

QSAR Studies on amino-succinamic acid derivatives sweeteners

Laszlo Tarko^{a*}, Irina Lupescu^b, Diana Constantinescu-Groposila^b

^a Center of Organic Chemistry "C.D.Nenitzescu" – Romanian Academy, Spl. Independentei
202B, 6th Sector, Bucharest, PO Box 35-108, MC 060023, Fax 3121601,

^b Center of Applied Biochemistry and Biotechnologies "BIOTECHNOL", B-dul Marasti 59,
MC 011464, 1st Sector, Bucharest, Fax 2242815

E-mail: ltarko@cco.ro

Abstract

This paper presents some QSAR studies realized with the PRECLAV computer program. The database contains 3-amino-succinamic acid derivatives. The outlier molecules included in the calibration set were identified with a specific criterion. Without outliers $N = 121$, $s = 0.3989$, $r^2 = 0.8468$, $F = 128.2$, $K_{CV} = 0.7700$. The virtual molecular fragments that lead to a significant increase of the sweetness power SP are -CN (cyano) and -C₆H₄-NHCONH- (aryl-substituted urea). The non-conjugated or weakly conjugated virtual fragment -NH₂ leads to a significant decrease of the SP value. The sweetness power is favorably influenced by the size of the molecule. With a view to external validation, the calibration set includes 97 molecules ($s = 0.4248$, $r^2 = 0.8301$, $F = 89.9$, $K_{CV} = 0.7560$) and the validation set includes 24 molecules ($s = 0.3836$, $r^2 = 0.8580$, $K = 0.7609$). Identification of molecules in validation set with high estimated value of SP is correct enough to have practical value, even if the calibration/validation set contains 3-amino-succinamic acid derivatives with very different chemical structures.

Keywords: Sweeteners, QSAR, dipeptides, PRECLAV

Introduction

The discovery of L- α -aspartyl-L-phenylalanine 1-methyl ester sweet taste pointed out the sweetness potential of aspartyl-dipeptide derivatives and led to new researches on this class of compounds.^{1-15, 25-70} These researches focus on QSAR (Quantitative Structure Activity Relationship) studies, molecular modeling and conformational analysis and aim to establish the structural features of aspartyl-dipeptide derivatives that favour the activation of sweet taste receptors and to predict the sweetness potency of such molecules. Recent studies on molecular bases of sweet taste in the family of sweet taste dipeptide ligands revealed the fact that some N-arylalkyl and other substituted aspartyl dipeptide derivatives have a sweetness potency of

1.000– 50.000 higher than sucrose. This property could justify intensive research with the aim to obtain a compound close in quality to the ideal artificial sweetener: very high sweetness potency, pleasant taste, lack of toxicity, water solubility, thermal and chemical stability.

The following definitions are used throughout this article:

calibration set – a group of molecules that contains molecules with a known structure and known values of the sweetness power, a group of molecules used in calculating the QSAR; in the literature there are also used as synonyms “learning set” or “training set”

validation set – a group of molecules that contains molecules with a known structure and known values of the sweetness power, a group of molecules that is NOT used in calculating the QSAR; the equation obtained using the calibration set is used for calculating the sweetness power of the validation set; the concordance between the calculated values and the experimental values for the validation set is a measure of the accuracy of the algorithm (program) used.

prediction set - a group of molecules that contains molecules with a known structure and unknown values of the sweetness power; this group includes new structures, not synthesized yet; due to the fact that the values of the sweetness power for these molecules are unknown, the prediction set cannot be used in calculating the QSAR or for validation purposes; in literature we can find as synonyms “testing set”

descriptor – any molecular characteristic whose value is to be calculated

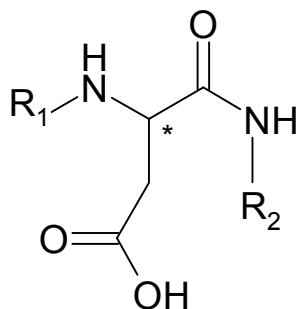
predictor – a descriptor present in the highest quality equation, the only equation used for prediction

This article presents the results of some QSAR studies done without a prediction set, with a calibration set and a validation set including derivatives of the 3-amino-succinamic acid.

Methods and formulae

The dependent property was sweetness power $SP = \text{Log}(1+RS)$, where RS is the relative sweetness to sucrose. The starting point of the computation was the database (136 derivatives of the 3-amino-succinamic acid) shown in Table 1, with the values of RS found in literature (see last column). Where two bibliographic sources indicate different values of RS for the same molecule, the highest value was taken into account. The configuration of the asymmetrical carbon atoms marked in the table refers to the configuration of the asymmetric carbon atom in the picture and the asymmetric carbon atoms from the R_2 group.

The molecules have been virtually constructed using the molecular mechanics program, PCMODEL.¹⁶ The geometry of the minimum energy conformer was obtained with the MMX force field and conformational analysis with GMMX algorithm included.

**Table 1.** 3-Amino-succinamic acid derivatives used in QSAR computations

Index	R ₁	R ₂	Config.	SP	Ref.
1	F ₃ C-CO-	C ₆ H ₅	S	1.114	25
2	F ₃ C-CO-	C ₆ H ₄ - <i>p</i> -F	S	0.301	26
3	F ₃ C-CO-	C ₆ H ₄ - <i>p</i> -CS	S	2.083	26
4	F ₃ C-CO-	C ₆ H ₄ - <i>p</i> -Br	S	2.083	27
5	F ₃ C-CO-	C ₆ H ₄ - <i>p</i> -CN	S	3.477	25
6	Cl ₃ C-CO-	C ₆ H ₅	S	0.301	25
7 ^b	Cl₃C-CO-	C₆H₄-<i>p</i>-F	S	0.845	27
8	Cl ₃ C-CO-	C ₆ H ₄ - <i>p</i> -CS	S	0.845	27
9	Cl ₃ C-CO-	C ₆ H ₄ - <i>p</i> -CN	S	3.477	27
10 ^{*a}	H	C ₆ H ₄ - <i>p</i> -CN	S	1.041	27
11	H	CH(CH ₂ C ₆ H ₅)COOCH ₃	S-S	2.303	1
12	H	CH(CH ₂ C ₆ H ₅)COOC ₂ H ₅	S-S	1.041	1
13 [*]	H	CH(CH ₂ C ₆ H ₅)COOC ₃ H ₇	S-S	0.301	1
14 [*]	H	CH(CH ₂ C ₆ H ₅)COO- <i>iso</i> Pr	S-S	0.301	1
15 [*]	H	CH(CH ₂ C ₆ H ₅)COO- <i>tert</i> Bu	S-S	0.301	1
16 [*]	H	CH(CH ₂ C ₆ H ₄ <i>p</i> OH)COOCH ₃	S-S	1.041	2
17	H	CH(CH ₂ CH ₂ SCH ₃)COOCH ₃	S-S	2.004	1
18	H	CH(CH ₃)COOCH ₃	S-R	1.415	2
19	H	CH(CH ₃)COOC ₂ H ₅	S-R	1.908	2
20	H	CH(CH ₃)COOC ₃ H ₇	S-R	2.233	2
21	H	CH(CH ₃)COO- <i>iso</i> Pr	S-R	2.100	2
22	H	CH(CH ₃)COOC ₄ H ₉	S-R	1.041	2
23	H	CH(CH ₃)COO- <i>cic</i> Hex	S-R	2.303	28
24	H	CH(CH₃)COO--2,6-diMe-<i>cic</i>Hex	S-R	2.303	28
25 [*]	H	CH(CH ₂ -2-furyS)COOCH ₃	S-R	4.041	6
26	H	CH(C ₂ H ₅)COO- <i>iso</i> Pr	S-R	2.233	2
27	H	CH(<i>iso</i>Pr)COO-<i>iso</i>Pr	S-R	2.233	2
28	H	CH(C₃H₇)COO-<i>iso</i>Pr	S-R	1.255	2

