**ORIGINAL ARTICLE**

# A SINGLE- AND MULTIPARAMETRIC QSAR STUDY OF AMINOPEPTIDASE N INHIBITORS

OLDŘICH FARSA[1]*, MATĚJ HALUSKA[1,2]

[1]*Department of Chemical Drugs, Faculty of Pharmacy, University of Veterinary and Pharmaceutical Sciences Brno, Palackého 1/3, Brno, 612 42, Czech Republic*
[2]*State Institute for Drug Control, Regional Department OKL 406 - Brno, Stará 25, Brno, 602 00, Czech Republic*

**corresponding author: farsao@vfu.cz*

*Manuscript received: October 2015*

**Abstract**

A collection of 52 compounds with inhibitory activities, expressed as the inhibition constant $K_i$, against aminopeptidase N (APN) was assembled from diverse literature resources. The relationships between activities and physicochemical parameters derived from the structures were then searched using linear or multilinear regression analysis. Regression models of 1st and 2nd orders were used for this purpose. Three single-parametric model of dependence of log $K_i$ on the partition coefficient logarithm (log P), dipole moment (p) and molecular surface area (MSA) respectively were significant. Two bi-parametric models containing log P and p or MSA respectively reached higher significance. The only tri-parametric model containing all the three above parameters was shown to be the best. This model is useful for activities predictions of compounds proposed for synthesis as potential APN inhibitors.

**Rezumat**

O colecție de 52 de compuși cu activitate inhibitoare asupra aminopeptidazei N (APN) exprimată ca inhibiție constantă Kj, a fost asamblată din resurse diferite din literatură. Relațiile dintre activitățile și parametrii fizico-chimici, derivate din structuri, au fost apoi cercetate folosind analiza de regresie liniară sau multiliniară. În acest scop au fost folosite modelele de regresie de ordinul 1 și al doilea. Modelul parametric ce exprimă dependența log Ki de logaritmul coeficientului de partiție (log P), momentul de dipol (p) și aria suprafeței moleculare (MSA) a fost semnificativ. Două modele bi-parametrice care conțin log P, respectiv MSA a atins o semnificație mai mare. Model tri-parametric, ce conține toți cei trei parametri de mai sus, s-a dovedit a fi cel mai bun. Acest model este util pentru predicția activității unor compuși nou sintetizați ca potențialial inhibitori APN.

**Keywords**: aminopeptidase N (APN/CD13), inhibition, Quantitative Structure-Activity Relationships (QSAR), linear and multilinear regression, single- and multi-parametric models

## Introduction

Aminopeptidase N (APN), (EC 3.4.11.2), also known as CD13 antigen or membrane alanyl aminopeptidase (mAAP), belongs to the MA class of zinc metallo-peptidases, in which zinc cations are bound by two histidine residues in the sequence Hys-Glu-Xaa-Xaa-Hys, where Xaa can be any amino acid residue [15]. APN has broad substrate specificity – it cleaves amino acids from the N-terminus of mainly unsubstituted oligopeptides, or from amides. The order of amino acid residues, i.e., how easily they are hydrolysed, is Ala > Phe > Tyr > Leu > Arg > Thr > Trp > Lys > Ser > Asp > Hys. Val, Pro and Glu residues are hydrolysed very poorly. If a proline residue is preceded by a bulky hydrophobic residue, such as Leu, Tyr or Trp, an unusual reaction can occur, and the sequence Xaa-Pro is released as the intact dipeptide. The optimal pH values for activity ranges between 7.0-7.5, although the activity can be high even at a pH close to 9.0 if the concentration of a substrate is appropriate. Biologically active substrates of APN are neuropeptides (Met- and Leu-enkephalins, neurokinine A, Met-Lys-bradykinin and endorphins, such as spinorphin), vasoactive peptides (kallidin, somatostatin and angiotensin III) and chemotactic peptides (monocyte chemotactic protein 1, MCP-1) [2, 10, 15]. APN is widely distributed in many tissues of various species, although the tissues richest in APN are the brush border membranes of the kidney, and the mucosal cells of the small intestine and liver. This enzyme takes part in the extracellular catabolism of glutathione and in the terminal stages of protein and peptide digestion in the intestine. APN is also abundant in the choroid plexus of the brain and most likely protects brain tissue against access by damaging neuropeptides from the blood. It is also present in hematopoietic cells, where it is known as CD13. There, APN is primarily expressed in myeloid cells, but it also occurs on the surface of antigen-presenting cells, melanoma cells and lymphocytes. Imbalanced APN expression has been observed in many diseases.

