Donors (HBD)**
Phe	221.26	5	4	1
4-AcNH-Phe	278.31	6	6	2
4-NH2-Phe	236.27	5	5	2
4-N3-Phe	263.28	6	7	2
4-Br-Phe	300.15	5	4	1
4-CN-Phe	246.27	5	5	1
4-Cl-Phe	255.70	5	4	1
4-F-Phe	239.25	5	4	1
4-CF3-Phe	289.25	6	4	1
(Me)Tyr	251.28	6	5	1

Table S10. Summary of molecular descriptors of studied tryptophan derivatives for Models 1 and 2, estimated with DataWarrior.

Model 1				
Amino acid code	**Molecular Weight (MW) (g/mol)**	**Rotatable Bonds (RB)**	**Hydrogen Bond Acceptors (HBA)**	**Hydrogen Bond Donors (HBD)**
Trp	259.31	4	5	3
1-CH3-Trp	273.34	4	5	2
5-NH2-Trp	274.32	4	6	4
5-Br-Trp	338.20	4	5	3
5-Cl-Trp	293.75	4	5	3
5-F-Trp	277.30	4	5	3
5-OH-Trp	275.31	4	6	4
5-CH3-Trp	273.34	4	5	3
6-Br-Trp	338.20	4	5	3
Model 2				
Amino acid code	**Molecular Weight (MW) (g/mol)**	**Rotatable Bonds (RB)**	**Hydrogen Bond Acceptors (HBA)**	**Hydrogen Bond Donors (HBD)**
Trp	260.29	5	5	2
1-CH3-Trp	274.32	5	5	1
5-NH2-Trp	275.31	5	6	3
5-Br-Trp	339.19	5	5	2
5-Cl-Trp	294.74	5	5	2
5-F-Trp	278.28	5	5	2
5-OH-Trp	276.29	5	6	3
5-CH3-Trp	274.32	5	5	2
6-Br-Trp	339.19	5	5	2

Table S11. Summary of molecular descriptors of studied lysine derivatives for Models 1 and 2, estimated with DataWarrior.

Model 1				
Amino acid code	**Molecular Weight (MW) (g/mol)**	**Rotatable Bonds (RB)**	**Hydrogen Bond Acceptors (HBA)**	**Hydrogen Bond Donors (HBD)**
Lys	301.39	9	7	3
(Boc)Lys	216.30	6	4	2
Sac	269.30	8	7	3
(Pro)Lys	243.31	7	6	3
(Ac)Lys	301.39	9	7	3
Model 2				
Amino acid code	**Molecular Weight (MW) (g/mol)**	**Rotatable Bonds (RB)**	**Hydrogen Bond Acceptors (HBA)**	**Hydrogen Bond Donors (HBD)**
Lys	302.37	10	7	2
(Boc)Lys	217.29	7	4	1
Sac	270.28	9	7	2
(Pro)Lys	244.29	8	6	2
(Ac)Lys	302.37	10	7	2

Table S12. Summary of molecular descriptors of studied proline derivatives for Models 1 and 2, estimated with DataWarrior.

Model 1				
Amino acid code	**Molecular Weight (MW) (g/mol)**	**Rotatable Bonds (RB)**	**Hydrogen Bond Acceptors (HBA)**	**Hydrogen Bond Donors (HBD)**
(Bn)Hyp	276,.34	4	5	1
Oic	224.30	1	4	1
(Boc)Amp	285.34	4	7	2
2TfmPro	238.21	2	4	1
3TfmPro	238.21	2	4	1
4TfmProA	238.21	2	4	1
4TfmProB	238.21	2	4	1
5TfmPro	238.21	2	4	1
Ash	196.25	1	4	1
cF2Ash	232.23	1	4	1
tF2Ash	232.23	1	4	1
2Mep	184.24	1	4	1
3Mep	184.24	1	4	1
Mep	184.24	1	4	1
5Mep	184.24	1	4	1
Cys[ΨPro]	188.25	1	4	1
Dfp	206.19	1	4	1
(Me)rPrc	228.25	3	6	1
Dhp	168.20	1	4	1
Flp	188.20	1	4	1
(Me)mPdc	228.25	3	6	1
flp	188.20	1	4	1
(Ac)Amp	227.26	2	6	2
HypA	186.21	1	5	2
HypB	186.21	1	5	2
Model 2				

