MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **83** (2020) 261-284

ISSN 0340 - 6253

A Metric Space Approach on the Molecular vs. Chemical Similarity of Some Analgesic and Euphoric Compounds

Demetrios Xenides[1], Dionisia Fostiropoulou, Dimitrios Vlachos

Laboratory of Mathematical Modeling and Complex Systems, Department of Computer Science and Technology, Faculty of Sciences and Technology, University of Peloponnese, GR-22100 Tripolis, Greece

(Received September 16, 2019)

Abstract

When discussing similarity of compounds two different types of it should be mentioned: *Molecular Similarity* and *Chemical Similarity*. The former is based solely on structural features while the latter on the physicochemical properties of compounds. The present study has been set towards a fresh approach on the calculation of the *Chemical Similarity*. For this reason a set of molecules ranging from opium alcaloids and their synthetic analogues, to sugar, nicotine and caffeine, has been formed as to study their connections. The study emerged from the available clinical studies, in small samples though, that revealed that analgesic drugs and short term pleasure factors exhibit the same psychokinetic or psychomimetic effect. To perform a rigorous classification we have implemented two independent approaches with solid mathematical grounds: thus we relied on the information theory and string matching tools (*Tanimoto* coefficient (T_c) and *Tversky* index (T_v)). The former accounts for chemical similarity (cs) whereas the latter for molecular similarity (ms). Both indices led to the formation of clusters of compounds similar as: A_{ms}=(*morphine*, *heroine*, *codeine*), of B_{ms}=(*fentanyl*, *carfentanil*, *furanylfentanyl*) and of C_{ms}=(*endomorphine-1*, *endomorphine-2*), in great agreement to the T_v's. The information theory approach not only predicted these clusters but also brought forth further ones such as A_{cs}=(*caffeine*, *serotonin*, *adrenaline*), B_{cs}=(*endomorphine-1*, *endomorphine-2*), and C_{cs}=(*THC*, *fentanyl*, *cocaine*, *ecstasy*, *carfentanil*, *furanylfentanyl*, *codeine*, *heroine*, *morphine*, *LSD*, *methadone*) in agreement to clinical observations. The C_{cs} cluster that describes the strong connections of the active substance of marijuana (namely THC) to cocaine, morphine, methadone and fentanyl is an evidence that offers another break on the wall between "soft" and "hard" drugs. We also note that sucrose seemed to be quite dissimilar to all other compounds.

[1]Author to whom all correspondece should be addressed, E-mail: xenides@uop.gr

1 Introduction

It is well known that the question of similarity in nature rarely can be answered in one and only one way [1]. On the other hand, there is a relentless effort in science on the finding of the best alternative as to replace one object with another (or a process with another), thus special effort has been put into making the endeavour both accurate and effective and at the same time as source reserving as it can be. To this extend, tools as to, in a first step, screening a set of candidates, and after some cycles, finding the one that best matches a certain pattern or shows similar behaviour led to the formation of the *Quantitave or Qualitative Structure Activity Relationship* field, that covers chemistry, biology and engineering [2].

There is an ongoing discusion on the possible resemblance of the procedures that accelerated in the brain when someone consumes compounds that are on the top of the nutrition list of humans, e.g. sugar and caffeine as well as nicotine, and those known as narcotics that are either opioids or hallucinogens. In parallel the debate on "soft" and "hard" drugs is also quite active. Given the importance of the subject we have formed a set of nineteen (19) compounds ranging from opium alcaloids and their synthetic analogues, to sugar, caffeine and nicotine as to study their connections. What is intriguing is the fact that clinical findings have shown that there are compounds that share little or no structural similarity, they are accelerating more or less the same psychokinetic or psychomimetic effect. Therefore, it is expected that a theoretical study could shed some light on the subject. Thus, we followed two independent paths, which are the metric space approach, as to account for chemical similarity of them, and the other one, the calculation of *Tanimoto* and *Tversky* indices, that both classify similarity through chemical structure screening. What we do expect from the present study is on the one hand to validate the metric space approach, by being able to at least predict the same results as the indices do, and on the other hand to support the clininal observations and to initiate further studies on the subject.

Details of both theoretical approaches will be given in the *Theory* section. Our findins along with the discussion of them could be found in the *Results and Discussion* section. Finally, we conclude with the take home messages in the *Conclusion*.

2 Theory

2.1 Metric Space Approach

To apply the metric space approach we should set the space, that is formed by the theoretical descriptions and in particular physicochemical properties of compounds of interest. Every single object of this space turned into a vector, with the linear independent properties (meaning that there is not a direct relation from one property to the other) of the object playing the role of the coordinates of the vector. Thus an $n-$dimensional vector space is constructed. The major difficulty is that for $n > 3$ neither the presentation nor the searching for distances among the objects could be characterized as an easy task. At this point it is the use of a *metric*, that is a function that satisfies the minimal properties we might expect of a distance [3], that allows for a direct comparison between vectors. In this study the Minkowski metric [4] $d(i,j) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ (n equals to the dimensions of the space) has been implemented:

$$d(i,j) = \left(\sum_{\alpha} \frac{(Q_{i,\alpha} - Q_{j,\alpha})^p}{(\max_{i,j} (Q_{i,\alpha} - Q_{j,\alpha}))^p} \right)^{\frac{1}{p}} \tag{1}$$

Let I_c be a set that contains the studied compounds and I_p a set that contains the properties used to describe the compounds. In this equation Q stands for the value of the property $\alpha \in I_p$ that compounds i and $j \in I_c$ have. For example: Let α be e.g. *polar surface* then i would be compound $\mathcal{A}$ and j compound $\mathcal{B}$ (it is obvious that at some point $i = j$, therefore their distance is zero).

Since Minkowski metric is a generalization of different distances, in this study the Euclidean distance, thus p=2, has been employed as to calculate the point to point distances. The indices i, j refer to two different vectors (here compounds), whereas α corresponds to the property. Thus, the denominator $\max_{i,j} (Q_{i,\alpha} - Q_{j,\alpha})$ is the maximum difference between values of the same property and its role is to normalize $d_{i,j}$, respectively. In the context of the present article we might also use the $S_{i,j}$ that could be easily derived from $d_{i,j}$ as:

$$S_{i,j} = 1 - \frac{d_{i,j}}{d_{i,j}^{max}},\ 0 \leq S_{ij} \leq 1 \tag{2}$$

This methodology has been introduced in the pioneering papers by Maroulis et. al [5] and has been succesfully applied in numerous studies since then (see for example [6, 7]).