Increased levels of APN were reported in cancer cells in melanomas, cancers of the kidney, pancreas, colon, prostate, stomach and thyroid gland. APN activity also significantly increases in patients suffering from feochromocytoma. Its activity correlates so closely with the stage and prognosis of breast, thyroid and non-small cell lung cancers that it can be used as a diagnostic and prognostic marker [13, 18]. APN is one of the key regulators of the neovascularization of cancers and other abnormal tissues. Pasqualini *et al.* [14] demonstrated that APN is expressed specifically in endothelial and subendothelial cells of cancer tissues undergoing angiogenesis. There, it serves as a specific receptor for tumour-homing NGR peptides, i.e., specific tumour-binding peptides containing a Asp-Gly-Arg motif. They also demonstrated that both an anti-APN monoclonal antibody and APN inhibitors, such as bestatin (Compound **6**, Table 1) and actinonin (Compound 1, Figure 1), significantly suppressed vessel growth, which had been triggered by the addition of basic fibroblast growth factor (bFGF) to the chicken chorioallantoic membrane (CAM). This implies that APN inhibitors could be potent angiogenesis suppressors. Because angiogenesis is a key process in cancer growth and metastasis, APN inhibitors may be considered as potential anticancer agents. Bestatin, being produced by actinomycetes, was found to inhibit the invasion and activity of mouse and human metastatic cells [21]. This effect was subsequently explained to be due to APN inhibition. Bestatin also underwent clinical trials for the treatment of several types of cancer and was used for the treatment of lung cancer and acute myeloid leukaemia in Japan for several years [7, 17]. Recently, tosedostat, CHR-2797 (Compound 2, Figure 1), another aminopeptidase inhibitor, was intensely researched. Although this compound is also a potent APN inhibitor, its anticancer activity is attributed to inhibition of other aminopeptidases located inside the cell. Tosedostat is a cyclopentyl ester prodrug, and the free carboxylic acid form CHR-79888 (Compound 3, Figure 1), arising from its de-esterification, is responsible for its activity. This reaction is, however, performed by intra-cellular esterases. The free acid form can then have difficulty moving back outside the cell to interact with APN, which is localized on the outer side of the cell membrane [9].

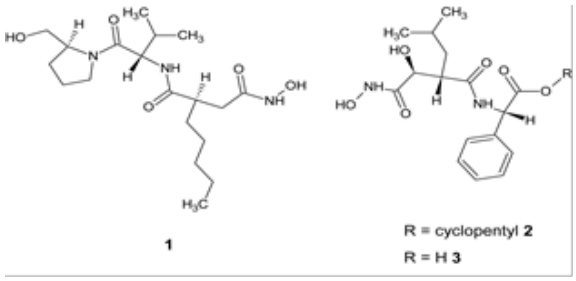



**Figure 1.**
Structures of inhibitors of aminopeptidases with anticancer activity: actinonin 1, tosedostat 2 and CHR-79888 3

APN also serves also as a cellular receptor enabling the entrance of some coronaviruses into a cell. Coronaviruses are encapsulated human and animal RNA viruses, which cause mainly enteric and respiratory diseases, such as Severe Acute Respiratory Syndrome (SARS) and approximately 10-20% of common colds. In these viruses, the envelope spike glycoprotein (S) mediates the attachment of the virus particles to APN and subsequent cell entry, which can be blocked by neutralizing anti-APN antibodies [16]. The use of APN as an entrance receptor was confirmed for transmissible gastro-enteritis virus, which causes fatal diarrhoea in new-born pigs, and for porcine respiratory coronavirus [3, 16]. APN also appears to mediate cytomegalovirus (CMV) infection [19]. APN/CD13 has been detected, using antibodies, in viral envelopes [6]. It was shown that infection by both human (HCMV) and murine (MCMV) cytomegaloviruses resulted in formation of anti-APN antibodies [8]. The afore-mentioned results imply that potent APN inhibitors could be effective therapeutics for cancers and viral infections. This is why we attempted to create mathematical models enabling the estimation of the proposed compounds' inhibitory activities against APN.

## Materials and Methods

*Software and computation equipment*

HyperChem Professional Version 6.03, Hypercube, Inc., was used for calculation of hydration energy ($DH_H$) values. MarvinSketch 6.2.0, ChemAxon, Ltd. was used for calculation of molecular surface area (MSA), polar surface area (PSA), isoelectric point (pI) and dipole moment values. ChemSketch 12.01, Advance Chemistry Development, Inc. was used for calculation of partition coefficient logarithm (log P) values for the octanol/water system, molar refractivity (MR), molar volume (MV) and parachor (Pr) values. QC. Expert 2.5,