29	H	CH(C ₄ H ₉)COO-CH ₃	S-R	1.613	2
30	H	CH(C₅H₁₁)COO-CH₃	S-R	1.706	4
31	H	CH(CH(CH ₃)C ₃ H ₇)COO-CH ₃	S-R	1.785	12
32	H	CH(C ₆ H ₁₃)COO-CH ₃	S-R	1.881	12
33	H	CH(CH ₂ <i>cic</i> Hex)COO-CH ₃	S-R	2.354	29
34	H	CH(C ₂ H ₅)COO-CH(C ₂ H ₅)CH ₂ C ₆ H ₅	S-R	2.846	30
35	H	CH(C ₆ H ₅)COO-CH ₃	S-S	2.149	65
36	H	C(CH₃)(C₆H₅)COO-CH₃	S-S	2.179	66
37	H	C(CH ₃)(CH ₂ C ₆ H ₅)COO-CH ₃	S-S	2.303	67
38	H	CH(CH ₂ CH ₂ C ₆ H ₅)COO-CH ₃	S-S	2.004	65
39	H	C(CH ₃)(CH ₂ CH ₂ C ₆ H ₅)COO-CH ₃	S-S	2.179	66
40	H	1-metil-1-(4-metil-2,6,7-trioxa-biciclo[2,2,2]octan)ethyl	S-R	2.904	68
41	H	CH(CH ₂ bicicloC ₇ H ₁₁)COO-CH ₃	S-S	2.904	41
42	H	CH(CH ₂ -6Me- bicicloC ₇ H ₁₀)COO-CH ₃	S-S	3.000	41
43	H	CH(CH ₂ biciclo-7,7' diMe-C ₇ H ₉)COO-CH ₃	S-S	3.398	41
44	H	CH(CH ₂ biciclo-2Me-7,7' diMe-C ₇ H ₈)COO-CH ₃	S-S	3.602	41
45*	H	CH(CH ₂ -6Me-2,2' diMe- bicicloC ₇ H ₁₀)COO-CH ₃	S-S	1.322	41
46	H	CH(COO- <i>iso</i> Pe)COO-CH ₃	S-S	2.700	32
47	H	CH(COO- <i>cic</i> Pe)COO-CH ₃	S-S	2.779	32
48	H	CH(COO-<i>cic</i>Hex)COO-CH₃	S-S	2.955	32
49*	H	CH(COO-2Me- <i>cic</i> Hex)COO-CH ₃	S-S	3.845	32
50	H	CH(COO- biciclo-2Me-7,7' diMe-C ₇ H ₈)COO-CH ₃	S-S	4.521	32
51	H	CH(COO- <i>cic</i> Pe)COO-C ₂ H ₅	S-S	2.207	32
52	H	CH(COO-<i>cic</i>Hex)COO-C₂H₅	S-S	2.449	32
53	H	CH(COO-2Me- <i>cic</i> Hex)COO-C ₂ H ₅	S-S	2.814	32
54	H	CH(COO- biciclo-2Me-7,7' diMe-C ₇ H ₈)COO-C ₂ H ₅	S-S	3.716	32
55	(CH₂)₂C(CH₃)₂CH₃	C(CH₃)(CH₂C₆H₅-<i>p</i>OH)COO-CH₃	S-S	2.846	33
56	(CH ₂) ₂ C(CH ₃) ₃	C(CH ₃)(CH ₂ C ₆ H ₅)COO-CH ₃	S-S	3.740	33
57*	(CH ₂) ₂ C(CH ₃) ₃	CH(CH ₂ C ₆ H ₅)COO-CH ₃	S-S	4.000	7
58*	(CH ₂) ₂ C(CH ₃) ₃	CH(CH ₂ C ₆ H ₁₁)COO-CH ₃	S-S	4.079	9
59	(CH ₂) ₃ C ₆ H ₅	CH(CH ₂ C ₆ H ₅)COO-CH ₃	S-S	3.000	14
60	(CH ₂) ₂ C(CH ₃) ₂ C ₆ H ₅	CH(CH ₂ C ₆ H ₅)COO-CH ₃	S-S	3.602	14
61	(CH ₂) ₃ C ₆ H ₅ . <i>m</i> OH- <i>p</i> OMe	CH(CH ₂ C ₆ H ₅)COO-CH ₃	S-S	4.398	14
62	(CH₂)₂C(CH₃)₂	CH(CH₂C₆H₅)COO-CH₃	S-S	4.699	14