2.2 Structural Similarity Indices (Tanimoto coefficient (T_c) and Tversky index T_v).

In order to quantify the structural similarity of the compounds we have implemented three different indices or coefficients. The procedure starts with the transformation of every single 2D structured chemical formula into a vector that is called *fingerprint*. To form this vector a mathematical trick has been used as to make it binary: the presence, takes "1", and the absence takes "0" of a specific fragment. The *fingerprints* of compounds have been used as to find common features of $\mathcal{A}$ and $\mathcal{B}$ ($f(\mathcal{A} \cap \mathcal{B})$), features that are present in $\mathcal{A}$ and not in $\mathcal{B}$ ($f(\mathcal{A} - \mathcal{B})$) and in $\mathcal{B}$ and not in $\mathcal{A}$ ($f(\mathcal{B} - \mathcal{A})$).

$$S_{\mathcal{A},\mathcal{B}} = \frac{f(\mathcal{A} \cap \mathcal{B})}{f(\mathcal{A} \cap \mathcal{B}) + \alpha f(\mathcal{A} - \mathcal{B}) + \beta f(\mathcal{B} - \mathcal{A})} \tag{3}$$

The coefficients α and β obey the rule $\alpha + \beta = 1$. However, by setting $\alpha = \beta = 1$ the T_v=T_c that is the Tanimoto coefficient [8], which is based on the *Jaccard* index [9] a well known tool in the field of *data mining*. For this reason the input is the SMILE structure of the compound and by using FP2 fingerprint we have calculated T_C similarities by performing computations via the opensource package *Open Babel* [11] (more information on the calculation of the index can be found in [10]). Given that T_c's index counts for similiraty between compounds a value of 1, should be interpeted as a perfect match, while lower values are a sign of dissimilarity. A further similarity analysis based on the *SkelSpheres* descriptor, as it is described in *Datawarrior* [12], has also been performed and the respective T_c's have been calculated. Two other indices that measure the structural similarity of compounds are the *Tversky* index (T_v) [1] and *Dice* coefficient T_D [13], respectively. Both of them have been computed with an evaluation copy of *Schrödinger* suite [14]. The values of α varied as 0.1, 0.2, 0.3, 0.5, 0.7, 0.8, 0.9 (and the β's respectively) and we note that when $\alpha = \beta = 0.5$ the T_v corresponds to T_D.

2.3 Minimum Spanning Tree

By using the calculated distances one can draw an acyclic graph (a connected graph where there is a unique path between every pair of vertices) that connects vertices (in the present study compounds). The total cost, that is the sum of all the distances, of the path that is formed is the minimum, thus the compounds that are the most similar (i.e. less distant) form neighbourhoods. This evaluation has been done by using the Prim's Algorithm [15]: In brief, given the distance matrix that contains the distances between all pairs of i and

j, it starts from an arbitrary compound (vertex A) and finds the compound (vertex B) with the least distance. At this point, it starts by either A or B as to find the next closest compound (vertex C) without creating a closed path. The procedure, continues till the summation over all paths (that is distances) is the minimum of all other possible ones.

2.3.1 Computational protocol

A target set of nineteen compounds, all related to pain relief, pleasure and euphoric effects as well as psychokinetic and psychomimetic activity, used in this study as to form the set I_c =(*cocaine*, *codeine*, *heroine*,*methadone*, *ecstasy*, *mescaline*, *caffeine*, *serotonin*, *adrenaline*, *morphine*, *lysergide acid (LSD)*, *nicotine*, *marijuana (THC)*,*sugar*, *endorphine-1*, *endorphine-2*, *fentanyl*, *carfentanil*, *furanylfentanyl*) (see Fig. 1). The properties are: I_p =(*polarizability*, *polar surface*, *density*, *flash point*, *boiling point*, *molecular weight*) and have been tabulated in Table 1 (all values have been taken from *www.chemspider.com*). These properties have been chosen (for their validity see Ref. [16]) as to construct the space of the theoretical descriptions (TD) on the basis that they describe both electronic structure as well as structural features of compounds.

Table 1. Properties that construct the space of theoretical descriptions

	Polar Surface $Å^2$	Polarizability $(10^{-24}\ cm^3)$	Density (g/cm^3)	Flash Point (^oC)	Boiling Point (^oC in mmHg)	Molecular Weight (Da)
Cocaine	56	32.2	1.2	192.8	395.2	303.353
Codeine	42	32.8	1.3	232.2	462.0	299.364
Heroine	65	38.5	1.4	251.9	492.9	369.411
Methadone	20	38.0	1.0	126.5	423.7	309.445
Ecstasy	30	21.7	1.1	113.2	283.4	193.242
Mescaline	54	23.5	1.1	145.8	312.1	211.258
Coffeine	58	20.0	1.5	205.9	416.8	194.191
Serotonin	62	21.2	1.3	205.4	416.1	176.094
Adrenaline	73	19.6	1.3	207.9	413.1	183.204
Morphine	53	30.9	1.4	241.8	476.2	285.338
LSD	39	38.5	1.2	281.2	541.3	323.432
Nicotine	16	19.5	1.0	101.7	244.4	162.115
THC	29	37.9	1.0	149.3	390.4	314.462
Sugar	190	28.1	1.8	375.4	697.1	342.297
Endomorphine-1	184	67.4	1.3	590.5	1052.8	610.7
Endomorphine-2	168	62.5	1.3	541.9	972.4	571.67
Fentanyl	24	41.1	1.1	185.8	466.2	336.47
Carfentanil	50	45.4	1.1	261.1	508.1	394.51
Furanylfentanyl	37	44.3	1.2	266.6	517.2	374.48

(all values have been taken from *www.chemspider.com.*)