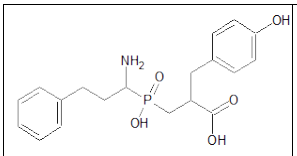
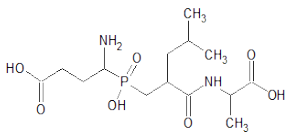
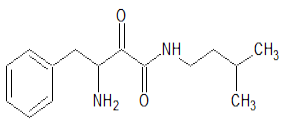
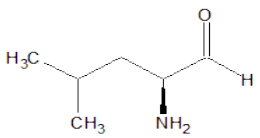
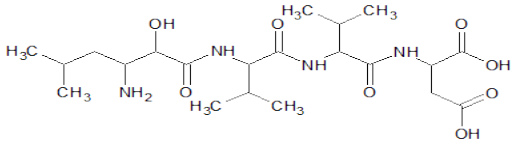
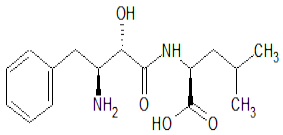
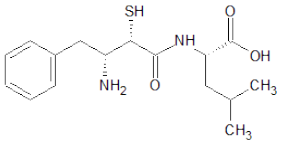
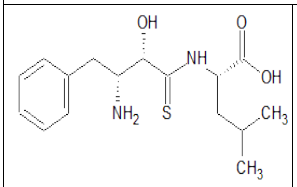
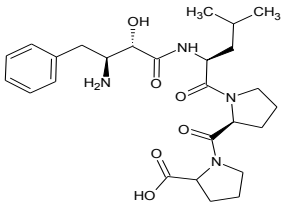
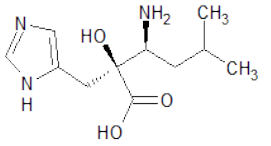
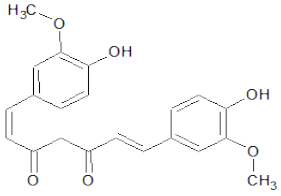
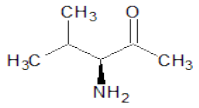
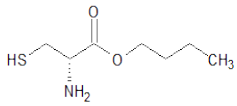
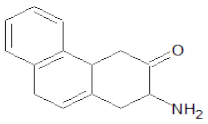
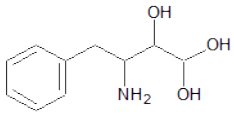
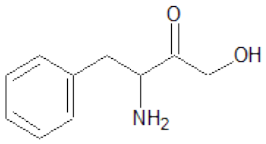
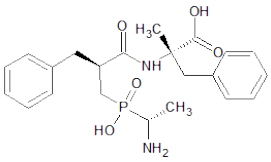
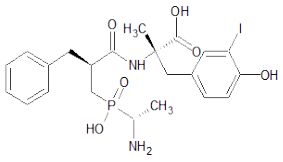
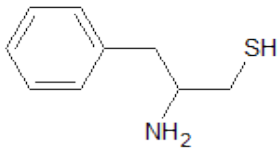
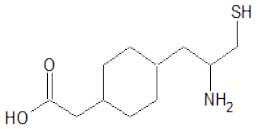
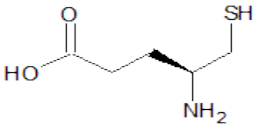
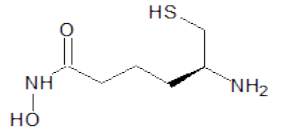
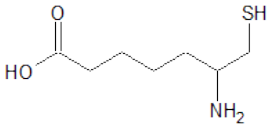
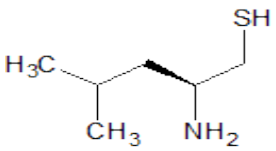
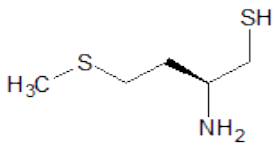
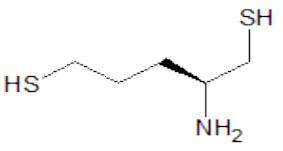
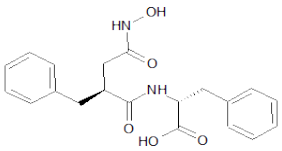
Trilobyte s.r.o., was the statistical software used for regression analyses. All computations were performed on an Acer Extensa 5220 laptop equipped with an Intel T1400 1.73 GHz CPU and 1.00 GB of RAM.

*Set of tested compounds*

Both the structures and activities, expressed as inhibition constants ($K_i$), of the APN inhibitors were taken from original articles published in literature [1, 11, 12, 20]. Each compound was given a number. If more than one $K_i$ value for a compound had been found in the literature, their mean value was used. Because $K_i$ values had been expressed in different units in various articles, they were first recalculated to μmol*$L^{-1}$ units. $K_i$ values were also transformed to log $K_i$ values, which were found to be more useful in linear correlations than the $K_i$ values themselves. The structures of the tested compounds are presented in Table I, while their corresponding calculated structural parameters are depicted in Table II.

**Table I**

Structures of tested compounds with their assigned numbers

| | | | |
|---|---|---|---|
|  **1** |  **2** |  **3** |  **4** |
|  **5** |  **6** |  **7** | |
|  **8** |  **9** |  **10** |  **11** |
|  **12** |  **13** |  **14** |  **15** |
|  **16** |  **17** |  **18** |  **19** |
|  **20** |  **21** |  **22** |  **23** |
|  **24** |  **25** |  **26** |  **27** |