Amino acid code	Molecular Weight (MW) (g/mol)	Rotatable Bonds (RB)	Hydrogen Bond Acceptors (HBA)	Hydrogen Bond Donors (HBD)
(Bn)Hyp	277.32	5	5	0
Oic	225.29	2	4	0
(Boc)Amp	286.33	5	7	1
2TfmPro	239.19	3	4	0
3TfmPro	239.19	3	4	0
4TfmProA	239.19	3	4	0
4TfmProB	239.19	3	4	0
5TfmPro	239.19	3	4	0
Ash	197.23	2	4	0
cF2Ash	233.21	2	4	0
tF2Ash	233.21	2	4	0
2Mep	185.22	2	4	0
3Mep	185.22	2	4	0
Mep	185.22	2	4	0
5Mep	185.22	2	4	0
Cys[ΨPro]	189.23	2	4	0
Dfp	207.18	2	4	0
(Me)rPrc	229.23	4	6	0
Dhp	169.18	2	4	0
Flp	189.19	2	4	0
(Me)mPdc	229.23	4	6	0
flp	189.19	2	4	0
(Ac)Amp	228.25	3	6	1
HypA	187.19	2	5	1
HypB	187.19	2	5	1

Table S13. log P n-octanol/water for the 20 *N*-acetyl-L-amino-acid-*N*-methyl amides.

Amino acid	log *P*	log *P* backbone
GLY	-1.26	-1.04
ALA	-0.56	-0.82
LEU	0.90	-0.62
ILE	1.51	-0.55
MET	0.31	-0.49
VAL	-0.54	-0.52
PRO	-0.23	-0.58
PHE	1.26	-0.55
TYR	0.16	-0.66
TRP	1.02	-0.54
HIS	-0.95	-0.71
SER	-1.25	-0.50
THR	-0.78	-0.44
CYS	0.42	-0.77
ASN	-2.15	-0.75
GLN	-1.73	-0.66
ARG	-1.68	-0.81
GLU	-0.79	-0.63
LYS	0.17	-0.94
ASP	-1.57	-0.63

Table S14. Summary of logP values performed for methionine derivatives for Models 1 and 2.

Model 1				
Amino acid code	**logP (exptl.)**	**logP (calc.)**	**logP (calc. corr.)[a]**	**Δlog P (exptl − calc. corr.)**
Met	0.13	-0.16	0.01	0.12
Nle	0.93	0.47	0.64	0.29
Tfnle	0.72	0.42	0.59	0.13
Eth	0.56	0.03	0.20	0.36
Nva	0.40	0.43	0.60	-0.20
Aha	-0.11	-1.37	-1.20	1.09
Mox	-0.65	-0.82	-0.65	0.00
MetO	-1.79	-1.27	-1.10	-0.69
Model 2				
Amino acid code	**logP (exptl.)**	**logP (calc.)**	**logP (calc. corr.)[b]**	**Δlog P (exptl − calc. corr.)**
Met	0.13	1.05	0.27	-0.14
Nle	0.93	1.61	0.83	0.10
Tfnle	0.72	2.60	1.82	-1.10
Eth	0.56	1.99	1.21	-0.65
Nva	0.40	1.82	1.04	-0.64
Aha	-0.11	-0.12	-0.90	0.79
Mox	-0.65	1.24	0.46	-1.11
MetO	-1.79	-0.36	-1.14	-0.65

[a]logP corrected by adding +0.17 logP units. Difference of calculated Glycine reported in Zamora et al. 2019 against experimental value reported by Fauchere and Pliska in 1983. [b]logP corrected by subtracting -0.78 logP units. Difference comparing IEFPCM/MST calculated Glycine derivate against experimental value reported by Kubyshkin.

Table S15. Summary of logP values performed for tyrosine derivatives for Models 1 and 2.

Model 1				
Amino acid code	**logP (exptl.)**	**logP (calc.)**	**logP (calc. corr.)[a]**	**Δlog P (exptl − calc. corr.)**
Tyr	0.29	-0.19	-0.02	0.31
2,3,5,6-tetraFTyr	0.98	1.49	1.66	-0.68
3-NO2-Tyr	0.76	0.93	1.10	-0.34
3-F-Tyr	0.48	0.19	0.36	0.12
Dopa	-0.21	-0.95	-0.78	0.57
Model 2				
Amino acid code	**logP (exptl.)**	**logP (calc.)**	**logP (calc. corr.)[b]**	**Δlog P (exptl − calc. corr.)**
Tyr	0.29	1.30	0.52	-0.23
2,3,5,6-tetraFTyr	0.98	2.88	2.10	-1.12
3-NO2-Tyr	0.76	2.20	1.42	-0.66
3-F-Tyr	0.48	1.37	0.59	-0.11
Dopa	-0.21	0.55	0.52	-0.23

[a]logP corrected by adding +0.17 logP units. Difference of calculated Glycine reported in Zamora et al. 2019 against experimental value reported by Fauchere and Pliska in 1983.
[b]logP corrected by subtracting -0.78 logP units. Difference comparing IEFPCM/MST calculated Glycine derivate against experimental value reported by Kubyshkin.