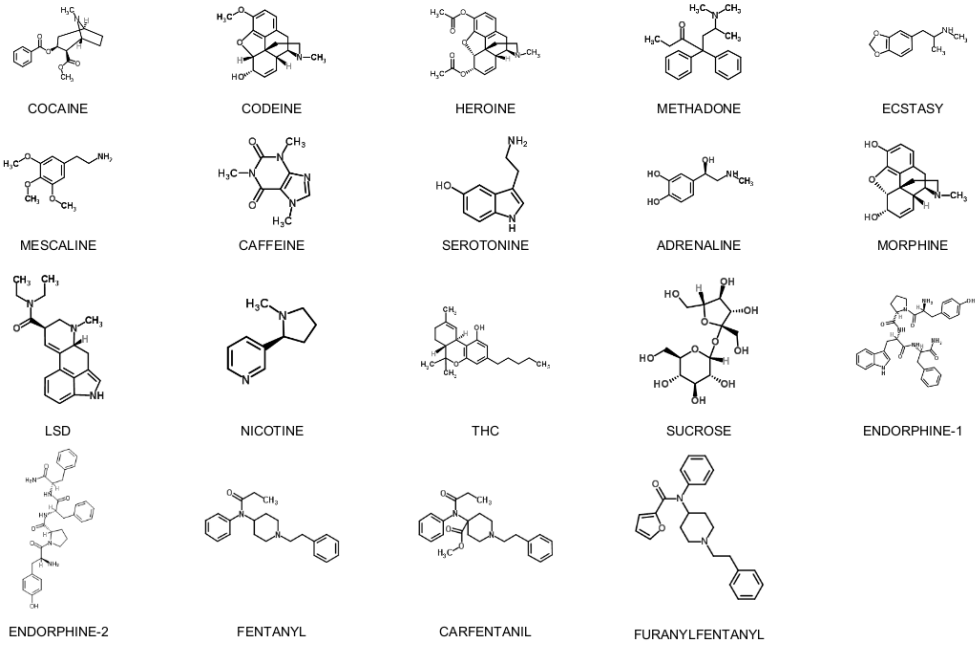

Figure 1. Chemical Structures of the Studied Compounds

It is worth noting at this point that T_c and T_v indices assign similarity only to common structural resemblance, while *metric space* approach takes into account overall characteristics of a certain object. Thus, it might be of some importance to check the relative performance, by evaluating their findings on the basis of how rational are the formed neighbourhoods (i.e. clusters of compounds that are the most similar) given what is known in the literature. To picture the value of the calculated similarities or distances we have constructed a minimum spanning tree (MST) that clusters compounds that are the less distant.

3 Results and Discussion

Every member of the group of compounds that has been formed does respect the saying that "*the dose makes the poison*" [17]. However nicotine, sugar, and caffeine are legally available, whereas for the vast majority of the others they are reachable only after a medical prescription. Therefore, a large part of the following discussion will be devoted to the interconnections between members of these two groups.

3.1 Similarities Based on Structural Patterns

The results of the calculated similarities using Tanimoto index (based on two different fingerprints, namely FP2 and Skelspheres), Tversky index (by varying the weight of the importance of $f(\mathcal{A}-\mathcal{B})$ by setting α values to 0.1, 0,2, 0.3, 0,7, 0.8 and 0.9 ($\beta = 1-\alpha$)) and Dice coefficient when $\alpha = 0.5$, are listed in Table 2. Only values that equal or greater of 0.80 are significant, however when available we have included lower values as to show the differences between the values of different indices.

Since the similarity of other compounds to sucrose or of sucrose to other compounds are well below the 0.80 threshold it is fair to assume that is the less simililar to all other compounds used in the study. Nevertheless, clinical findings suggest that there is a kind of addiction to sugar [18, 19]. Furthermore it is interesting the fact that it is not only causing the same rewarding effect as some of the common drugs but it is more intense in the case of sugar consumption [18, 19], though not by using the same pathways [18, 20].

On that side, of the less structural similarity, are the results of both nicotine and caffeine. Nevertheless, it has been observed that nicotine exhibits an antistress activity [23–25], is causing illusions [26] as well as it causes or it could cause addiction [27], and is

widely known that the vast majority of the studied compounds are causing similar effects. When the discussion comes to caffeine, that causes dopamine release in specific regions of the brain [27, 28], a matching to cocaine that acts similar in the brain, or due to its calm down effect [29–32], connection to serotonin and endomorphines could have been observed, as well.

Another interesting group of compounds can be formed from those known as analgesics. The very well known morphine has an almost identical structure with both codeine and heroine. Thus, T_c values should be, and they are, rather high: in particular *FP2* based calculations gave $S_{10,2} = 0.99$ and $S_{10,3} = 0.90$, whereas *SkelSpheres* result is $S_{10,2} = 0.92$ and not above the threshold for $S_{10,3}$. It is noted that that both morphine and codeine are opium extracts. On the other hand it is well known that heroine is a synthetic analogue of morphine which was introduced as to fight morphine abuse [33].

The chemical offsprings of the morphine to heroine parents are the compounds that form the fentanyl group. It has been found that fentanyl is as 60-80 times more potent that morphine [34]. For this group of molecules *FP2* based $S_{i,j}$'s are beyond the threshold (however the largest in the specific columns of Table 2). On the other hand the *SkelSpheres* results are $S_{17,18} = 0.84$ and $S_{17,19} = 0.89$.

The last set of similar compounds is the one that contains Endomorphines-1 and 2, with the *FP2* value $S_{15,16} = 0.70$ being below the 0.80 threshold, whereas *SkelSpheres* is $S_{15,16} = 0.94$, which is more rational. In this place it should be mentioned that none of the indices has proven any kind of similarity between endomorphine's and morphine or any of their chemical realtives.

Table 2. Similarities Based on Structural Patterns

Sij	Tanimoto FP2	Tanimoto Skelspheres	Tversky $\alpha = 0.1,\ \beta = 0.9$	Tversky $\alpha = 0.2,\ \beta = 0.8$	Tversky $\alpha = 0.3,\ \beta = 0.7$	Dice $\alpha = 0.5,\ \beta = 0.5$	Tversky $\alpha = 0.7,\ \beta = 0.3$	Tversky $\alpha = 0.8,\ \beta = 0.2$	Tversky $\alpha = 0.9,\ \beta = 0.1$
(2,3)	0.90	0.82	0.89	0.88	0.86	0.84	0.82	0.81	0.80
(2,10)	0.99	0.92	0.91	0.91	0.91	0.92	0.93	0.93	0.94
(3,2)	0.90	0.82	0.80	0.81	0.82	0.84	0.86	0.88	0.89
(3,10)	0.90	none	0.76	0.77	0.78	0.81	0.83	0.85	0.86
(10,2)	0.99	0.92	0.94	0.93	0.93	0.92	0.91	0.91	0.91
(10,3)	0.90	none	0.86	0.85	0.83	0.81	0.78	0.77	0.76
(15,16)	0.70	0.94	0.61	0.64	0.66	0.73	0.80	0.84	0.89
(16,15)	0.70	0.94	0.89	0.84	0.80	0.73	0.66	0.64	0.61
(17,18)	0.67	0.85	0.49	0.47	0.46	0.43	0.40	0.39	0.38
(17,19)	0.60	0.89	0.80	0.76	0.73	0.67	0.62	0.59	0.57
(18,17)	0.67	0.85	0.38	0.39	0.40	0.43	0.46	0.47	0.49
(18,19)	0.60	none	0.30	0.29	0.29	0.28	0.28	0.28	0.27
(19,17)	0.60	0.89	0.57	0.59	0.62	0.67	0.73	0.76	0.80
(19,18)	0.60	none	0.27	0.28	0.28	0.28	0.29	0.29	0.30

3.2 Euclidean distances/similarities

The similarities obtained from the other leg of the present study have been tabulated in Table 3. What is profound after reading these values is that many more connections-similarities have been predicted and will be discussed in the following paragraphs.