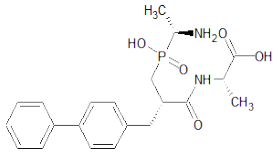
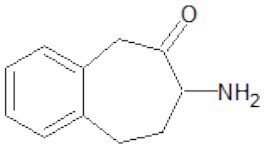
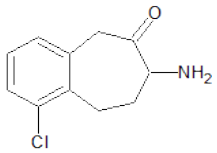
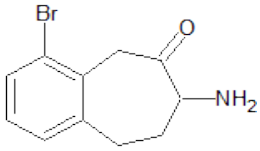
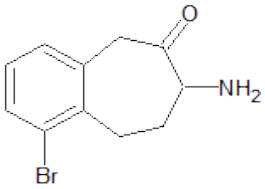
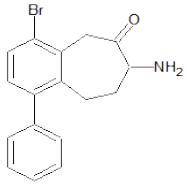
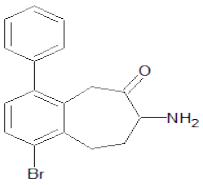
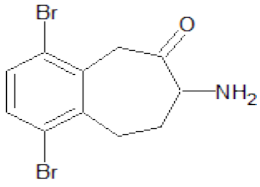
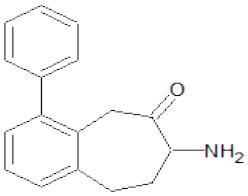
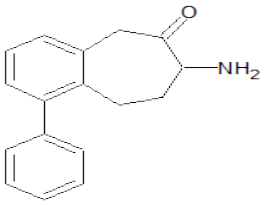
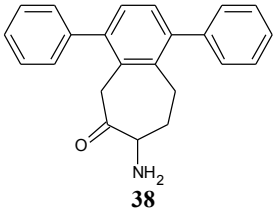
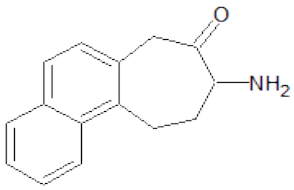
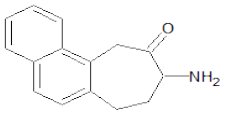
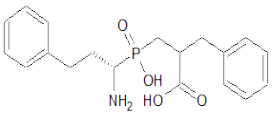
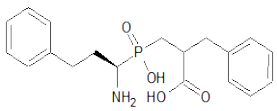
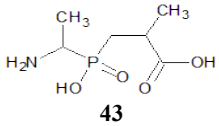
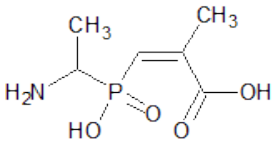
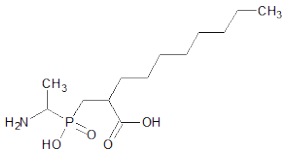
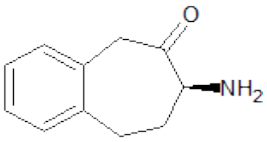
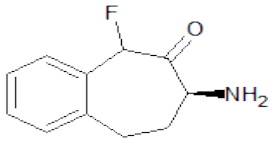
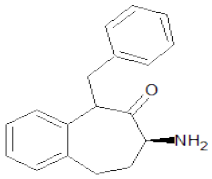
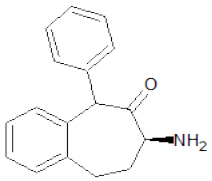
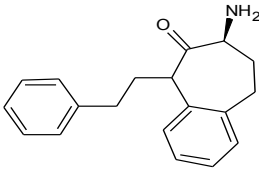
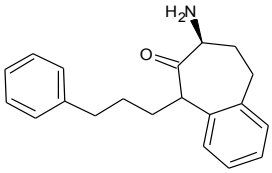
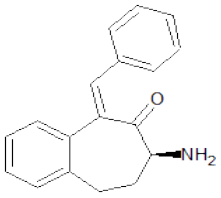
| | | | |
|---|---|---|---|
|  **28** |  **29** |  **30** |  **31** |
|  **32** |  **33** |  **34** |  **35** |
|  **36** |  **37** |  **38** |  **39** |
|  **40** |  **41** |  **42** |  **43** |
|  **44** |  **45** |  **46** |  **47** |
|  **48** |  **49** |  **50** |  **51** |
|  **52** | | | |

**Table II**

Set of tested compounds with their activities ($K_i$) and all calculated structural parameters

| Compound No. | $K_i$ ($\mu mol*L^{-1}$) | log $K_i$ | log P | MR ($cm^3$) | MV ($cm^3$) | Pr ($cm^3$) | $\Delta H_H$ ($kJ*mol^{-1}$) | pI | p. D | PSA ($10^{-2}$ $nm^2$) | MSA ($10^{-2}$ $nm^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | 0.03600 | -1.444 | 2.84 | 98.76 | 284.6 | 804.3 | -28.291 | 2.44 | 3.57 | 120.85 | 544.00 |
| 2. | 31.00000 | 1.491 | 0.33 | 85.44 | 281.6 | 773.1 | -33.564 | 2.12 | 6.66 | 167.02 | 557.20 |
| 3. | 2.50000 | 0.398 | 2.29 | 75.33 | 245.2 | 624.2 | -6.194 | 10.04 | 2.63 | 72.19 | 435.03 |
| 4. | 0.76000 | -0.119 | 0.86 | 33.48 | 129.3 | 303.5 | 42.227 | 13.84 | 2.51 | 43.09 | 221.63 |
| 5. | 0.05100 | -1.292 | 2.40 | 118.08 | 386.4 | 1035.1 | -29.337 | 4.45 | 2.49 | 208.15 | 749.03 |
| 6. | 3.26500 | 0.514 | 2.64 | 82.95 | 257.5 | 692.2 | -21.511 | 6.18 | 5.05 | 112.65 | 481.45 |