Table S16. Summary of logP values performed for phenylalanine derivatives for Models 1 and 2.

Model 1				
Amino acid code	**logP (exptl.)**	**logP (calc.)**	**logP (calc. corr.)[a]**	**Δlog P (exptl − calc. corr.)**
Phe	0.92	0.44	0.61	0.31
4-CF3-Phe	1.96	1.75	1.92	0.04
4-Br-Phe	1.92	2.75	2.92	-1.00
4-Cl-Phe	1.76	1.65	1.82	-0.06
4-N3-Phe	1.58	-0.72	-0.55	2.13
4-F-Phe	1.16	1.35	1.52	-0.36
(Me)Tyr	0.92	1.25	1.42	-0.50
4-CN-Phe	0.58	-0.30	-0.13	0.71
4-AcNH-Phe	0.02	-0.52	-0.35	0.37
4-NH2-Phe	-0.33	-1.07	-0.90	0.57
Model 2				
Amino acid code	**logP (exptl.)**	**logP (calc.)**	**logP (calc. corr.)[b]**	**Δlog P (exptl − calc. corr.)**
Phe	0.92	2.38	1.60	-0.68
4-CF3-Phe	1.96	3.59	2.81	-0.85
4-Br-Phe	1.92	4.70	3.92	-2.00
4-Cl-Phe	1.76	3.09	2.31	-0.55
4-N3-Phe	1.58	1.04	0.26	1.32
4-F-Phe	1.16	2.57	1.79	-0.63
(Me)Tyr	0.92	2.26	1.48	-0.56
4-CN-Phe	0.58	1.18	0.40	0.18
4-AcNH-Phe	0.02	1.96	1.18	-1.16
4-NH2-Phe	-0.33	0.61	-0.17	-0.16

[a]logP corrected by adding +0.17 logP units. Difference of calculated Glycine reported in Zamora et al. 2019 against experimental value reported by Fauchere and Pliska in 1983. [b]logP corrected by subtracting -0.78 logP units. Difference comparing IEFPCM/MST calculated Glycine derivate against experimental value reported by Kubyshkin.

Table S17. Summary of logP values performed for tryptophan derivatives for Models 1 and 2.

Model 1				
Amino acid code	**logP (exptl.)**	**logP (calc.)**	**logP (calc. corr.)[a]**	**Δlog P (exptl − calc. corr.)**
Trp	1.20	1.20	1.37	-0.17
6-Br-Trp	2.37	3.54	3.71	-1.34
5-Br-Trp	2.34	3.90	4.07	-1.73
5-Cl-Trp	2.17	1.83	2.00	0.17
5-CH3-Trp	1.65	1.20	1.37	0.28
1-CH3-Trp	1.53	2.07	2.24	-0.71
5-F-Trp	1.48	1.60	1.77	-0.29
5-OH-Trp	0.24	0.25	0.42	-0.18
5-NH2-Trp	-0.26	-0.41	-0.24	-0.02
Model 2				
Amino acid code	**logP (exptl.)**	**logP (calc.)**	**logP (calc. corr.)[b]**	**Δlog P (exptl − calc. corr.)**
Trp	1.20	2.49	1.71	-0.51
6-Br-Trp	2.37	5.44	4.66	-2.29
5-Br-Trp	2.34	4.68	3.90	-1.56
5-Cl-Trp	2.17	2.57	1.79	0.38
5-CH3-Trp	1.65	2.81	2.03	-0.38
1-CH3-Trp	1.53	3.27	2.49	-0.96
5-F-Trp	1.48	3.13	2.35	-0.87
5-OH-Trp	0.24	1.06	0.28	-0.04
5-NH2-Trp	-0.26	0.79	0.01	-0.27

[a]logP corrected by adding +0.17 logP units. Difference of calculated Glycine reported in Zamora et al. 2019 against experimental value reported by Fauchere and Pliska in 1983. [b]logP corrected by subtracting -0.78 logP units. Difference comparing IEFPCM/MST calculated Glycine derivate against experimental value reported by Kubyshkin.