The validity of this approach could be assumed by the fact that the molecular similarity discussed in the previous section is also present in this place where chemical similarity is discussed. Another evidence is the fact that the results are in accord with the findings of clinical studies showing that there are connections among compounds that are not characterized by molecular similarity. To procceed with the discussion it is fair to set a threshold of 0.70 as to filter out important (when $S_{i,j} \geq 0.70$) to the less important (when $S_{i,j} < 0.70$) connections, since there two distinct regions of values in almost every column.

We will start the discussion of the values of Table 3 with the compounds that are of high consumption on a daily basis. Nicotine's matching to mescaline, a natural compound known for its elucidative and mood altering effects has been long known [41]. The resemblance to MDMA (ecstasy), though the marginal value of $S_{12,5} = 0.79$, well supported from a study [42] that shows that hallucinogens might be used to assist tobacco cessation, while another review suggests that they exhibit quite the same psychobiological effects [43]. The value of $S_{12,8} = 0.76$ to serotonin has been the subject of some studies (see [23–25] and refs therin) and it has been shown that nicotine acts as an antagonist to serotonin, in more than one ways.

Then, the discussion comes to caffeine that better matches to serotonin $S_{7,8} = 0.89$, a relation that has been studied [44] and found that caffeine alters the serotonin receptors. It has also been found proximal to adrenaline [45, 46] ($S_{7,9} = 0.88$) where caffeine induces adrenaline secretion, and to morphine [47–49] ($S_{7,10} = 0.85$) since it has been found that supports the analgesic effect of morphine.

Of all the compounds with medical uses it is the morphine that shows an extremely high matching to codeine and heroine, which is expected since they are chemical relatives, and to cocaine that has also been observed [50, 51] since they are both inducing same receptors while not via the same mechanisms.

The widely known and abused marijuana, contains the psycoactive tetrahydrocannabinol (THC), very well matches to methadone ($S_{13,4} = 0.96$), and this has been the subject of previous clinical studies [52, 53], a chemical compound that has been introduced as to

help heroine addicts to rehab. Some further matching to fentanyl ($S_{13,17} = 0.91$) and the other two members of this family ($S_{13,18} = 0.82$, $S_{13,19} = 0.81$) has not been examined yet. It is worth noting that (THC) shows a high matching to both cocaine ($S_{13,1} = 0.85$) and MDMA ($S_{13,5} = 0.85$) that has been well depicted in a recent study on the dopamine release and the reward effect caused by cannabis [54, 55].

The first round of this discussion ends with the fentanyl group of compounds; while the matching to each other seems reasonable it is worth noting the one between fentanyl and both THC and methadone. Concerning the similarity to the latter lots of studies have shown that methadone and fentanyl might be the best alternative to morphine when the question comes to pain relief without the undesired side effects [56, 57].

Then we formed a set of molecules containing only drugs (nicotine, caffeine and sucrose excluded) and we performed again the calculation of their similarities (see Table 4).

Table 3. Euclidean distance based similiraties: (1=cocaine, 2=codeine, 3=heroine, 4=methadone, 5=ecstasy, 6=mescaline, 7=caffeine, 8=serotonin, 9=adrenaline, 10=morphine, 11=LSD, 12=nicotine, 13=THC, 14=Sugar, 15=endomorphine-1, 16=endomorphine-2, 17=fentanyl, 18=carfentanil, 19=furanylfentanyl).

S_{ij}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	0.92	0.84	0.83	0.81	0.85	0.77	0.83	0.82	0.87	0.86	0.74	0.85	0.47	0.25	0.34	0.86	0.82	0.83
2	0.92	1.00	0.88	0.79	0.77	0.80	0.80	0.83	0.81	0.93	0.90	0.69	0.80	0.49	0.27	0.36	0.85	0.81	0.85
3	0.84	0.88	1.00	0.72	0.69	0.70	0.74	0.74	0.73	0.89	0.86	0.58	0.73	0.57	0.36	0.45	0.79	0.82	0.86
4	0.83	0.79	0.72	1.00	0.85	0.79	0.64	0.71	0.69	0.73	0.80	0.75	0.96	0.32	0.19	0.28	0.91	0.81	0.80
5	0.81	0.77	0.69	0.85	1.00	0.82	0.68	0.75	0.72	0.73	0.75	0.79	0.85	0.32	0.13	0.22	0.81	0.72	0.73
6	0.85	0.80	0.70	0.79	0.82	1.00	0.76	0.86	0.85	0.77	0.74	0.86	0.81	0.39	0.12	0.22	0.76	0.69	0.70
7	0.77	0.80	0.74	0.64	0.68	0.76	1.00	0.89	0.88	0.85	0.71	0.67	0.64	0.54	0.15	0.24	0.66	0.61	0.65
8	0.83	0.83	0.74	0.71	0.75	0.86	0.89	1.00	0.97	0.84	0.75	0.76	0.72	0.49	0.16	0.25	0.71	0.66	0.69
9	0.82	0.81	0.73	0.69	0.72	0.85	0.88	0.97	1.00	0.83	0.74	0.74	0.70	0.51	0.16	0.26	0.70	0.65	0.67
10	0.87	0.93	0.89	0.73	0.73	0.77	0.85	0.84	0.83	1.00	0.85	0.66	0.74	0.55	0.27	0.36	0.78	0.76	0.80
11	0.86	0.90	0.86	0.80	0.75	0.74	0.71	0.75	0.74	0.85	1.00	0.65	0.81	0.47	0.33	0.42	0.88	0.88	0.92
12	0.74	0.69	0.58	0.75	0.79	0.86	0.67	0.76	0.74	0.66	0.65	1.00	0.75	0.25	0.00	0.09	0.69	0.60	0.60
13	0.85	0.80	0.73	0.96	0.85	0.81	0.64	0.72	0.70	0.74	0.81	0.75	1.00	0.33	0.20	0.29	0.91	0.82	0.81
14	0.47	0.49	0.57	0.32	0.32	0.39	0.54	0.49	0.51	0.55	0.47	0.25	0.33	1.00	0.40	0.47	0.37	0.43	0.45
15	0.25	0.27	0.36	0.19	0.13	0.12	0.15	0.16	0.16	0.27	0.33	0.00	0.20	0.40	1.00	0.90	0.26	0.37	0.35
16	0.34	0.36	0.45	0.28	0.22	0.22	0.24	0.25	0.26	0.36	0.42	0.09	0.29	0.47	0.90	1.00	0.35	0.46	0.45
17	0.86	0.85	0.79	0.91	0.81	0.76	0.66	0.71	0.70	0.78	0.88	0.69	0.91	0.37	0.26	0.35	1.00	0.88	0.89
18	0.82	0.81	0.82	0.81	0.72	0.69	0.61	0.66	0.65	0.76	0.88	0.60	0.82	0.43	0.37	0.46	0.88	1.00	0.93
19	0.83	0.85	0.86	0.80	0.73	0.70	0.65	0.69	0.67	0.80	0.92	0.60	0.81	0.45	0.35	0.45	0.89	0.93	1.00