| Compound No. | $K_i$ ($\mu mol*L^{-1}$) | log $K_i$ | log P | MR ($cm^3$) | MV ($cm^3$) | Pr ($cm^3$) | $\Delta H_H$ ($kJ*mol^{-1}$) | pI | p. D | PSA ($10^{-2}$ $nm^2$) | MSA ($10^{-2}$ $nm^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 7. | 4.40000 | 0.643 | 3.31 | 89.32 | 274.0 | 727.3 | -21.218 | 6.19 | 6.21 | 131.22 | 491.44 |
| 8. | 40.30000 | 1.605 | 3.75 | 90.21 | 265.4 | 733.7 | -23.352 | 6.33 | 6.45 | 127.67 | 486.93 |
| 9. | 0.01900 | -1.721 | 2.41 | 131.98 | 391.7 | 1099.2 | -25.026 | 6.05 | 7.80 | 153.27 | 768.80 |
| 10. | 0.23000 | -0.638 | 0.54 | 62.64 | 189.4 | 533.7 | -24.775 | 8.30 | 2.85 | 112.23 | 378.29 |
| 11. | 11.20000 | 1.049 | 2.92 | 104.04 | 287.8 | 781.5 | -27.286 | 2.17 | 12.00 | 93.06 | 507.87 |
| 12. | 0.55000 | -0.260 | 0.33 | 33.32 | 129.1 | 299.7 | -25.277 | 13.15 | 3.63 | 43.09 | 221.97 |
| 13. | 0.18000 | -0.745 | 2.00 | 47.64 | 165.6 | 413.6 | -54.698 | 8.51 | 3.21 | 91.12 | 290.41 |
| 14. | 0.50000 | -0.301 | 1.71 | 63.08 | 178.6 | 478.1 | -41.641 | 12.90 | 3.73 | 43.09 | 306.06 |
| 15. | 3.00000 | 0.477 | -0.12 | 53.08 | 152.5 | 434.8 | -83.574 | 10.03 | 3.02 | 86.71 | 295.60 |
| 16. | 1.00000 | 0.000 | 0.95 | 50.12 | 153.9 | 411.6 | -30.592 | 10.56 | 3.62 | 63.32 | 278.38 |
| 17. | 0.00060 | -3.222 | 2.85 | 114.36 | 341.6 | 936.8 | -2.511 | 2.32 | 1.04 | 139.53 | 642.41 |
| 18. | 0.00095 | -3.022 | 3.29 | 129.15 | 362.2 | 1027.4 | -10.379 | 1.94 | 3.49 | 159.76 | 679.29 |
| 19. | 0.00500 | -2.301 | 2.11 | 51.82 | 156.8 | 401.9 | -19.502 | 9.70 | 0.79 | 64.82 | 259.32 |
| 20. | 0.03500 | -1.456 | 1.89 | 63.91 | 210.8 | 548.3 | -27.663 | 6.98 | 2.04 | 102.12 | 372.72 |
| 21. | 0.12000 | -0.921 | 0.13 | 38.16 | 123.7 | 331.1 | -33.647 | 6.65 | 2.11 | 102.12 | 221.73 |
| 22. | 0.03700 | -1.432 | -1.14 | 46.39 | 149.1 | 398.8 | -69.136 | 9.30 | 4.49 | 114.15 | 270.91 |
| 23. | 0.28000 | -0.553 | 0.40 | 47.43 | 156.8 | 410.7 | -33.647 | 6.95 | 0.53 | 102.12 | 282.65 |
| 24. | 0.02200 | -1.658 | 1.79 | 41.19 | 146.0 | 346.5 | -49.885 | 10.04 | 0.32 | 64.82 | 245.71 |
| 25. | 0.00800 | -2.097 | 1.19 | 44.60 | 143.1 | 360.0 | -59.887 | 10.03 | 0.58 | 90.12 | 238.66 |
| 26. | 0.03200 | -1.495 | 1.13 | 44.49 | 143.2 | 361.2 | -63.110 | 9.71 | 2.66 | 103.62 | 233.85 |
| 27. | 0.12000 | -0.921 | 1.39 | 98.35 | 286.7 | 794.0 | -17.661 | 1.01 | 3.32 | 115.73 | 530.75 |
| 28. | 0.00290 | -2.538 | 2.42 | 109.83 | 330.1 | 900.5 | -28.500 | 2.33 | 6.17 | 139.53 | 612.52 |
| 29. | 1.00000 | 0.000 | 0.91 | 51.33 | 158.0 | 405.5 | -25.989 | 11.33 | 3.80 | 43.09 | 270.67 |
| 30. | 0.09000 | -1.046 | 1.51 | 56.22 | 170.0 | 442.6 | -26.240 | 11.02 | 3.23 | 43.09 | 285.77 |
| 31. | 0.04000 | -1.398 | 1.69 | 59.02 | 174.2 | 456.5 | -25.989 | 10.88 | 4.57 | 43.09 | 289.39 |
| 32. | 0.02000 | -1.699 | 1.69 | 59.02 | 174.2 | 456.5 | -26.156 | 11.08 | 3.30 | 43.09 | 289.47 |
| 33. | 0.07000 | -1.155 | 3.58 | 83.61 | 239.5 | 628.6 | -25.780 | 10.75 | 4.67 | 43.09 | 395.57 |
| 34. | 0.00006 | -4.222 | 3.58 | 83.61 | 239.5 | 628.6 | -25.026 | 10.86 | 3.32 | 43.09 | 395.36 |
| 35. | 0.00600 | -2.222 | 2.49 | 66.71 | 190.4 | 507.6 | -26.114 | 10.63 | 3.87 | 43.09 | 307.82 |
| 36. | 0.00700 | -2.155 | 2.67 | 75.92 | 223.3 | 577.5 | -24.901 | 11.07 | 3.80 | 43.09 | 376.73 |
| 37. | 0.25000 | -0.602 | 2.67 | 75.92 | 223.3 | 577.5 | -25.821 | 11.20 | 3.81 | 43.09 | 376.77 |
| 38. | 0.04000 | -1.398 | 4.33 | 100.51 | 288.6 | 749.6 | -24.650 | 10.97 | 3.77 | 43.09 | 483.51 |
| 39. | 0.04000 | -1.398 | 2.15 | 69.17 | 192.1 | 510.4 | -26.073 | 11.16 | 3.81 | 43.09 | 332.40 |
| 40. | 0.10000 | -1.000 | 2.15 | 69.17 | 192.1 | 510.4 | -26.031 | 11.07 | 3.37 | 43.09 | 299.56 |
| 41. | 0.15800 | -0.801 | 3.57 | 96.87 | 286.1 | 789.3 | -16.782 | 2.65 | 6.67 | 110.43 | 533.16 |
| 42. | 4.894 | 0.690 | 2.15 | 69.17 | 192.1 | 510.4 | -26.031 | 11.07 | 3.37 | 43.09 | 299.56 |
| 43. | 10.00000 | 1.000 | -0.32 | 43.27 | 148.2 | 405.2 | -42.394 | 2.76 | 4.18 | 110.43 | 291.58 |
| 44. | 1.00000 | 0.000 | -1.12 | 43.21 | 141.3 | 391.3 | -42.394 | 3.67 | 6.29 | 110.43 | 263.72 |
| 45. | 1.10000 | 0.041 | 3.40 | 75.70 | 263.8 | 683.7 | -43.984 | 2.79 | 5.38 | 110.43 | 505.36 |
| 46. | 1.00000 | 0.000 | 0.91 | 51.33 | 158.0 | 405.5 | -25.989 | 11.33 | 3.80 | 43.09 | 270.67 |
| 47. | 0.30000 | -0.523 | 0.92 | 51.66 | 160.3 | 412.6 | -24.524 | 11.44 | 2.48 | 43.09 | 272.94 |
| 48. | 0.08300 | -1.081 | 3.05 | 80.50 | 236.3 | 617.4 | -14.313 | 11.59 | 3.64 | 43.09 | 404.82 |
| 49. | 0.06000 | -1.222 | 2.37 | 75.77 | 222.2 | 577.3 | -21.051 | 11.93 | 3.72 | 43.09 | 374.04 |
| 50. | 55.0000 | 1.740 | 3.45 | 85.14 | 256.3 | 657.4 | -20.758 | 11.58 | 3.52 | 43.09 | 435.56 |
| 51. | 12.00000 | 1.079 | 3.93 | 89.78 | 272.8 | 697.5 | -19.544 | 11.74 | 3.55 | 43.09 | 466.21 |
| 52. | 80.00000 | 1.903 | 3.77 | 81.97 | 225.8 | 603.3 | -14.313 | 12.68 | 6.22 | 43.09 | 378.24 |