	C₆H₃mOH-pOMe				
63	(CH ₂) ₃ C ₆ H ₃ .oOH- pOH	CH(CH ₂ C ₆ H ₅)COO-CH ₃	S-S	4.000	69
64	(CH ₂) ₃ C ₆ H ₃ .mOH- pOH	CH(CH ₂ C ₆ H ₅)COO-CH ₃	S-S	4.176	69
65	(CH ₂) ₃ C ₆ H ₂ .oOH. mOH-pOH	CH(CH ₂ C ₆ H ₅)COO-CH ₃	S-S	4.000	69
66	(CH₂)₃ C₆H₂. mOH- m'OH-pOH	CH(CH₂C₆H₅)COO-CH₃	S-S	4.398	69
67	(CH ₂) ₂ C(CH ₃) ₂ C ₆ H ₃ . mOH-pOH	CH(CH ₂ C ₆ H ₅)COO-CH ₃	S-S	4.699	69
68	H	CH(CH ₃)CONHC ₆ H ₅	S-R	1.881	34
69	H	CH(CH ₃)CONH-2,6diMe-C ₆ H ₃	S-R	2.700	34
70	H	CH(CH ₃)CONHCH(CH ₃)C ₆ H ₅	S-R-S	2.258	35
71	H	CH(CH ₃)CONHCH(C ₆ H ₅) ₂	S-R	2.258	35
72	H	CH(CH ₃)CONH(2,2',4,4'-tetraMe-4-tio- ciclobutan)	S-R	3.301	36
73	H	CH(C ₂ H ₅)CONHCH(C ₂ H ₅)C ₆ H ₄ -pOH	S-R-S	3.301	35
74	H	CH(C ₂ H ₅)CONHCH(CH ₃)C ₆ H ₅	S-R-S	2.558	35
75	H	CH(C₂H₅)CONHCH(cicPr)₂	S-R	2.700	36
76*	H	CH(CH ₃)CONHCH(CH ₃)C ₆ H ₅	S-R-S	3.398	35
77	H	CH(C₂H₅)CONHCH(CH₂OCH₃)C₆H₅	S-R-R	3.398	14
78*	H	CH(C ₂ H ₅)CONHCH(C ₃ H ₇)C ₆ H ₅	S-R-S	3.301	38
79	H	C(CH ₃) ₂ CONHCH(C ₂ H ₅)C ₆ H ₅	S-R-S	2.303	10
80	H	CH(CH ₃)CONHCH(CH ₂ OCH ₃)C ₆ H ₅	S-R-R	2.479	10
81	H	CH(C₃H₇)CONHCH(CH₃)C₆H₅	S-R-S	1.959	35
82	H	CH(isoPr)CONHCH(CH₃)C₆H₅	S-R-S	2.733	35
83	H	CH(isoPr)CONHCH(cicPr)₂	S-R-S	2.045	36
84	H	CH(isoPr)CONHCH(C ₂ H ₅)C ₆ H ₅	S-R-S	3.176	35
85	H	CH(isoPr)CONHCH(CH ₂ OCH ₃)C ₆ H ₅	S-R-R	3.114	10
86	H	CH(isoPr)CONHCH(CH ₂ OC ₂ H ₅)C ₆ H ₅	S-R-R	2.700	10
87	H	CH(isoPr)CONHCH(isoPr)CH ₂ OCH ₃	S-R-S	1.613	10
88	H	C(CH ₃) (C ₂ H ₅) CONHCH(C ₂ H ₅)C ₆ H ₅	S-R-S	2.004	10
89	H	C(CH ₃) (CH ₃) CONHCH(C ₂ H ₅)C ₆ H ₅	S-S	3.000	10
90	H	CH(CH ₂ isoPr) CONHCH(C ₂ H ₅)C ₆ H ₅	S-R-S	2.004	64
91	H	CH(CH(CH ₃)C ₂ H ₅)CONHCH(C ₂ H ₅)C ₆ H ₅	S-R-S	2.700	64
92	H	CH(CH(CH ₃)C ₂ H ₅)CONHCH(CH ₂ OCH ₃)C ₆ H ₅	S-R-R	2.700	64
93	H	CH(cicPe)CONHCH(C ₂ H ₅)C ₆ H ₅	S-R-S	3.000	10
94	H	CH(cicHex)CONHCH(C₂H₅)C₆H₅	S-R-S	3.000	10

95	H	CH(cicHex)CONHCH(CH ₂ OCH ₃)C ₆ H ₅	S-R-R	2.400	64
96	H	CH(C ₆ H ₅)CONHCH(CH ₂ OCH ₃)C ₆ H ₅	S-R-R	2.700	64
97	H	CH(2-furil) CONHCH(C ₂ H ₅)C ₆ H ₅	S-R-S	3.080	10
98	H	CH(2-furil) CONHCH(CH ₂ OCH ₃)C ₆ H ₅	S-R-R	3.080	10
99	H	CH(C ₆ H ₅)CONHCH(CH ₃)C ₆ H ₅	S-R-S	2.433	35
100	(CH ₂) ₂ C(CH ₃) ₂ CH ₃	CH(C ₂ H ₅)CONHCH(CH ₃)C ₆ H ₅	S-R-S	3.301	39
101	(CH₂)₂C(CH₃)₂CH₃	CH(C₂H₅)CONHCH(C₂H₅)C₆H₅	S-R-S	3.903	39
102	(CH ₂) ₂ C(CH ₃) ₂ CH ₃	CH(C ₂ H ₅)CONHCH(CH ₂ OCH ₃)C ₆ H ₅	S-R-R	3.602	39
103	(CH₂)₂C(CH₃)₂CH₃	CH(<i>isoPr</i>)CONHCH(C₂H₅)C₆H₅	S-R-S	3.477	39
104	(CH₂)₂C(CH₃)₂CH₃	CH(<i>isoPr</i>)CONHCH(CH₂OCH₃)C₆H₅	S-R-R	3.602	39
105	(CH ₂) ₂ C(CH ₃) ₂ CH ₃	CH(CH ₃)CONHC(cicPr) ₂ CH ₃	S-R-R	3.398	40
106	H	CH(CH)₃-CH(OH)- CH₂-biciclo-2Me-6,6'diMe-C₇H₆)	S-R-S	3.256	41
107	H	CH(CH) ₃ -CH(OH)- CH ₂ -biciclo-C ₇ H ₉)	S-R-S	2.303	41
108	H	CH(CH) ₃ -CH(OH)- CH ₂ -biciclo-6,6'diMe-C ₇ H ₇)	S-R-S	2.779	41
109	H	CH(CH) ₃ -CH(OH)- CH ₂ -biciclo-2Me-C ₇ H ₈)	S-R-S	2.904	41
110	H	CH(CH) ₃ -CH(OH)- CH ₂ -biciclo-6Me-C ₇ H ₈)	S-R-S	2.930	41
111	HCO-NH-CO-	CH(CH ₂ C ₆ H ₅)(COOCH ₃)	S-S	2.303	63
112	HCO-NH-CO-	C(CH₃)(CH₂C₆H₅)(COOCH₃)	S-S	2.303	63
113	HCO-NH-CO-	CH(C ₆ H ₅)(COOCH ₃)	S-S	2.303	62
114	HCO-NH-CO-	OH instead of NH-R ₂	S	3.477	42
115	<i>p</i> -NC-C ₆ H ₄ -NH-CO-	CH(CH ₂ C ₆ H ₅)(COOCH ₃)	S-S	3.903	42
116	<i>p</i> -NO ₂ -C ₆ H ₄ -NH-CO-	C ₆ H ₄ - <i>p</i> -Cl	S	3.732	43
117*	<i>p</i> -NO ₂ -C ₆ H ₄ -NH-CO-	C ₆ H ₄ - <i>p</i> -CN	S	3.681	43
118	<i>p</i> -NC-C ₆ H ₄ -NH-CO-	C ₆ H ₄ - <i>m</i> -CF ₃	S	4.602	44
119	<i>p</i> -NC-C ₆ H ₄ -NH-CO-	C ₆ H ₄ - <i>m</i> -CH ₃	S	4.477	44
120	<i>p</i> -NC-C ₆ H ₄ -NH-CO-	C ₆ H ₄ - <i>m</i> -Cl	S	4.477	44
121	<i>p</i>-NC-C₆H₄-NH-CO-	C₆H₄-<i>m</i>-OH	S	4.301	44
122	<i>p</i> -NC-C ₆ H ₄ -NH-CO-	C ₆ H ₄ - <i>m</i> -OCH ₃	S	4.301	44
123	<i>p</i>-NC-C₆H₄-NH-CO-	C₆H₄-<i>m</i>-CH₂OH	S	4.000	44
124	<i>p</i> -NC-C ₆ H ₄ -NH-CO-	C ₆ H ₄ - <i>m</i> -F	S	4.000	44
125	<i>p</i> -NC-C ₆ H ₄ -NH-CO-	C ₆ H ₄ - <i>m</i> -NO ₂	S	4.000	44
126	<i>p</i> -NC-C ₆ H ₄ -NH-CO-	C ₆ H ₄ - <i>m</i> -CH ₂ CH ₃	S	3.954	44
127	<i>p</i> -NC-C ₆ H ₄ -NH-CO-	C ₆ H ₄ - <i>m</i> -Br	S	3.845	44
128	<i>p</i> -NC-(2-piridil)-NH-CO	C ₆ H ₄ - <i>m</i> -CH ₃	S	4.301	45
129	<i>p</i> -NC-(2-piridil)-NH-	C ₆ H ₄ - <i>m</i> -Cl	S	4.301	45