Table 4. Euclidean distance based similiraties: (1=cocaine, 2=codeine, 3=heroine, 4=methadone, 5=ecstasy, 6=mescaline, 8=serotonin, 9=adrenaline, 10=morphine, 11=LSD, 13=THC, 15=endomorphine-1, 16=endomorphine-2, 17=fentanyl, 18=carfentanil, 19=furanylfentanyl).

Sij	1	2	3	4	5	6	8	9	10	11	13	15	16	17	18	19
1	1.00	0.86	0.73	0.72	0.77	0.81	0.78	0.78	0.75	0.85	0.74	0.15	0.26	0.81	0.77	0.81
2	0.86	1.00	0.83	0.62	0.67	0.69	0.81	0.79	0.87	0.85	0.62	0.18	0.29	0.73	0.71	0.81
3	0.73	0.83	1.00	0.48	0.54	0.55	0.69	0.69	0.87	0.74	0.49	0.27	0.38	0.61	0.63	0.74
4	0.72	0.62	0.48	1.00	0.80	0.75	0.55	0.54	0.49	0.70	0.96	0.04	0.13	0.86	0.76	0.70
5	0.77	0.67	0.54	0.80	1.00	0.81	0.65	0.63	0.57	0.70	0.80	0.00	0.10	0.79	0.69	0.68
6	0.81	0.69	0.55	0.75	0.81	1.00	0.74	0.73	0.60	0.70	0.76	0.00	0.10	0.74	0.66	0.65
8	0.78	0.81	0.69	0.55	0.65	0.74	1.00	0.96	0.80	0.70	0.56	0.06	0.17	0.62	0.57	0.64
9	0.78	0.79	0.69	0.54	0.63	0.73	0.96	1.00	0.79	0.69	0.55	0.07	0.18	0.61	0.57	0.63
10	0.75	0.87	0.87	0.49	0.57	0.60	0.80	0.79	1.00	0.74	0.50	0.18	0.28	0.61	0.59	0.70
11	0.85	0.85	0.74	0.70	0.70	0.70	0.70	0.69	0.74	1.00	0.71	0.24	0.34	0.83	0.84	0.92
12	0.74	0.62	0.49	0.96	0.80	0.76	0.56	0.55	0.50	0.71	1.00	0.05	0.15	0.86	0.78	0.71
15	0.15	0.18	0.27	0.04	0.00	0.00	0.06	0.07	0.18	0.24	0.05	1.00	0.89	0.14	0.27	0.27
16	0.26	0.29	0.38	0.13	0.10	0.10	0.17	0.18	0.28	0.34	0.15	0.89	1.00	0.24	0.36	0.37
17	0.81	0.73	0.61	0.86	0.79	0.74	0.62	0.61	0.61	0.83	0.86	0.14	0.24	1.00	0.87	0.84
18	0.77	0.71	0.63	0.76	0.69	0.66	0.57	0.57	0.59	0.84	0.78	0.27	0.36	0.87	1.00	0.87
19	0.81	0.81	0.74	0.70	0.68	0.65	0.64	0.63	0.70	0.92	0.71	0.27	0.37	0.84	0.87	1.00

After a careful reading of the values listed in Table 4, where nicotine, caffeine and sucrose have been left out of the computations, one can easily see the proximity among morphine, codeine and heroine ($S_{10,2} = 0.87$ and $S_{10,3} = 0.87$). The matchings that are also of interest are these of morphine to serotonin $S_{10,7} = 0.80$ and to adrenaline $S_{10,8} = 0.79$, which have been studied in the frame of the opioids association with serotonin syndrome ([62, 63]). Speaking about serotonin its proximity to adrenaline ($S_{7,8} = 0.96$) seems fair due to the fact that are both neurotransmitters that contribute to how exercise affects brain function [64].

The case of cocaine also deserves some discussion since it is quite similar to codeine ($S_{1,2} = 0.86$), a connection that has been studied [65], to LSD $S_{1,11} = 0.85$ while it is known that they are following similar ways of toxiticity in the central nervous system [66], Fentanyl $S_{1,17} = 0.81$ and Furanylfentanyl $S_{1,2} = 0.81$, while the literature on Fentanyls are not so rich since they have been recently introduced in the class of abused drugs.

To provide a more effective way on reading the findings listed in Tables 3 and 4 we procceed with the construction of the minimum spanning trees. The outcome of this analysis has been depicted in Figs. **??** and **??**. In brief, we used the distance values between the full set of compounds and then these of the one that produced after nicotine, caffeine and sugar having being left out and by applying the Prim's algorithm we found the (acyclic) path that joins the most relevant compounds. What follows the construction of the trees is the pruning of these branches that are above a threshold. It is obvious that the lower the threshold the tighter the survived connections (branches) thus the higher the chemical similarity of the compounds. In both figures **??**, **??** clusters have been formed after having prune all distances that are above 0.30 (D_T=0.30).

In Fig. **??** where the full set of molecules is being presented all the abused drugs are forming a neighbourhood. The nicotine, serotonin and adrenaline are grouped together, when the two types of endomorphine form an another cluster.

In the filtered set of compounds, see Fig. **??**, cocaine, morphine, codeine and heroine are forming a group with tight connections. Another cluster with the same property has been formed from LSD, fentanyl, carfentanil, furanylfentanil, THC, methadone. While nicotine is not presented serotonin and adrenaline are still together. What is also happens for endomorphines.