Table S18. Summary of logP values performed for lysine derivatives for Models 1 and 2.

Model 1				
Amino acid code	**logP (exptl.)**	**logP (calc.)**	**logP (calc. corr.)[a]**	**Δlog P (exptl − calc. corr.)**
Lys	0.97[c]	0.17[d]	1.11[d]	-0.14
(Boc)Lys	1.15	0.75	0.92	0.23
Sac	0.61	1.17	1.34	-0.73
(Pro)Lys	0.17	-0.62	-0.45	0.62
(Ac)Lys	-0.89	-1.81	-1.64	0.75
Model 2				
Amino acid code	**logP (exptl.)**	**logP (calc.)**	**logP (calc. corr.)[b]**	**Δlog P (exptl − calc. corr.)**
Lys	0.97[c]	-	0.97[c]	-
(Boc)Lys	1.15	1.65	0.87	0.28
Sac	0.61	2.07	1.29	-0.68
(Pro)Lys	0.17	1.35	0.57	-0.40
(Ac)Lys	-0.89	0.39	-0.39	-0.50

[a]logP corrected by adding +0.17 logP units. Difference of calculated Glycine reported in Zamora et al. 2019 against experimental value reported by Fauchere and Pliska in 1983.

[b]logP corrected by subtracting -0.78 logP units. Difference comparing IEFPCM/MST calculated Glycine derivate against experimental value reported by Kubyshkin.

[c]Experimental value of logP for butylamine (similar to side chain of Lys) reported in ref. 18 from the main text

[d]For Lysine in Model 1, +0.17 value is obtained by taking the logP value (considering capping groups) from Table S19 and subtracting the value -0.94 (referring to the backbone) from Table S19. Therefore, the value +1.11 refers only to the side chain.

Table S19. Summary of logP values performed for proline derivatives for Models 1 and 2.

Model 1				
Amino acid code	**logP (exptl.)**	**logP (calc.)**	**logP (calc. corr.)[a]**	**Δlog P (exptl − calc. corr.)**
Pro	-0.50	-0.23[c]	0.35[c]	-0.85
(Bn)Hyp	1.09	1.93	1.15	-0.06
Oic	1.03	2.46	1.68	-0.65
(Boc)Amp	0.66	0.59	-0.19	0.85
2TfmPro	0.41	1.78	1.00	-0.59
3TfmPro	0.35	1.25	0.47	-0.12
4TfmProA	0.23	1.24	0.46	-0.23
4TfmProB	0.24	0.81	0.03	0.21
5TfmPro	0.28	0.52	-0.26	0.54
Ash	0.19	0.85	0.07	0.12
cF2Ash	0.18	1.19	0.41	-0.23
tF2Ash	0.03	0.94	0.16	-0.13
2Mep	−0.06	0.78	0.00	-0.06
3Mep	−0.04	0.83	0.05	-0.09
Mep	−0.06	0.64	-0.14	0.08
5Mep	−0.14	0.44	-0.34	0.20
Cys[ΨPro]	−0.31	-0.09	-0.87	0.56
Dfp	−0.29	0.61	-0.17	-0.12
(Me)rPrc	−0.43	0.12	-0.66	0.23
Dhp	−0.47	-1.95	-2.73	2.26
Flp	−0.66	0.13	-0.65	-0.01
(Me)mPdc	−0.74	1.96	1.18	-1.92
flp	−0.84	-0.34	-1.12	0.28
(Ac)Amp	−1.23	-0.63	-1.41	0.18
HypA	−1.24	-0.87	-1.65	0.41
HypB	−1.43	-0.93	-1.71	0.28
Model 2				
Amino acid	**logP**	**logP**	**logP**	**Δlog P**