	CO				
130	<i>p</i> -NC-C ₆ H ₄ -NH-CO-	C ₆ H ₄ - <i>m</i> -CH ₃ - <i>p</i> CH ₃	S	4.699	45
131	-CO-(1-tetralinil)	(3-piridinil)-2CN	S	4.845	46
132 [*]	H	CH(CH ₃)CONHCH(CH ₃)C ₆ H ₅	S-R-R	0.778	35
133 [*]	H	CH(<i>iso</i> Pr)CONHCH(CH ₂ OCH ₃)C ₆ H ₅	S-R-S	1.041	10
134	H	CH(<i>iso</i> Pr)CONHCH(<i>iso</i> Pr)CH ₂ OCH ₃	S-R-R	1.491	10
135	H	CH(CH ₃) ₃ -CH(OH)- CH ₂ -biciclo-2Me-6,6' diMe-C ₇ H ₆)	S-R-R	4.097	41
136	H	CH(CH ₃) ₃ -CH(OH)- CH ₂ -biciclo-6Me-C ₇ H ₈)	S-R-R	3.398	41

^a index with asterisk * - outlier molecules identified in QSAR study #1

^b bold rows – molecules of validation set in QSAR study #3

Afterwards, the geometry was more rigorously optimized with the quantum mechanics program MOPAC,¹⁷ using the keyword string: “pm3 pulay gnorm=0.01 shift=50 geo-ok mmok camp-king bonds vectors”. The computations have been done for the neutral molecules, not for the cations, anions or zwitterions, although these ionic species are, probably, present in the aqueous solution. Thus the errors in geometry optimization are smaller and the predictive power of QSARs is better.

For the statistical computations an improved version of the PRECLAV^{18,19} (Property Evaluation by Class Variables) has been used. The output files created by MOPAC for each analyzed molecule are input files for PRECLAV and they contain the values of some descriptors. Using the data from the files generated by MOPAC, PRECLAV has computed most of the descriptors and has done the statistical analysis.

We have used only “whole molecule” PRECLAV descriptors.³¹ The methods for identifying “significant” descriptors, the quality criteria for the descriptors, the method for grouping the significant descriptors in sets and the quality criteria for the calculated QSARs have been presented in previous work.^{18,31} In the situation where there is no validation/prediction set, the “significant” descriptors are the descriptors that are sufficiently correlated with the dependent property ($r^2 > 4/N$, N is number of molecules in calibration set).

The computed QSARs are multilinear.

$$SP = c_0 + \sum c_k \cdot p_k \quad (1)$$

Weighting factors c_k of predictors p_k are computed by Ordinary Least Square Method.

Identification of the outlier molecules

The outlier molecules are molecules for whom the QSAR resulted from computations offers only a poor estimation of the sweetness power, although for the rest of the molecules in the calibration set the estimates have been good. The presence of the outlier molecules lowers the predictive

quality of the whole calibration set and often determines the inclusion into the final equation of a different set of predictors.

In order to identify these molecules, PRECLAV uses a specific criteria, called COIN (*Combined Outlier Index*).

COIN value is the product of two factors:

$$\text{COIN} = O_{\text{value}} \cdot O_{\text{rank}} \quad (2)$$

where:

O_{value} is a usual criterion for identifying outlier molecules based on comparing the calculated and the experimental values of the dependent property

$$O_{\text{value}} = \Delta_{\text{value}} / (2 \cdot s) \quad (3)$$

$$\Delta_{\text{value}} = |\text{SP}_{\text{obs}} - \text{SP}_{\text{calc}}|_{\text{value}}$$

$s = [\sum (\text{SP}_{\text{obs}} - \text{SP}_{\text{calc}})^2 / (N - p)]^{1/2}$ is standard error of estimation

N is the number of molecules in the calibration set

p is the number of predictors

O_{rank} is an identification criterion for the outlier molecules based on comparing the rank of the molecules in the set ordered by the calculated or experimental values

$$O_{\text{rank}} = \Delta_{\text{rank}} / [4 \cdot (N - p)]^{1/2} \quad (4)$$

$$\Delta_{\text{rank}} = |\text{SP}_{\text{obs}} - \text{SP}_{\text{calc}}|_{\text{rank}}$$

If the difference Δ_{value} between the calculated and the experimental value is big – compared with the standard error of estimation s – then O_{value} will be large. The molecule for which the value of $O_{\text{value}} > 1$ may be considered “outlier by residue”. If the difference Δ_{rank} between the calculated and experimental values of the ranks is big – compared with the number of degrees of freedom of the equation – then the value of O_{rank} is also big. The molecule for which the value of $O_{\text{rank}} > 1$ may be considered “outlier by rank”.

Some molecules in the calibration set may give very high values of O_{value} , but low values of O_{rank} or the other way round. Here only the molecules with $\text{COIN} > 1$ have been considered outliers and have been eliminated from further computations.

Obtaining the homogenized calibration set

The homogenized calibration set has been obtained by eliminating the outlier molecules from the original database.

Identifying the significant virtual molecular fragments

PRECLAV divides the analyzed molecules into virtual fragments, using an algorithm presented in some previous papers.^{20,21} Two bonded heavy atoms (by a chemical bond of bond order value B) are included into the same fragment if $B > k$ (k limit value depend of computation method for B). The classic conjugated functional groups form, according to the program, a single fragment, of higher mass. Any fragment identified by the program is non-conjugated or has low conjugation with the neighboring fragments. The virtual fragments identified by PRECLAV do not always coincide with the classical functional groups.

For each fragment the program calculates the weight ratio in the analyzed molecule. For a certain fragment, within the N molecules in the calibration set, the ratios p_1, p_2, \dots, p_N are calculated. The program calculates the linear correlation r between the values of p_k ratios and the values of the sweetness power. If the value of r^2 is greater than a predefined limit, then the fragment is considered as "significant". The presence of a significant fragment in the molecule greatly influences (in a positive or negative way) the sweetness power of the molecule.