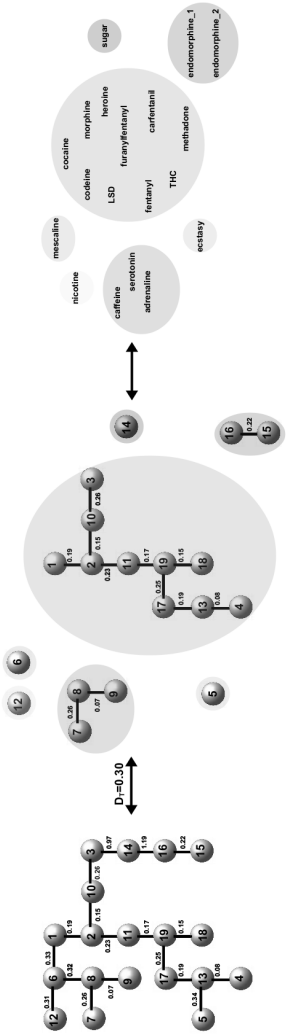

Figure 2. Minimum Spanning Tree Based on Euclidean Distances (full set of nineteen compounds).

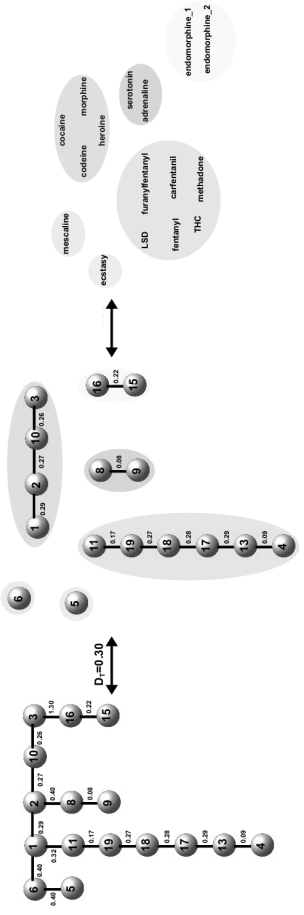

Figure 3. Minimum Spanning Tree Based on Euclidean Distances (nicotine, caffeine and sugar not included).

4 Conclusion

The present study reports on the findings of two independent approaches: one leg is based solely on the structural resemblance see *molecular similarity* via different indices i.e. *Tanimoto*, *Tversky* and *Dice* as well as different descriptors i.e. *FP2* and *Skelspheres*, and the other leg that is *Metric Space Approach* assigns similarities after performing calculations that are based on an number of linear independent characteristics (here physicochemical properties, thus *chemical similarity*) of the objects (here compounds). These two different types of analyses brought fourth relations that have been observed in clinical experiments that are in most cases based on small samples.

The findings of the molecular similarity apporach suggest that compounds could be clustered as A_T=(*morphine, heroine, codeine*), of B_T=(*fentanyl, carfentanil, furanylfentanyl*) and of C_T=(*endomorphine1, endomorphine2*), in great agreement to the T_v's. On the other hand the results of the chemical similarity endeavour not only led to the formation of these clusters but also predicted further connections as: A_E=(*caffeine*, *codeine*, *serotonin*, *adrenaline* and *morphine*), B_E=(*nicotine*, *mescaline*), C_E=(*THC*, *methadone*, *fentanyl*, *cocaine*, *ecstasy*, *carfentanil*, *furanylfentanyl*) and quite similar results for MDMA, and LSD, that are in agreement to clinical observations. The C_E cluster that describes the strong connections of marijuana (THC) to cocaine, morphine, methadone and fentanyl is an evidence that offers another break on the artificial wall between "soft" and "hard" drugs. To conclude, sucrose though included in the present study has not been found similar to the other compounds by none of the two approaches.

References

[1] A. Tversky, Features of similarity, *Psychol. Rev.* **84** (1977) 327–352.

[2] C. D. Selassie, History of quantitative structure–activity relationships, in: D. J. Abraham (Ed.), *Burger's Medicinal Chemistry and Drug Discovery*, Wiley, Hoboken, 2003, pp 1–48.

[3] W. P. Ziemer, *Modern Real Analysis*, Dept. Math., Indiana Univ., Bloomington, 2017.

[4] E. Deza, M. M. Deza, *Encyclopedia of Distances*, Springer, New York, 2009, pp. 94–94.

[5] G. Maroulis, M. Sana, G. Leroy, Molecular properties and basis set quality: an approach based on information theory, *Int. J. Quantum Chem.* **19** (1981) 43–60; G. Maroulis, M. Sana, G. Leroy, Information theory and basis set quality: automatic classification and systematic improvement of approximate molecular wavefunctions; A study on OH radical, *J. Mol. Struct. (Theochem)* **110** (1985) 107–122; G. Maroulis, M. Sana, G. Leroy, Information theory and basis set quality, automatic classification and systematic improvement of approximate wavefunctions: III. A study on H_2O ($1\ a_1^2\ 2\ a_1^2\ 1\ b_2^2\ 3\ a_1^2\ 1\ b_1^2,\ ^1A_1$), *J. Mol. Struct. (Theochem)* **121** (1985) 69–78.

[6] G. Maroulis, A systematic study of basis set, electron correlation, and geometry effects on the electric multipole moments, polarizability, and hyperpolarizability of HCl, *J. Chem. Phys.* **108** (1998) 5432–5448; G. Maroulis, Applying conventional ab initio and density functional theory approaches to electric property calculations. Quantitative aspects and perspectives, *Struct. Bond.* **149** (2012) 95–130.

[7] G. Maroulis, D. Xenides, On the performance of DFT methods on electric polarizability and hyperpolarizability calculations for the lithium tetramer, *Comput. Lett.* **1** (2005) 246–252; D. Xenides, On the performance of DFT methods in (hyper)polarizability calculations: N4 (T_d) as a test case, *J. Molec. Struct. (Theochem)* **804** (2007) 41–46; C. Christodouleas, D. Xenides, T. E. Simos, Trends of the bonding Effect on the performance of DFT methods in electric properties calculations: A pattern recognition and metric space approach on some XY_2 (X = O, S and Y = H, O, F, S, Cl) molecules, *J. Comput. Chem.* **31** (2010) 412–420.

[8] T. T. Tanimoto, IBM Internal Report 17th Nov, 1957; D. J. Rogers, T. T. Tanimoto, A computer program for classifying plants, *Science* **132** (1960) 1115–1118.

[9] P. Jaccard, Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines, *Bull. Soc. Vaudoise Sci. Nat.* **37** (1901) 241–272.

[10] T. G. Kristensen, J. Nielsen, C. N. Pedersen, A tree-based method for the rapid screening of chemical fingerprints,*Algorit. Mol. Biol.* **5** (2010) 9–18.

[11] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open Babel: An open chemical toolbox, *J. Cheminf.* **3** (2011) 1–14.

[12] T. Sander, J. Freyss, M. von Korff, C. Rufener, DataWarrior: An open-source program for chemistry aware data visualization and analysis, *J. Chem. Inf. Model.* **55** (2015) 460–473.