Legend: inhibition constants ($K_i$); partition coefficient logarithm (log P); molar refractivity (MR); molar volume (MV); parachor (Pr); isoelectric point (pI); polar surface area (PSA); molecular surface area (MSA)

*Linear regression procedure*

The first regression analysis, using the simple linear model, was performed with a complete set of $K_i$ values for each structural parameter. The first regression model was thus acquired, usually with poor values for the linear correlation coefficient R, Fisher-Snedecor test of model significance F and mean error of prediction (MEP). In accordance with our previous experiences [4], a test for influential points using Atkinson distance, Andrews-Pregibon statistics, and the $Y^2$ parameter was performed simultaneously. The point indicated by these tests as the most influential one was then excluded as an outlier and the regression analysis, accompanied by an influential points search, was repeated. If the statistical parameters improved, this step was repeated until the best regression model was obtained. Then, the same procedure was performed with a quadratic model. Subsequently, both stages mentioned above were repeated for the set of log $K_i$

values, which gave mostly better results. All of these procedures enabled us to select structural parameters having a significant impact on the $K_i$ values. Such parameters were then combined into bi- or tri-parametric models, which were developed as the previous ones, except that linear regression was replaced with multilinear regression.

**Results and Discussion**

*Single parametric models*

Linear regression by both the first and the second order regression models did not result in a significant regression model of the dependence of $K_i$ on log P, $\Delta H_H$, p, MR, MV, Pr, pI, PSA or MSA. The transformation of $K_i$ into log $K_i$ enabled to find significant regression models for log P, p and MSA. None of other parameters correlated with log $K_i$ with the linear correlation coefficient ≥ 0.5.

*Lipophilicity expressed as log P*

The correlation between $K_i$ and log P was not found for the first order model, but for the second order one it was successfully found. 7 points had to be excluded as outliers to reach the following model:

$$\log K_i = 0.507(\log P)^2 - 1.774\log P + 0.333 \quad (1)$$

n = 45, R = 0.621, F = 13.179, MEP = 0.868.

*Dipole moment*

The exclusion of only 3 compounds as outliers enabled to get a simple first order regression model with following parameters:

$$\log K_i = 0.325p - 1.811 \quad (2)$$

n = 49, R = 0.523, F = 17.736, MEP = 1.147.

*Molecular surface area*

The correlation between log $K_i$ and MSA was found in a second order after exclusion of 4 compounds as outliers:

$$\log K_i = -4.076*10^{-5}(MSA)^2 + 0.033MSA - 6.724 \quad (3)$$

n = 48, R = 0.502, F = 7.594, MEP = 1.197.

Regression models for $\Delta H_H$, MR, MV, Pr, pI and PSA were insignificant in the first and in the second order. These are the best regression models reached for them:

$$\log K_i = -6.104*10^{-4}(MR)^2 - 0.089MR - 3.537 \quad (4)$$

n = 51, R = 0.370 F = 3.803, MEP = 1.426.

This model evidently tends to be constant.

$$\log K_i = -1.288*10^{-4}(MV)^2 - 0.058MV - 6.545 \quad (5)$$

n = 49, R = 0.399, F = 4.365, MEP = 1.420.

$$\log K_i = -8.675*10^{-6}(Pr^2) - 0.011Pr - 3.594 \quad (6)$$

n = 52, R = 0.322, F = 2.838, MEP = 1.687.

This is also a virtually constant model.

$$\log K_i = 2.332*10^{-3}(DH_H)^2 - 0.170DH_H - 3.348 \quad (7)$$

n = 46, R = 0.431, F = 4.798, MEP = 1.243.

$$\log K_i = 7.982*10^{-2}pI - 1.537 \quad (8)$$

n = 49, R = 0.222, F = 2.445, MEP = 1.600.

*Multi-parametric models*

Multilinear regression analyses were performed in accordance with experiences with previous single parametric models. The log $K_i$ values sets were only correlated with the corresponding sets of two or three structural parameters. Hydrophobicity is generally considered to be the most important structural parameter influencing biological activity of a compound. That is why our bi-parametric models were constructed from log P and one additional structural parameter.