The program calculates a type (1) QSPR equation where the variables (maximum 10) are the p_k ratios of the significant fragments. The value of the sum in equation (1) is, in this case, the value of the descriptor "QSPR of mass fragment percents".

Identification of the parabolic descriptors

PRECLAV calculates for the sweetness power SP multilinear equations of type (1). Although it is difficult to believe that the sweetness power dependency on various molecular features (descriptors) is linear, these equations have a reasonable predictive value. This is due to the following facts:

- the presence of a certain number of predictors in the equations (the errors induced by each may compensate)
- the intrinsic nonlinear character of some predictors
- acceptable approximations, at least on certain intervals, of some nonlinear functions (exponential, logarithmic, hyperbolic, etc.) by linear functions

If the dependence of the sweetness power on a certain descriptor is parabolic, and the parabola has a peak point (maximum or minimum) between the minimum and maximum value of the descriptor, then the linear function of the descriptor has low predictive abilities. In other words, if there is an optimal value of the analyzed descriptor from the point of view of the sweetness power, the linear function is incapable of finding this optimal value.

The quality q_{lin} of the linear function and the quality q_{par} of the parabolic function of certain descriptor are calculated in a specific way by PRECLAV.¹⁸ The replacement of the linear function with a parabolic function is done only if this replacement leads to a significant increase in the quality of the descriptor, according to (5):

$$q_{par} > 1.5 \cdot q_{lin} \quad (5)$$

The relative utility of the predictors

In the case when there is no validation/prediction set, the program calculates the relative utility U of the predictors, using equation (6):

$$U = (R^2 - r^2) / (1 - r^2) \quad (6)$$

where:

R^2 is the square of the Pearson correlation between the experimental and calculated values of SP (values calculated using an equation with p predictors)

r^2 is the square of the Pearson correlation between the experimental and calculated values of SP (values calculated using an equation with $p-1$ predictors, that is the equation that does not contain the analyzed predictor)

After calculating the value of U for all the predictors, these values are normalized according to the highest U (the highest value becomes 1000). The predictors with a high value of U ($U > 400$) may be considered very useful in calculating the sweetness power SP. These predictors are useful as they correlate very well with SP and do not correlate with the other predictors. Each “useful” predictor explains (quite) a lot of the SP variation and, in the same time, a different thing as the other predictors.

Validation of the computation procedure

For the validation of the method, we have proceeded to a QSAR study with a validation set and reduced calibration set.

The validation set was extracted from the homogenized calibration set. For the extraction, the molecules in this set have been ordered according to the observed value of SP, starting from the lowest values. For equal values, the order was arbitrary. In this set, the molecules with rank 3, 8, 13, 18, 23, 28, 33, ..., etc. have constituted the validation set. The remaining molecules form the reduced calibration set. The validation set includes approximately 20% of the molecules in the homogenized calibration set. We can assume that the reduced calibration set obtained in this way is a representative sample for the homogenized calibration set. The quality of the prediction for the validation set was considered a measure of the quality of the computation method.

The external validation procedure presented here must not be mistaken with the LOO (*Leave One Out*) internal cross-validation method used by PRECLAV in various stages of the computation.

Results and Discussion

Identification of the outlier molecules

In order to identify the outlier molecules, an initial QSAR study had been done using as calibration set the initial database.

QSAR study #1

Calibration set: 136 molecules (all Table 1 molecules)

Validation set: none

Number of significant descriptors: 251

The type (1) QSAR for prediction:

$$c_0 = -3.8874$$

$$c_1 = .0579$$

p_1 – Platt topologic index (U = 1000)

$$c_2 = .3517$$

p_2 – QSPR of mass fragment percents (U = 315)

$$c_3 = 5.8061$$

p_3 – Maximum free valence of N atoms (U = 473)

$$c_4 = 30.4586$$

p_4 – Average free valence of C atoms (U = 233)

Standard error s: 0.6516

Pearson square correlation r^2 : 0.6547

Fisher function F: 62.6

Kendall rank correlation K = .6769

Kendall cross-validated rank correlation K_{CV} : 0.6638

The lowest correlation with SP is calculated for predictor p_4 ($r^2 = 0.0945$). The highest intercorrelation between predictors is calculated for the pair p_1, p_4 ($r^2 = 0.3359$).

Considering the values of s , r^2 , K and K_{CV} the estimative quality of SP for the whole molecular set is mediocre.

All practical QSAR studies are faced with the problem of correctly identifying the outlier molecules in the calibration set. The presence of these molecules makes it difficult to obtain results with practical value, respectively:

- a) correct identification of the molecular characteristics with high influence over the dependent property, very necessary in drug design
- b) correct identification in the prediction set of the molecules “recommended for synthesis”

The process of identifying and eliminating the outlier molecules precedes any other computation in “drug design” procedures. Consequently, the authors have considered that QSAR studies (with or without a validation set) will lead to erroneous results if the calibration/validation sets contain outlier molecules. Studies #1 and #2 have been performed in order to identify these outlier molecules and to underline the effects of their removal from the calibration set.

In the group of molecules in Table 1, 15 outlier molecules have been identified (COIN >1), that is 11% of the molecules in the calibration set. The index in Table 1 corresponding to these molecules has been marked with asterisks (first column in Table 1).

The outlier molecules are:

10 – the only aromatic amido group that has $R_1 = H$

25 – the only furyl group, with a very high SP value

13, 14, 15, 16, 45, 132, 133 – quite plain in structure, but with very small values of SP

49, 57, 58, 76, 78, 117 – quite plain in structure, but with very high values of SP

None of the molecules containing $R_1 = X_3C-CO$ has been identified as outlier. Also molecule **114**, the only diacid in the set, hasn't been identified as outlier.

Obtaining the homogenized calibration set

The homogenized calibration set (121 molecules) has been obtained by eliminating the 15 outlier molecules from the initial database.

Identification of the significant virtual molecular fragments

Another QSAR analysis has been performed, using the homogenized calibration set.

QSAR study #2

Calibration set: 121 molecules (Table 1 molecules, without outliers)

Validation set: none

Number of significant descriptors: 258

The type (1) QSAR for prediction:

$$c_0 = -1.6131$$

$$c_1 = .0633$$

p_1 – Platt topologic index (U = 1000)

$$c_2 = -.9008$$

p_2 – Maximum electrostatic force in NH bonds - parabolic function (U = 176)

$$c_3 = .3799$$

p_3 – QSPR of mass fragment percents (U = 455)

$$c_4 = 4.908$$

p_4 – Maximum free valence of N atoms (U = 448)

$$c_5 = 28.5438$$

p_5 – Average free valence of C atoms (U = 297)

Standard error s: 0.3989

Pearson square correlation r^2 : 0.8468

Fisher F function: 128.2

Kendall rank correlation K = .7821

Kendall cross-validated rank correlation K_{CV} : 0.7700

The lowest correlation with SP is calculated for predictor p_2 ($r^2 = 0.0383$). The highest intercorrelation between predictors is calculated for the pair p_1, p_5 ($r^2 = 0.3505$). Considering the values of s , r^2 , K and K_{CV} the estimative quality of SP for the whole molecular set significantly improved. The decrease in scattering of the experimental/calculated values after eliminating the outlier molecules is obvious (Fig. 1 and Fig. 2).