[13] L. R. Dice, Measures of the amount of ecologic association between species, *Ecology* **26** (1945) 297–302; T. Sørensen, A method of establishing groups of equal amplitude

in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons, *Kongelige Danske Videnskabernes Selskab* **4** (1948) 1–34.

[14] J. Duan, S. L. Dixon, J. F. Lowrie, W. Sherman, Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods, *J. Mol. Graph. Model.* **29** (2010) 157–170; M. Sastry, J. F. Lowrie, S. L. Dixon, W. Sherman, Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments, *J. Chem. Inf. Model.* **50** (2010) 771–784.

[15] R. C. Prim, Shortest connection networks and some generalizations *Bell Syst. Tech. J.* **36** (1957) 1839–1401; H. T. Cormen, E. C. Leiserson, L. R. Rivest, C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, 2001.

[16] A. G Maldonado, J. P. Douchet, M. Petitjean, B. T. Fan, Molecular similarity and diversity in chemoinformatics: from theory to applications, *Mol. Diver.* **10** (2006) 39–79.

[17] T. Paracelsus, Die dritte Defension wegen des Schreibens der neuen Rezepte, Septem Defensiones 1538. Werke Bd. 2, Darmstadt, pp. 510, 1965.

[18] N. M. Avena, P. Rada, B. G. Hoebel, Evidence for sugar addiction: Behavioral and neurochemical effects of intermittent, excessive sugar intake, *Neurosci. Biobehav. Rev.* **32** (2008) 20–39.

[19] S. H. Ahmed, K. Guillem, Y. Vandaele, Sugar Addiction: pushing the drug-sugar analogy to the limit, *Curr. Opin. Clin. Nutr. Metab. Care* **16** (2013) 434–439.

[20] C. Rob Markus, P. R. J. Rogers, F. Brouns, R. Schepers, Eating dependence and weight gain; no human evidence for a 'sugar-addiction' model of overweight, *Appetite* **114** (2017) 64–72.

[21] S. K. Jash, D. Gorai, Sugar derivatives of morphine: a new window for the development of potent anesthetic drugs, *Nat. Prod. Bioprospect.* **5** (2015) 111–127.

[22] S. Galanie, K. Thodey, I. J. Trenchard, M. Filsinger Interrante, C. D. Smolke, Complete biosynthesis of opioids in yeast, *Science* **349** (2015) 1095–1100.

[23] S. Singer, S. Rossi, S. Verzosa, A. Hashim, R. Lonow, T. Cooper, H. Sershen, A. Lajtha, Nicotine-induced changes in neurotransmitter levels in brain areas associated with cognitive function, *Neurochem. Res.* **29** (2004) 1779–1792.

[24] H-G. A. Breitinger, N. Geetha, G. P. Hess, Inhibition of the serotonin 5-HT_3 receptor by nicotine, cocaine, and fluoxetine investigated by rapid chemical kinetic techniques, *Biochemistry* **40** (2001) 8419–8429.

[25] M. Pierucci, S. Chambers, L. Partridge, P. De Deurwaerdère, G. Di Giovanni, Role of central serotonin receptors in nicotine addiction, in: R. Lester (Ed.), *Presynaptic Nicotinic Acetylcholine Receptors: Subtypes and Functions*, Humana Press, New York, 2014, pp. 279–305.

[26] T. Isomura, T. Murai, M. Kano, Paradise lost: a new paradigm for explaining the interaction between neural and psychological changes in nicotine addiction patients, in: R. R. Watson, S. Zibadi (Eds.),*Addictive Substances and Neurological Disease*, Academic Press, London, 2017, pp. 281–291.

[27] R. A. Wise, M. A. Bozarth, A psychomotor stimulant theory of addiction, *Psychol. Rev.* **94** (1987) 469–492.

[28] F. E. Pontieri, G. Tanda, G. Di Chiara, Intravenous cocaine, morphine, and amphetamine preferentially increase extracellular dopamine in the "shell" as compared with the "core" of the rat nucleus accumbens, *Proc. Natl. Acad. Sci.* **92** (1995) 12304–12308.

[29] R. N. Hughes, Neurobehavioral aspects of caffeine, in: V. R. Preedy (Ed.), *In Neuropathology of Drug Addictions and Substance Misuse*, Academic Press, San Diego, 2016, pp. 775–784.

[30] P. A. Adeniyi, E. P. Omatsuli, A. J. Akinyemi, A. O. Ishola, Caffeine plus nicotine improves motor function, spatial and non-spatial working memory and functional indices in BALB/c male mice, *Pathophysiology* **23** (2016) 251–258.

[31] L. O. Porciúncula, C. Sallaberry, S. Mioranzza, P. H. S. Botton, D. B. Rosemberg, The Janus face of caffeine, *Neurochem. Int.* **63** (2013) 594–609.

[32] S. E. Meredith, L. M. Juliano, J. R. Hughes, R. R. Griffiths, caffeine use disorder: A comprehensive review and research agenda, *J. Caff. Res.* **3** (2013) 114–130.

[33] J. Merry, A social history of heroin addiction, *Addiction* **70** (1975) 307–310.

[34] A. Miller, A. McKee, C. D. Mazer, Sedation, analgesia, and related topics, in: D. Sidebotham, A. Mckee, M. Gillham, J. H. Levy (Eds.), *Cardiothoracic Critical Care*, Butterworth–Heinemann, Philadelphia, 2007, pp. 57–80 .

[35] C. Cadoni, V. Valentini, G. Di Chiara, Behavioral sensitization to Δ^9-tetrahydrocannabinol and cross-sensitization with morphine: differential changes in accumbal shell and core dopamine transmission, *J. Neurochem.* **106** (2008) 1586–1593.

[36] P. E. Gilbert, Comparison of THC, nantradol, nabilone, and morphine in the chronic spinal dog, *J. Pharmacol.* **21** (1981) 311S–319S.

[37] C. Colantuoni, P. Rada, J. McCarthy, C. Patten, N. M. Avena, A. Chadeayne, B. G. Hoebel, Evidence that intermittent, excessive sugar intake causes endogenous opioid dependence, *Obes. Res.* **10** (2002) 478–488.

[38] D. J.Mysels, M. A. Sullivan, The relationship between opioid and sugar intake: Review of evidence and clinical applications, *J. Opioid Manag.* **6** (2010) 445–452.

[39] S. Ullrich, Y. C. de Vries, S. Kühn, D. Repantis, M. Dresler, K. Ohla, Feeling smart: Effects of caffeine and glucose on cognition, mood and self-judgment, *Physiology & Behavior* **151** (2015) 629–637.