code	(exptl.)	(calc.)	(calc. corr.)[b]	(exptl − calc. corr.)
Pro	-0.50	1.36	0.58	-1.08
(Bn)Hyp	1.09	3.14	2.36	-1.27
Oic	1.03	2.83	2.05	-1.02
(Boc)Amp	0.66	1.18	0.40	0.26
2TfmPro	0.41	2.03	1.25	-0.84
3TfmPro	0.35	2.04	1.26	-0.91
4TfmProA	0.23	1.82	1.04	-0.81
4TfmProB	0.24	2.17	1.39	-1.15
5TfmPro	0.28	2.39	1.61	-1.33
Ash	0.19	1.99	1.21	-1.02
cF2Ash	0.18	2.03	1.25	-1.07
tF2Ash	0.03	2.48	1.70	-1.67
2Mep	−0.06	2.06	1.28	-1.34
3Mep	−0.04	1.59	0.81	-0.85
Mep	−0.06	1.30	0.52	-0.58
5Mep	−0.14	1.72	0.94	-1.08
Cys[ΨPro]	−0.31	1.28	0.50	-0.81
Dfp	−0.29	0.64	-0.14	-0.15
(Me)rPrc	−0.43	1.08	0.30	-0.73
Dhp	−0.47	1.30	0.52	-0.99
Flp	−0.66	0.86	0.08	-0.74
(Me)mPdc	−0.74	1.71	0.93	-1.67
flp	−0.84	0.89	0.11	-0.95
(Ac)Amp	−1.23	-0.15	-0.93	-0.30
HypA	−1.24	0.22	-0.56	-0.68
HypB	−1.43	-0.70	-1.48	0.05

[a]log*P* corrected by adding +0.17 logP units. Difference of calculated Glycine reported in Zamora et al. 2019 against experimental value reported by Fauchere and Pliska in 1983. [b]log*P* corrected by subtracting -0.78 logP units. Difference comparing IEFPCM/MST calculated Glycine derivate against experimental value reported by Kubyshkin.

[c]For Proline in Model 1, +0.35 value is obtained by taking the logP value (considering capping groups) from Table S19 and subtracting the value -0.58 (referring to the backbone) from Table S19. Therefore, the value +0.35 refers only to the side chain.

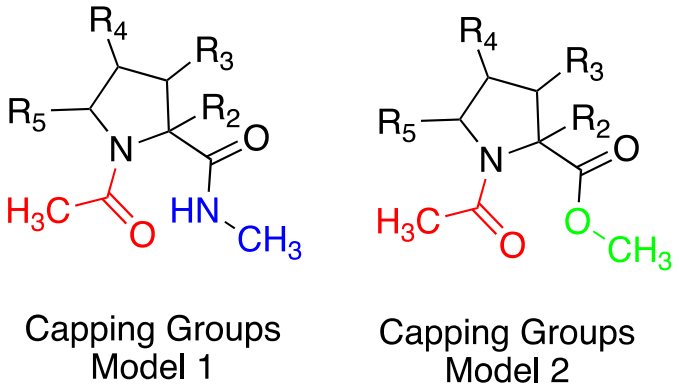


Figure S1. Chemical scaffolds of the capping group models used for the proline residues studied in this article: N-methyl (in blue), O-methyl (in green) and Acetyl group (in red).