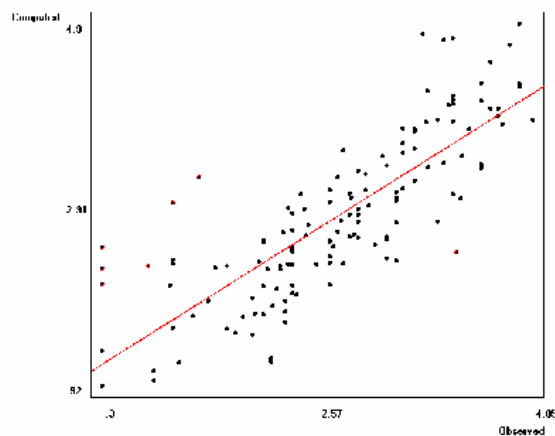


Figure 1

Scatter plot before elimination of outliers

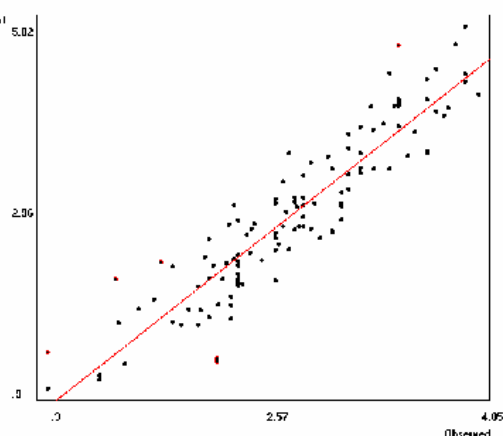


Figure 2

Scatter plot after elimination of outliers

Katritzky obtain⁷⁰, for a certain peptides group (N = 87), a sweetness QSAR with a moderate predictive power ($p = 5$, $r^2 = 0.6890$, $F = 35.7$). The predictors in Katritzky QSAR are “total entropy”, “maximum partial charge”, “number of chlorine atoms”, “maximum-n attraction in C-O bonds” and “minimum net atomic charge of N atoms”. Also is reviewed the molar volume contribution to the sweet taste.

We must underline the fact that the group of predictors in equation #1 is very similar to the group of predictors in equation #2. This suggests the fact that the outlier molecules belong to the same class as the other molecules. The fact that they are labeled as “outliers” might be owed not to a difference in the mechanism of the biochemical action, but, possibly, to an erroneous SP value used in the computations.

In the group of 121 molecules, the program identifies only 33 virtual fragments. We can say that there is not a significant variety from the structural point of view. On the other hand, there have been identified more than 10 significant structural fragments.

The virtual fragments that lead to a significant increase of the SP value are $-\text{CN}$ (cyan) and $-\text{C}_6\text{H}_4-\text{NHCONH}-$ (aryl-substituted urea). The last one is the fragment present in (for instance) molecules **114** - **127**. Due to the fact that the molecules containing a $-\text{CN}$ group have it only in the *para* position, we can not evaluate the influence of the position of this group on the aromatic nucleus.

A reduced favorable influence on the SP value is due to the presence of the $-\text{C}_6\text{H}_4-$, $-\text{NH}-$ (non-conjugated with neighboring groups) and $-\text{C}_6\text{H}_4\text{O}-$ fragments.

The virtual fragments that lead to a decrease in the values of SP are $-\text{NH}_2$, $-\text{NHCO}-$ and $-\text{COOH}$, non-conjugated with the neighboring fragments.

All molecules (with the exception of **114**) contain only one $-\text{COOH}$ group. Consequently, the unfavorable character of the “high proportion of COOH fragment” predictor must be interpreted as an unfavorable character of the “low molecular mass” predictor. The same

arguments apply to the presence of the -NHCO- fragment. The favorable character of high molecular mass is underlined also by the presence in equations #1 and #2 of highly “useful” predictor “Platt topological index”.

A reduced unfavorable influence is shown also for the fragments -Cl and >CH- (branched alkyl).

Identification of the parabolic descriptors

Equation #1 does not contain any parabolic descriptors. In QSAR study #2 only two parabolic descriptors are found on the list of 30 descriptors with the highest “signification”. In the final equation only descriptor p_2 is a parabolic function, and its calculated utility is rather reduced. We can conclude that the parabolic functions have lost the mathematical competition with the linear functions. The later are sufficient to describe the sweetness power of the molecules in Table 1, at least in the case when there is no validation/prediction set.

Utility of predictors

By the way it is defined,²²⁻²⁴ the Platt topological index is simultaneously a measure of the order and dimension of the molecular graph, that is of the size of the molecule and the degree of chain branching. For equal size (equal number of edges) the calculated value of this topological index is higher for the branched molecular graph (higher number of vertices with rank >2). Of the two graphs of equal size, the branched one is, necessarily, closer to the spherical shape. QSAR study #2 suggests that the weak unfavorable effect of chain branching can be compensated by a higher molecular mass.

The presence of predictor p_2 (in #1), respectively p_3 (in #2) suggests that the structure of virtual fragments has a certain influence over the sweetness power.

The calculated value for the free valence of N atoms is sensibly higher in the cyano groups than in the amide or amino groups. The presence of predictor p_3 (in #1), respectively p_4 (in #2) confirms the importance of -CN group in molecules with a high value of SP.

Validation of the computational method

Using the procedure described above a validation set of 24 molecules has been extracted from the homogenized calibration set. These molecules have been marked in Table 1 with bold characters. The remaining molecules (97 molecules) have formed the reduced calibration set. The reduced set and the validation set have been used in another QSAR study.

QSAR study #3

Calibration set: 97 molecules (Table 1 molecules, without outliers, without validation set)

Validation set: 24 molecules (Table 1 “bold” molecules)

Number of significant descriptors: 114

The type (1) QSAR for prediction:

$$c_0 = .7625$$

$$c_1 = .0322$$

p_1 – Zagreb topologic index

$$c_2 = -.3124$$

p_2 – Minimum net charge of N atoms - parabolic function

$$c_3 = -.7421$$

p_3 – p_x - p_y gap - parabolic function

p_x is sum of [net charge * exposed atomic surface] products (net charge > 0)

p_y is sum of [net charge * exposed atomic surface] products (net charge < 0)

$$c_4 = .3851$$

p_4 – QSPR of mass fragment percents

$$c_5 = 4.8328$$

p_5 – Maximum free valence of N atoms

Standard error s: 0.4248

Pearson square correlation r^2 : 0.8301

Fisher F function: 89.9

Kendall rank correlation K = .7728

Kendall cross-validated rank correlation K_{CV} : 0.7560

When working with a validation (or prediction) set, PRECLAV uses a specific criteria for selecting the significant descriptors¹⁸ by using the *Class* function. The program verifies if, from the point of view of the values of analyzed descriptor, the calibration set (now the reduced calibration set) is a representative sample of the whole molecular set (now the reduced calibration set + the validation set). The selection of descriptors is harsher. The resultant QSAR is different from the case when the validation/prediction set is missing, a thing that differentiates PRECLAV from other programs for QSAR computations.