[40] D. E. Walters, How are bitter and sweet tastes related? *Trends Food Sci. Technol.* **7** (1996) 399–403 .

[41] Castellano, Effects of mescaline and psilocin on acquisition, consolidation, and performance of light-dark discrimination in two inbred strains of mice, *C. Psychopharmacology* **59** (1978) 129–137.

[42] M. W. Johnson, A. Garcia-Romeu, M. P. Cosimano, R. R. Griffiths, Pilot study of the 5-HT_{2A}R agonist psilocybin in the treatment of tobacco addiction *J. Psychophar.* **28** (2014) 983–992.

[43] A. C. Parrot, Why all stimulant drugs are damaging to recreational users: an empirical overview and psychobiological explanation, *Hum. Psychopharmacol. Clin. Exp.* **30** (2015) 213–224.

[44] D. Shi, O. Nikodijević, K. A. Jacobson, J. W. Daly, Chronic caffeine alters the density of adenosine, adrenergic, cholinergic, GABA, and serotonin receptors and calcium channels in mouse brain, *Cell Mol. Neurobiol.* **13** (1993) 247–261 .

[45] Y. Yamada, Y. Nakazato, A. Ohga, The mode of action of caffeine on catecholamine release from perfused adrenal glands of cat, *Br. J. Pharmacol.* **98** (1989) 351–356.

[46] M. Van Soeren, T. Mohr, M. Kjaer, T. E. Graham, Acute effects of caffeine ingestion at rest in humans with impaired epinephrine responses, *J. Appl. Physiol.* **80** (1996) 999–1005.

[47] J. W. Bellville, Interaction of caffeine and morphine on respiration, *J. Pharmacol. Exp. Ther.* **143** (1964) 165–168.

[48] D.L. Person, I. Kissin, P. T. Brown, A. V. Xavier, H. R. Vinik, E. L. Bradley, Morphine-caffeine analgesic interaction in rats, *Anesth. Analg.* **64** (1985) 851–856.

[49] S. Mercadante, R. Serretta, A. Casuccio, Effects of caffeine as an adjuvant to morphine in advanced cancer patients: A randomized, double-blind, placebo-controlled, crossover study, *J. Pain Symptom Manag.* **21** (2001) 369–372.

[50] R. F. Kaiko, R. Kanner, K. M. Foley, S. L. Wallenstein, A.M. Canel, A. G. Rogers, R. W. Houde, Cocaine and morphine interaction in acute and chronic cancer pain, *Pain* **31** (1987) 35–45.

[51] N. M. Graziane, S. Sun, W. J. Wright, D. Jang, Z. Liu, Y. H. Huang, E. J. Nestler, Yu T. Wang, O. M. Schlüter, Yan Dong, Opposing mechanisms mediate morphine- and cocaine-induced generation of silent synapses, *Nat. Neurosci.* **19** (2016) 915–925.

[52] J. L. Scavone, R. C. Sterling, S. P. Weinstein, E. J. Van Bockstaele, Impact of cannabis use during stabilization on methadone maintenance treatment, *J. Am. Addict.* **22** (2013) 344–351.

[53] D. H. Epstein, K. L. Preston, Does cannabis use predict poor outcome for heroin-dependent patients on maintenance treatment? A review of past findings, and more evidence against, *Addiction* **98** (2003) 269–279.

[54] B. K. Madras, Dopamine challenge reveals neuroadaptive changes in marijuana abusers, *Proc. Natl. Acad. Sci.* **111** (2014) 11915–11916.

[55] E. van de Giessen, J. J. Weinstein, C. M. Cassidy, M. Haney, Z. Dong, R. Ghazzaoui, N. Ojeil, L. S. Kegeles, X. Xu, N. P. Vadhan, N. D. Volkow, M. Slifstein, A. Abi-Dargham, Deficits in striatal dopamine release in cannabis dependence, *Mol. Psychiatry* **22** (2017) 68–75.

[56] M. A. Benítez-Rosario, M. Feria, A. Salinas-Martín, L. P. Martínez-Castillo, J. J. Martín-Ortega, Opioid switching from transdermal fentanyl to oral methadone in patients with cancer pain, *Cancer* **101** (2004) 2866–2873.

[57] J. Haumann, J. W. Geurts , S. M. van Kuijk , B. Kremer, E. A. Joosten, M. H. J. van den Beuken-van Everdingen, Methadone is superior to fentanyl in treating neuropathic pain in patients with head-and-neck cancer, *Eur. J. Cancer* **65** (2016) 121–129.

[58] J. B. Appel, W. B. West, J. Buggy, LSD, 5-HT (serotonin), and the evolution of a behavioral assay, *Neurosci. Biobehav. Rev.* **27** (2004) 693–701.

[59] E. Hajduch, F. Rencurel, A. Balendran, I. H. Batty, C. P. Downes, H. S. Hundal, Serotonin (5-Hydroxytryptamine), a novel regulator of glucose transport in rat skeletal muscle, *J. Biol. Chem.* **274** (1999) 13563–13568,.

[60] J. O. Tweed, S. H. Hsia, K. Lutfy, T. C. Friedman, The endocrine effects of nicotine and cigarette smoke, *Trends Endocrin. Metab.* **23** (2012) 334–342.

[61] R. S. Moheimani, M. Bhetraratana, K. M. Peters, B. K. Yang, F. Yin, J. Gornbein, J. A. Araujo, H. R. Middlekauff, Sympathomimetic effects of acute e-cigarette use: role of nicotine and non-nicotine constituents, *J. Am. Heart Assoc.* **6** (2017) e006579.

[62] R. Rastogi, R. A. Swarm, T. A. Patel, Case scenario: opioid association with serotonin syndrome, *Anesthesiology* **12** (2011) 1291–1298.

[63] H. Mateo-Carrasco, E. M. Muñoz-Aguilera, J. M. García-Torrecillas, H. Abu Al-Robb, Serotonin syndrome probably triggered by a morphine-phenelzine interaction, *Pharmacotherapy* **35** (2015) e102–e105.

[64] T. W. Lin, Y. M. Kuo, Exercise benefits brain function: the monoamine connection, *Brain Sci.* **3** (2013) 39–53.

[65] R. J. Peters, M. Williams, M. W. Ross, J. Atkinson , G. S. Yacoubian, Codeine cough syrup use among african-american crack cocaine users, *J. Psychoactive Drugs* **39** (2007) 97–102.

[66] A. Vale, Drugs of abuse (amfetamines, BZP, cannabis, cocaine, GHB, LSD), *Medicine* **40** (2012) 84–87.