*Partition coefficient (log P) and dipole moment (p)*

Both structural parameters were taken into the multilinear regression analysis squared. After stepwise exclusion of 13 outliers, the following equation resulted:

$$\log K_i = 4.362*10^{-2}(\log P)^2 - 0.242p^2 + 0.211(\log P)p - 0.749\log P + 0.956p - 1.189 \quad (9)$$

n = 39, R = 0.875, F = 21.647, MEP = 0.403.

The regression coefficients of all of the variables except of those of $(\log P)^2$ and (logP)p are significant. The value of the multilinear correlation coefficient R is greater than the corresponding values of both partial single parametric models.

Partition coefficient (log P) and molecular surface area (MSA)

As above, both structural parameters were taken into the multilinear regression analysis squared. After stepwise exclusion of 11 outliers, the following regression model was acquired:

$$\log K_i = -0.134(\log P)^2 - 1.325(MSA)^2 + 1.747(\log P)(MSA) - 0.254\log P + 0.760MSA - 0.520 \quad (10)$$

n = 41, R = 0.828, F = 15.228, MEP = 0.602.

The regression coefficients of all variables except of logP and $(\log P)^2$ are significant. Thus, this implies the majority dependence of log $K_i$ on MSA with only smaller participation of log P which is surprising.

Tri-parametric model: partition coefficient (logP), dipole moment (p) and molecular surface area (MSA)

The partition coefficient can represent hydrophobic parameters, dipole moment electronic parameters and molecular surface area steric parameters, as they are presented in classical QSAR resources as well as in newer Medicinal Chemistry textbooks

[5]. All of the three structural parameters underwent a multilinear regression squared. After outliers had been excluded in a stepwise fashion, the following final regression model was acquired:

$$\log K_i = -0.484(\log P)^2 + 1.704*10^{-2}(p)^2 - 1.259(MSA)^2 - 0.185(\log P)(p) + 1.865(\log P)MSA + 0.316(p)MSA - 0.675\log P + 0.316p + 0.814MSA - 0.369 \quad (11)$$

n = 44, R = 0.721, F = 4.206, MEP = 1.392.
Regression coefficients for $(MSA)^2$, (logP)(MSA) and logP are significant. Despite the insignificance of some equation members, this equation can be useful for the estimation of activities of newly proposed compounds.

*Cross-validation of successful regression models*
The predictivity of the final models was checked by division of sets of data remaining after exclusion of outliers into sets of even-ordered and odd-ordered data for each regression model. Values of pair (Pearson) correlation coefficient and determination coefficient (RSQ) were then enumerated for both data subsets of each model. The results are summarized in Table III.

**Table I**

Correlation (R) and determination (RSQ) coefficients for even and odd data subsets of final single-, bi and tri-parametric regression models

| Model No. | Model description | $R_{odds}$ | $RSQ_{odds}$ | $R_{evens}$ | $RSQ_{evens}$ |
|---|---|---|---|---|---|
| 1 | $\log K_i = f[(\log P)^2]$ | 0.615 | 0.369 | 0.602 | 0.363 |
| 2 | $\log K_i = f(p)$ | 0.403 | 0.162 | 0.609 | 0.371 |
| 3 | $\log K_i = f[(MSA)^2]$ | 0.483 | 0.233 | 0.531 | 0.282 |
| 9 | $\log K_i = f[(\log P)^2, p^2]$ | 0.679 | 0.461 | 0.511 | 0.261 |
| 10 | $\log K_i = f[(\log P)^2, (MSA)^2]$ | 0.830 | 0.689 | 0.774 | 0.599 |
| 11 | $\log K_i = f[(\log P)^2, p^2, (MSA)^2]$ | 0.625 | 0.390 | 0.807 | 0.650 |

Although there are some differences between the coefficients for even and odd data subset of the same model, the coefficient values are satisfactory to demonstrate a good predictivity of all the above models.

## Conclusions

The aim of this study was to develop simple correlation models which would enable the rapid estimation of inhibition activity of a proposed compound against APN from only from the calculated structural parameters. Neither a linear, nor quadratic dependence of the inhibition constant $K_i$ on any of the tested parameters (log P, MR, MV, Pr, $\Delta H_H$, pI, p, PSA, MSA) was found. The transformation of $K_i$ into log $K_i$ values, however, enabled us to find a significant relationship equation for logP, p and MSA. The first order model (2) with satisfactory statistical significance was only obtained for the dipole moment (p), while the second order models (1) and (3) with satisfactory statistical significance were found for partition coefficient in the octanol /water system (log P) and molecular surface area (MSA) respectively. All other structure parameters did not exhibit linear correlation with $\log K_i$. The results for logP, p and MSA suggested that the dependence of the inhibition activity on structural features, which are represented by these parameters, was probable. That is why multi-parametric models, expressing dependence on more than one parameter, were then constructed. The bi-parametric models (9) and (10) where either p or MSA was added to log P showed better significance than any of original single-parametric models. The combination of all three parameters into one second-order bi-parametric model then led to the tri-parametric model (11) with the best statistical characteristic of them all. Here, partition coefficient logarithm, dipole moment and molecular surface area respectively play the part of a hydrophobic, an electronic and a sterical parameter. The predictivity of regression models believed successful was then tested by cross-validation. The remaining data sets after exclusion of outliers were divided each into a subset of even-ordered data and a subset of odd-ordered ones. Both subsets of each model were then characterized by pair or Pearson correlation coefficient and determination coefficient (RSQ). The values of these coefficients were satisfactory. In conclusion, mainly the latter equation (11) can be recommended for decision-making, if a proposed compound is interesting and will be synthesized as a potential APN inhibitor. The $K_i$ values, calculated in accordance with this model, can simply be compared with the experimental $K_i$ values for highly active or less active inhibitors from the literature.