In the case when there is a validation/prediction set, from the point of view of the program the characteristics of the equation (coefficients, number and type of predictors) used in prediction and the quality of the prediction for the calibration set are of less importance. What is in this case significant is the correlation between the estimated values and the experimental values of the dependent property for the molecules of the validation/prediction set. From the point of view of the considerations discussed here, the most significant is the correlation between the calculated and the experimental values of SP for the molecules in the validation set.

In Table 2 there are listed the calculated values (computed by equation #3) and the experimental values (from Table 1) of the sweetness power for the molecules in the validation set.

The correlation between the experimental and calculated values, presented in Table 2, was measured using some common statistical functions. The standard error of estimation ($s = 0.3836$), square of the Pearson correlation ($r^2 = 0.8580$) and the Kendall ranks correlation ($K =$

0.7609) have values that are comparable with the values calculated for the molecules in the calibration set

Table 2. Experimental/calculated values of SP for the molecules in the validation set

Index from Table 1	SP _{obs}	SP _{calc}	Index from Table 1	SP _{obs}	SP _{calc}
7	0.845	<u>0.989</u>	55	2.846	3.159
28	1.255	<u>1.680</u>	48	2.955	2.643
30	1.706	<u>1.592</u>	94	3.000	3.204
81	1.959	2.462	106	3.256	3.494
83	2.045	2.883	77	3.398	2.876
36	2.179	2.495	103	3.477	3.316
27	2.233	<u>1.665</u>	104	3.602	3.299
24	2.303	<u>2.364</u>	101	3.903	3.143
112	2.303	2.438	123	4.000	4.232
52	2.449	2.496	121	4.301	4.276
75	2.700	2.808	66	4.398	3.924
82	2.733	2.511	62	4.699	4.839

In the QSAR studies #2 and #3. We can state that the program has calculated for the molecules in the validation set values close to the experimental ones and has ordered the molecules in a sequence similar enough to the real one.

From the practical point of view, the labeling “high value of dependent property” and “low value of dependent property” is very important, a labeling that the program is always making¹⁸ for the molecules in the validation/prediction set. In Table 2 the calculated values identified by the program as “high” have been marked in bold letters, while the values identified as “low” have been underlined. Of the five molecules labeled as “high value”, four are the molecules that also have the highest experimental values of SP. Of the five molecules labeled as “low value” three are the molecules that have the lowest experimental SP value. This result, qualitatively correct, suggests that, in a QSAR analysis with prediction set, the program will identify with sufficient accuracy the molecules “suggested for synthesis”.

Conclusions

After eliminating the “outlier” molecules from the calibration set, using a specific criteria, PRECLAV is capable of offering information about

- the virtual molecular fragments that are significant from the point of view of the dependent property

- the molecular features useful in describing the variation of the dependent property
The virtual fragments that lead to a significant increase of the sweetness power of the derivatives of the 3-aminosuccinamic acid analyzed here are -CN (cyano) and -C₆H₄-NHCONH- (aryl-substituted urea).

The non-conjugated or weakly conjugated virtual fragment -NH₂ leads to a significant decrease of the sweetness power.

The sweetness power is favorably influenced by the size of the molecule.
The linear functions of the descriptors are enough to describe the sweetness power of the derivatives of the 3-aminosuccinamic acid, at least in the case when there is no validation/prediction set.

The tests using a validation set suggest that PRECLAV is capable to identify with sufficient accuracy the new, as yet non-synthesized, molecules with higher/lower values of the dependent property, included in a prediction set.

References

1. Mazur, R. H.; Goldkamp, A. H.; James, P. A.; Schlatter, J.M. *J. Med. Chem.* **1970**, *13*, 1217
2. Mazur, R. H.; Reuter, J. A.; Swiatek, K. A.; Schlatter J. M. *J. Med. Chem.* **1973**, *16*, 1284
3. MacRonald, S. A.; Willson, C. G.; Chorev, M.; Vernacchia, F. S.; Goodman, M. *J. Med. Chem.* **1980**, *23*, 413
4. Iwamura, H. *J. Med. Chem.* **1981**, *24*, 572
5. Mapelli, C.; Newton, M. G.; Ringold, C. E.; Stammer, C. H. *Int. J. Pept. Protein. Res.* **1987**, *30*, 498
6. Janusz, J. M.; Young, P. A.; Blum, R. B.; Riley, C. M. *J. Med. Chem.* **1990**, *33*, 1676
7. Prakash, I.; Chapeau, M.-C.R. US Patent 6,077,962, 2000.
8. Prakash, I.; Guo, Zhi US Patent 6,146,680, 2000.
9. Nofre, C.; Tinti, J.-M., EP 759,030 B1, 1998.
10. Takemoto, T.; Amino, Y.; Nakamura, R., EP 691,346, 1996.
11. Nofre, C.; Tinti, J.-M., US Patent 5,777,159, 1998.
12. Kawai, M.; Nyfeler, R.; Berman, J. M.; Goodman M. *J. Med. Chem.* **1982**, *25*, 397
13. Schumann, C.; Seyfarth, S.; Greiner, G.; Reissmann, S. *J. Pept. Res.* **2000**, *55*, 428
14. Goodman, M.; Del Vale, J.R.; Amino, Y.; Benedetti, E. *Pure Appl. Chem.* **2002**, *74*, 1109
15. Kim, Y. J; Han, S. J.; Kim, S. C.; Kang, Y. K. *Biopolymer* **1994**, *34*, 1037
16. Gilbert, K.; Gajewski, J. J. PCModel v.9.0. *Serena Software*, Box 3076, Bloomington, IN, USA
17. Stewart, J. J. P. MOPAC93. *Fujitsu Limited*, Tokyo, Japan
18. Tarko, L. *Rev. Chim. (București)*, **2005**, *56*, 639
19. PRECLAV program is available from Center of Organic Chemistry (CCO) - Bucharest, Romanian Academy; Director CCO E-mail address: pfilip@cco.ro