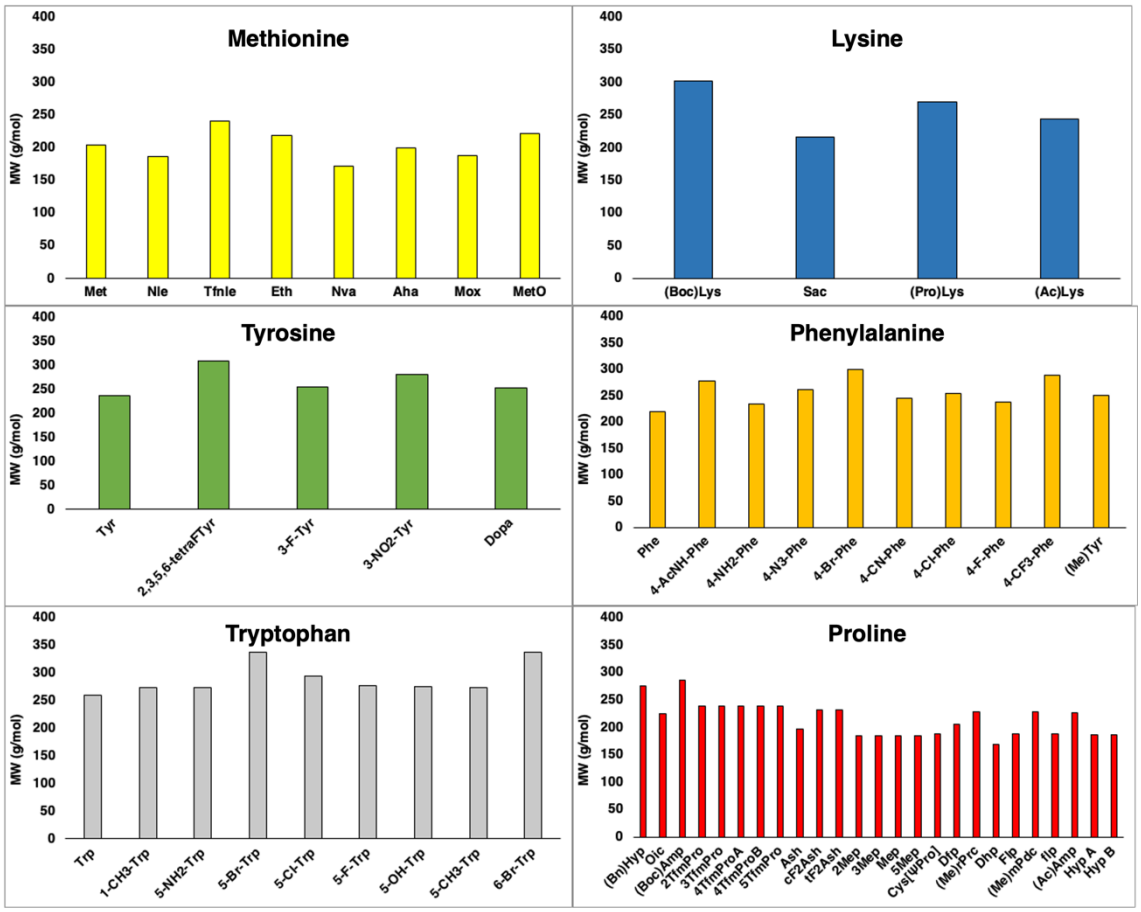


Figure S2. Distribution of values of molecular weight (MW) (in g/mol) from the studied compounds regarding Model 1, estimated with DataWarrior.