## References

1. Albrecht S., Salomon E., Defoin A., Tarnus C., Rapid and efficient synthesis of a novel series of substituted aminobenzosuberone derivatives as potent, selective, non-peptidic neutral aminopeptidase inhibitors. *Bioorg. Med. Chem.*, 2012; 20(16): 4942-4953.

2. Bavois B., Dauzonne D., Aminopeptidase-N/CD13 (EC 3.4.11.2) inhibitors: chemistry, biological evaluations, and therapeutic prospects. *Med. Res. Rev.*, 2006; 26(1): 88-130.
3. Delmas B., Gelfi J., L'Haridon R., Sjostromnoren H., Laude H., Aminopeptidase N is a major receptor for the enteropathogenic coronavirus TGEV. *Nature*, 1992; 357(6377): 417-420.
4. Florea A., Cristea C., Săndulescu R., MUC1 marker for the detection of ovarian cancer. A review. *Farmacia*, 2014; 62(1): 1-13.
5. Gareth T., Medicinal Chemistry. An Introduction. John Wiley and Sons, Chichester 2007: 94-105.
6. Giugni T.D., Soderberg C., Ham D.J., Bautista R.M., Hedlund K.O., Moller E., Zaia J.A., Neutralization of human cytomegalovirus by human cd13-specific antibodies. *J. Infect. Dis.*, 1996; 173(5): 1062-1071.
7. Ichinose Y., Genka K., Koike T., Kato H., Watanabe Y., Mori T., Iioka S., Sakuma A., Ohta M., Randomized double-blind placebo-controlled trial of bestatin in patients with resected stage I squamous-cell lung carcinoma. *J. Natl. Cancer. Inst.*, 2003; 95(8): 605-610.
8. Kasman L.M., CD13/aminopeptidase N and murine cytomegalovirus infection. *Virology*, 2005; 334(1): 1-9.
9. Krige D., Needham L.A., Bawden L.J., Flores N., Farmer H., Miles L.E.C., Stone E., Callaghan J., Chandler S., Clark V.L., Kirwin-Jones P., Legris V., Owen J., Patel T., Wood S., Box G., Laber D., Odedra R., Wright A., Wood M.L., Eccles S.A., Bone E.A., Ayscough A., Drummond AH., CHR-2797: an antiproliferative aminopeptidase inhibitor that leads to amino acid deprivation in human leukemic cells. *Cancer Res.*, 2008; 68(16): 6669-6679.
10. Langner J., Ansorge S., Peptidases in immune functions and diseases 2., Kluwer Academic/Plenum, New York 2000: 1-57.
11. Maiereanu C., Schmitt C., Schifano-Faux N., Le Nouen D., Defoin A., Tarnus C., A novel amino-benzosuberone derivative is a picomolar inhibitor of mammalian aminopeptidase N/CD13. *Bioorg. Med. Chem.*, 2011; 19(18): 5716-5733.
12. Mucha A., Lämmerhofer M., Lindner W., Pawełczak M., Kafarski P., Individual stereoisomers of phosphinic dipeptide inhibitor of leucine aminopeptidase. *Bioorg. Med. Chem. Lett.*, 2008; 18(5): 1550-1554.
13. Murakami H., Yokoyama A., Kondo K., Nakanishi S., Kohno N., Miyake M., Circulating aminopeptidase N/CD13 is an independent prognostic factor in patients with non–small cell lung cancer. *Clin. Cancer. Res.*, 2005; 11(24): 8674-8679.
14. Pasqualini R., Koivunen E., Kain R., Lahdenranta J., Sakamoto M., Stryhn A., Ashmun R.A., Shapiro L.H., Arap W., Ruoslahti E., Aminopeptidase N is a receptor for tumor-homing peptides and a target for inhibiting angiogenesis. *Cancer Res.*, 2000; 60(3): 722-727.
15. Rawlings N.D., Salvesen G. Handbook of Proteolytic Enzymes. 2nd edn., Elsevier Science, Amsterdam 2004: 289-292.
16. Reguera J., Santiago C., Mudgal G., Ordoño D., Enjuanes L., Casasnovas J.M., Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS Pathog.*, 2012; 8(8): e1002859.
17. Scornik O.A., Botbol V., Bestatin as an experimental tool in mammals. *Curr. Drug. Metab.*, 2001; 2(1): 67-85.
18. Severini G., Gentilini L., Tirelli C., Diagnostic evaluation of alanine aminopeptidase as serum marker for detecting cancer. *Cancer Biochem. Biophys.*, 1991; 12(3): 199-204.
19. Stănciulescu E.C., Pisoschi C.G., Rău G., Andrei A.-M., Baniță M., Proteasome inhibition: recent advances in antitumoral therapy. *Farmacia*, 2014; 62(2): 245-253.
20. Yang K., Golich F.C., Sigdel T.K., Crowder M.W., Phosphinate, sulfonate, and sulfonamidate dipeptides as potential inhibitors of *Escherichia coli* aminopeptidase N. *Bioorg. Med. Chem. Lett.*, 2005; 15(23): 5150-5153.
21. Yoneda J., Saiki I., Fujii H., Abe F., Kojima Y., Azuma I., Inhibition of tumor invasion and extracellular matrix degradation by ubenimex (bestatin). *Clin. Exp. Metastasis*, 1992; 10(1): 49-59.