20. Tarko, L. *Rev. Chim. (București)* **2004**, *55*, 539
21. Tarko, L. *ARKIVOC*, **2004**, *xiv*, 74
22. Platt, J. R. *J. Chem. Phys.* **1947**, *15*, 419
23. Platt, J. R. *J. Phys. Chem.* **1952**, *56*, 328
24. Diudea, M. V.; Ivanciuc, O. *TOPOLOGIE MOLECULARA*, Ed. Complex, Cluj, 1995; pp 58.
25. Henriksen, R. B.; Rolland, M.; Breddam, K.; Buchardt, O. *Prot. Pept. Lett.* **1998**, *5*, 141
26. Yamashita, Y.; Takahashi, Y.; Takayama, C.; Sumi, K.; Nakatsuka, K.; Ohkubo, T.; Abe, H.; Sasaki, S. *J. Med. Chem.* **1986**, *29*, 906
27. Frank, M.; Aitken, D. J. *Biosci. Biotechnol. Biochem.* **2000**, *64*, 1982
28. Fang, Y. Z.; Sheng, Z. T.; Yu S. J. *Chin. Chem. Lett.* **1995**, *6*, 369
29. Searle & Co. Fr. Patent 2,013,158, 1970.
30. Toyohito, T.; Takemoto, T.; Toshihide, Y. Jp Patent 7,285,991, 1995.
31. Tarko, L.; Lupescu, I.; Groposila-Constantinescu, D. *ARKIVOC*, **2005**, (x), 254
32. Ando, T.; Ota, M.; Kashiwagi, T.; Nagashima, N.; Ariyoshi, Y.; Chadha, R. K.; Yamazaki, T.; Goodman, M. *J. Am. Chem. Soc.* **1993**, *115*, 397
33. Nakamura, R.; Amino, Y.; Takemoto, T., US Patent 5,968,581, 1999.
34. Zeng, G.; Chen, J.; He, H.; Wang, Z.; Yan, J. *J. Agric. Food Chem.* **1991**, *39*, 782
35. D'Angelo, L.; Sweeny, J. G.; King, G. A. International Patent WO 17,418, 1995.
36. Brennan, T. M.; Hendrick, M. E. US Patent 4,411,925, 1983.
37. Takemoto, T.; Amino, Y.; Nakamura, R. EP 748,814, 1996.
38. King, G. A.; Sweeny, J. G. *Org. Prep. Proced. Int.* **1997**, *29*, 177
39. Takemoto, T.; Amino, Y.; Nakamura, R. Jp Patent 9,227,589, 1997.
40. Goodman, W. M.; Zhu, Q.; Kent, D.; Amino Y.; Iacovino, R.; Benedetti, E.; Santini, A. *J. Pept. Sci.* **1997**, *3*, 231
41. Yuasa, Y.; Watanabe, T.; Nagakura, A.; Tsuruta, H.; King III, G. A.; Sweeny, J. G.; Iacobucci, G. A. *Tetrahedron* **1992**, *48*, 3473
42. Grenby, T. H. *Advances in Sweeteners*; Ed. Blackie Academic&Professional: London, 1996, pp 11.
43. Tsushia Y. Jp Patent 260,052, 1986.
44. Nofre, C.; Tinti, J.-M. Fr Patent 2,744,122, 1997.
45. Nofre, C.; Tinti, J.-M. International Patent WO 22,971, 1996.
46. Tinti, J.-M.; Nofre, C. *Contrib. Low Volatile Mater. Flavour Foods*, 1996, p 77
47. van der Heijden A.; Brussel L. B. P.; Peer, H. G. *Chem. Senses*, **1979**, *4*, 141
48. Lee, G. I.; Nakai, S.; Clarklewis, I. *Food Res. Inter.*, **1994**, *27*, 483
49. Goodman, W. M.; Mattern, R. H.; Gantzel, P.; Santini, A.; Iacovino, R.; Saviano, M.; Benedetti, E. *J. Pept. Sci.* **1998**, *4*, 229
50. Brussel, L. B.; Peer, H. G.; van der Heijden, A. *Z. Lebensm Unters Forsch.* **1975**, *159*, 337
51. Temussi, P. A.; Lelj, F.; Tancredi, T. Structure-activity relationship of sweet molecules. In *Sweeteners, Discovery, Molecular Design and Chemoreception* Walters, D. E.; Orthofer,

51. F. T.; DuBois, G. E., Eds., ACS Symposium Series 450, pp. 143-161, **1991**
52. Goodman, M.; Coddington, J.; Mierke, D.F. *J. Am. Chem. Soc.* **1987**, *109*, 4712
53. Yuasa, Y.; Nagakure, A.; Tsuruta, H. *J. Agric. Food Chem.* **2001**, *49*, 5013
54. Polanski, J.; Jarzembek, K.; Gasteiger, J. *Comb. Chem. High Throughput Screen.* **2000**, *3*, 481
55. Xing, G. V.; Tian, G. L.; Ye, Y. H. *J. Pept. Res.* **1998**, *52*, 300
56. Kawashima, K.; Itoh, H.; Yoneda N.; Hagio, K.; Moriya, T.; Chibata, I. *J. Agric. Food Chem.* **1980**, *28*, 1338
57. Kawai, M.; Chorev, M.; Marin-Rose, J.; Goodman M. *J. Med. Chem.* **1980**, *23*, 420
58. Nelson, V.; Mastalerz, P. *J. Pharm. Sci.* **1984**, *73*, 1844
59. Rodriguez, M.; Goodman M. *J. Med. Chem.* 1984, *27*, 1668
60. Tsang, J. V.; Schmied, B.; Nyfeler, R.; Goodman, M. *J. Med. Chem.* 1984, *27*, 1663
61. Rodriguez, M.; Bland, J. M.; Tsang, J. W.; Goodman M. *J. Med. Chem.* **1985**, *28*, 1527
62. Boesten, W. J. H.; Schiepers, L. A. C. US Patent 4,371,464, 1983.
63. Hooper, N. M.; Hesp, R. J.; Tiek, S.; Boesten, W. H. J.; Toniolo, C.; Kamphuis, J. *J. Agric. Food Chem.* **1994**, *42*, 1397
64. Takemoto, T.; Amino, Y.; Nakamura, R., US Patent 5,723,165, 1998.
65. Mazur, R. H. *Aspartame: Physiology and Biochemistry*, Ed. Dekker: New York, 1984, p 3
66. Mossel, E.; Formaggio, F.; Crisma, M.; Toniolo, C.; Broxterman, Q. B.; Boesten, W. H. J.; Kamphuis, J.; Quaedflieg, P. J.; Temussi, P. *Tetrahedron: Asymmetry* **1997**, *8*, 1305
67. Polinelli, S.; Broxterman, Q. B.; Schoemaker, H. E.; Boesten, W. H. J.; Crisma, M.; Valle, G.; Toniolo, C.; Kamphuis, J. *Bioorg. Med. Chem. Lett.* **1992**, *2*, 453
68. Goodman, M.; Yeong, S.O. US Patent 5,298,272, 1994.
69. Amino, Y.; Takemoto, T.; Yuzawa, K.; Nakamura, R. US Patent 65,210 A1, 2003
70. Katritzky A. R., Petrukhin R., Perumal S., Karelson M., Prakash I., Desai N. *Croat. Chem. Acta.* **2002**, *75*, 475