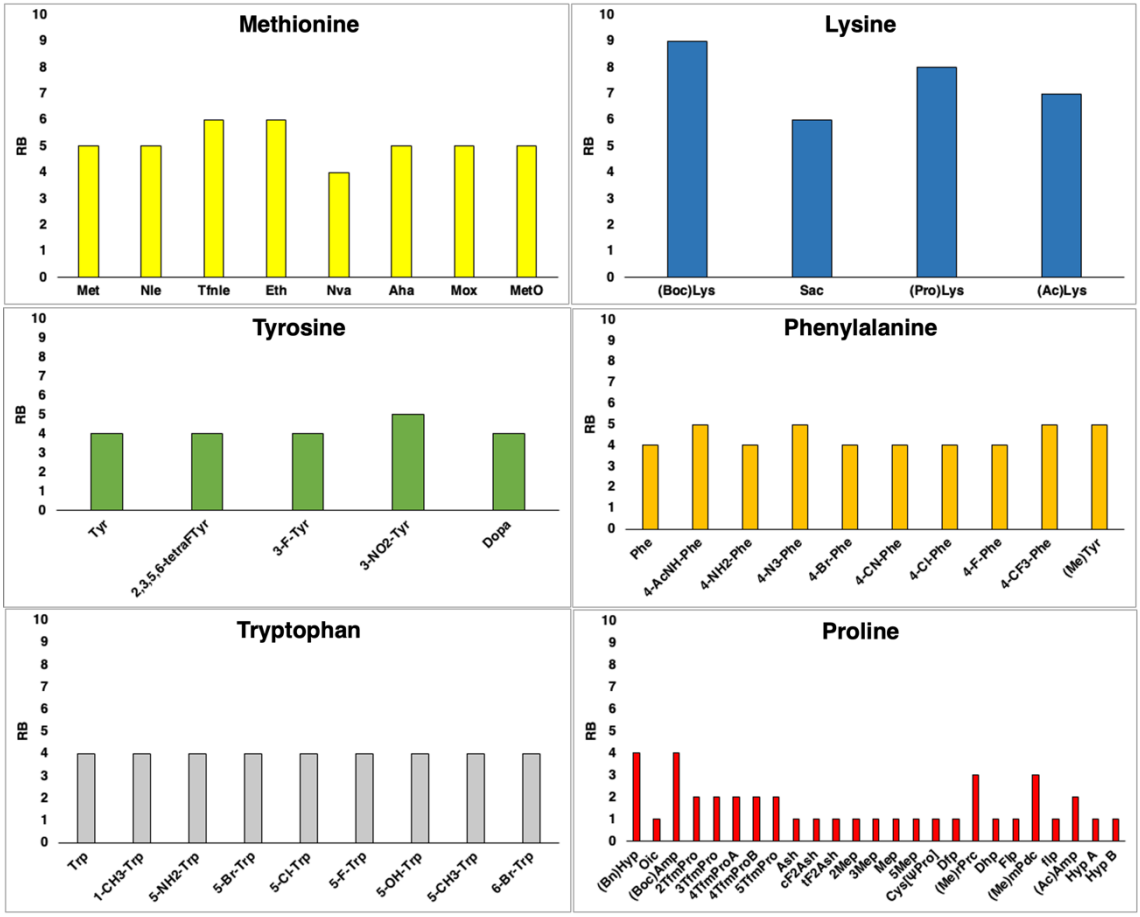


Figure S3. Distribution of values of number of rotatable bonds (RB) from the studied compounds, regarding Model 1, estimated with DataWarrior.

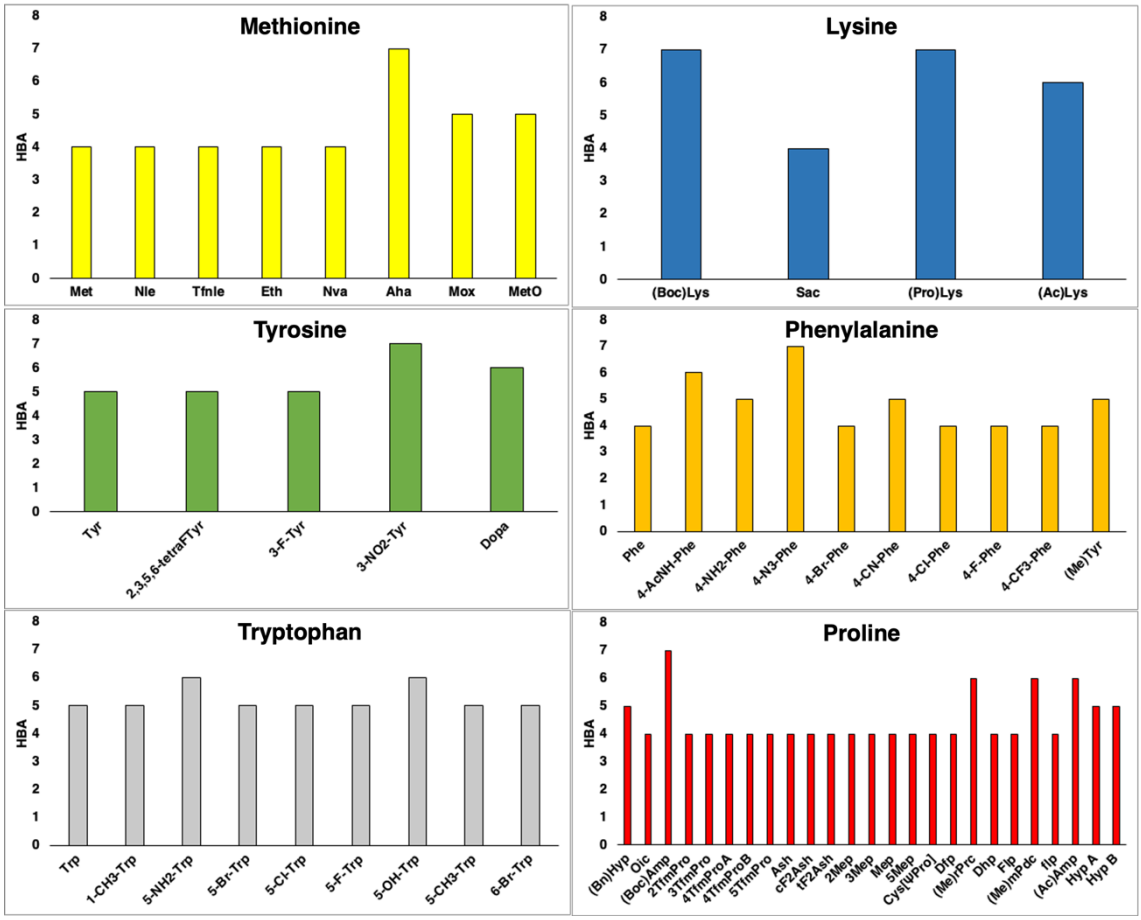


Figure S4. Distribution of values of number of hydrogen bond acceptors (HBA) from the studied compounds regarding Model 1, estimated with DataWarrior.

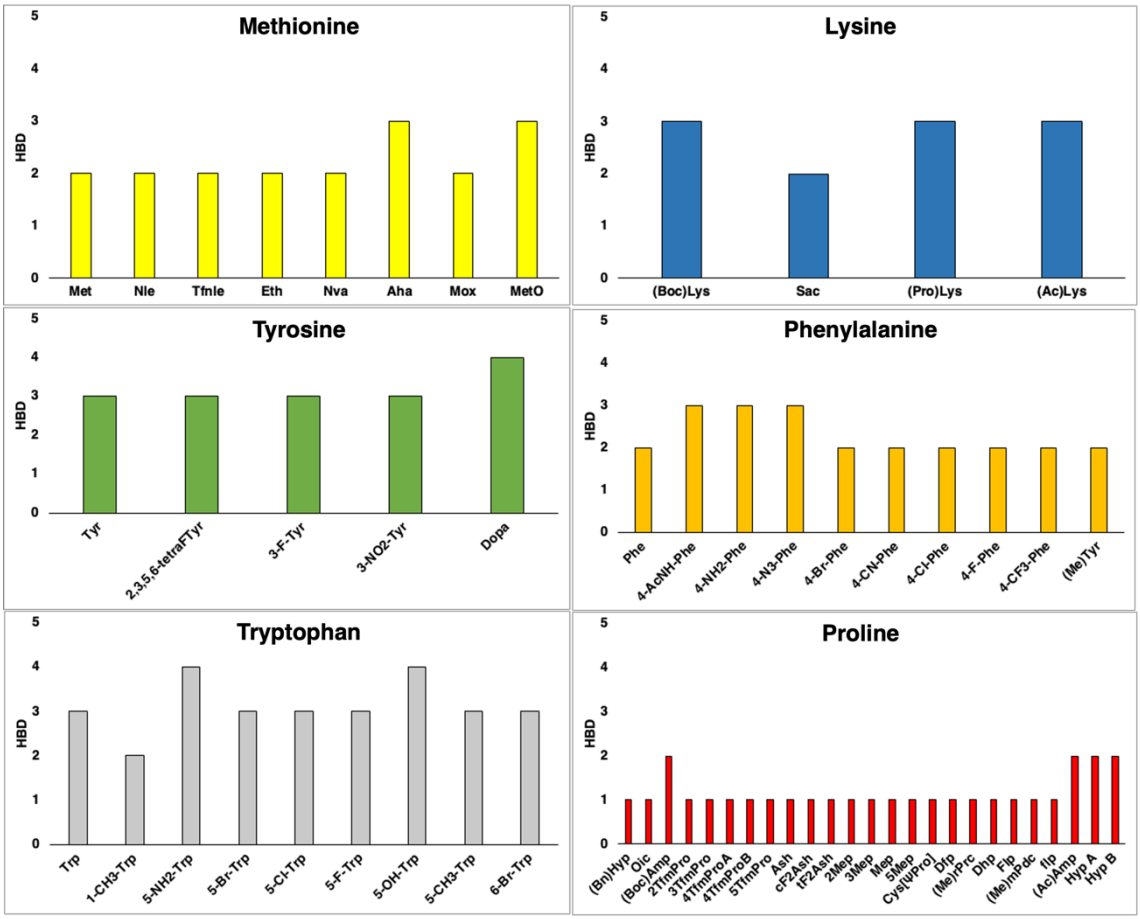


Figure S5. Distribution of values of number of hydrogen bond donors (HBD) from the studied compounds regarding Model 1, estimated with DataWarrior.

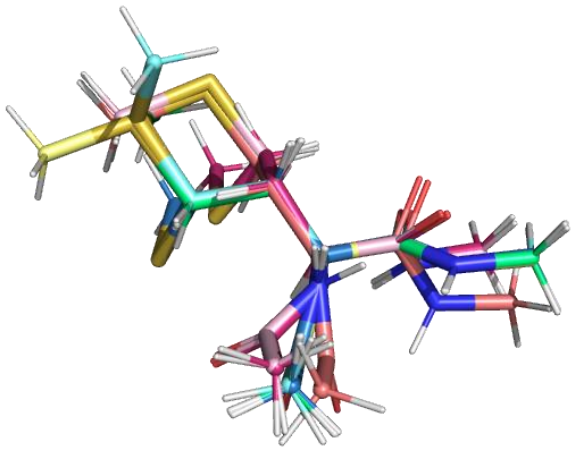


Figure S6. Water conformational landscape for amino acid methionine performed in this article. The rest of conformational families in water and n-octanol for both models can be found in the Github repository facilitated in the main text.