MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **92** (2024) 317–336
ISSN: 0340–6253
doi: 10.46793/match.92-2.317R

Comparative Cheminformatics Study on Structural and Physiological Similarity of Bridged Heterometallic Complexes

Izudin Redžepović[a,*], **Tanja V. Soldatović**[a]

[a] *Department of Natural-Mathematical Sciences, State University of Novi Pazar, Vuka Karadžića 9, 36300 Novi Pazar, Serbia*

iredzepovic@np.ac.rs, tsoldatovic@np.ac.rs

(Received February 5, 2024)

Abstract

In the chemical and pharmaceutical industries, many procedures are based on the principle that compounds with similar structures tend to exhibit similar activities. Therefore, the development of computational methods for decision support in similarity assessment is of utmost importance. In this proof-of-concept study, we introduce a computational framework for quantitative computation of agreement between structural and physiological similarity of compounds. As model molecules, we use bridged heterometallic complexes. To derive physiological binary vectors we use rate constants for the first substitution reaction between our compounds and some biomolecules. It is believed that such treatment of metal complexes allows a better understanding of the structure-activity relationship and can serve as a guide for new synthetic targets.

1 Introduction

In recent years the development of heterometallic complexes gained significant importance. While the underlying properties of monometallic complexes are relatively well-explored, complexes containing more than one

*Corresponding author.

metal atom often exhibit unexpected and potentially valuable properties. For example, they have been quite helpful in bimetallic catalytic processes, whereas a two-metal catalyst can perform two or more mechanistically distinct transformations [1, 2]. Also, their significantly increased antitumor activity, compared to cisplatin, makes them potential drug candidates [3]. Generally, incorporating different metal ions that differ in Lewis acidity has a synergic effect [4]. The compounds of two metal ions with distinct coordination geometry and kinetic properties will have distinguishable affinity and reactivity towards biologically relevant nucleophiles [3–5]. The ability to coordinate into diverse geometries depending on the specific arrangement of donor atoms in biomolecules increases the cytotoxicity of these molecules [3, 6]. All of this has caused the flourishing of the chemistry of the heterometallic complexes.

Quite recently, four novel complexes [{*cis*-PtCl(NH_3)(μ-4,4'-bipyridyl)ZnCl(terpy)}](ClO_4)$_2$ (C1), [{*trans*-PtCl(NH_3)(μ-4,4'-bipyridyl)ZnCl(terpy)}](ClO_4)$_2$ (C2), [{*cis*-PtCl(NH_3)(μ-pyrazine)ZnCl(terpy)}](ClO_4)$_2$ (C3), and [{*trans*-PtCl(NH_3)(μ-pyrazine)ZnCl(terpy)}](ClO_4)$_2$ (C4) (where terpy = 2,2':6',2"-terpyridine) have been synthesized [6]. The structures of C1-C4 have been depicted in Figure 1.

It has been shown that the *trans*-Pt-L-Zn complexes are more reactive than the *cis*-Pt-L-Zn complexes. Namely, the antiproliferative effect of the complexes, such as the expression of apoptosis mRNA and repair-related genes after cancer cell treatment, indicates that the newly synthesized C2 exhibits highly selective cytotoxicity against HCT116 colon cancer cells [6]. In general, drug design relies on optimizing activity and selectivity for the target biomolecule by structural tuning of a drug [7]. For this reason, molecular similarity plays an important role both in drug design and cheminformatics [8, 9]. It is one of the most heavily exploited concepts, resulting in many computational approaches developed for this purpose [10, 11]. The most recent computational methods include network analysis of molecular clusters obtained by calculating Euclidean distance or by utilizing similarity of physicochemical features [12, 13]. Molecular similarity techniques have been applied to diverse problems, some of them include toxicology screening [14], design of novel zeolite materials [15],

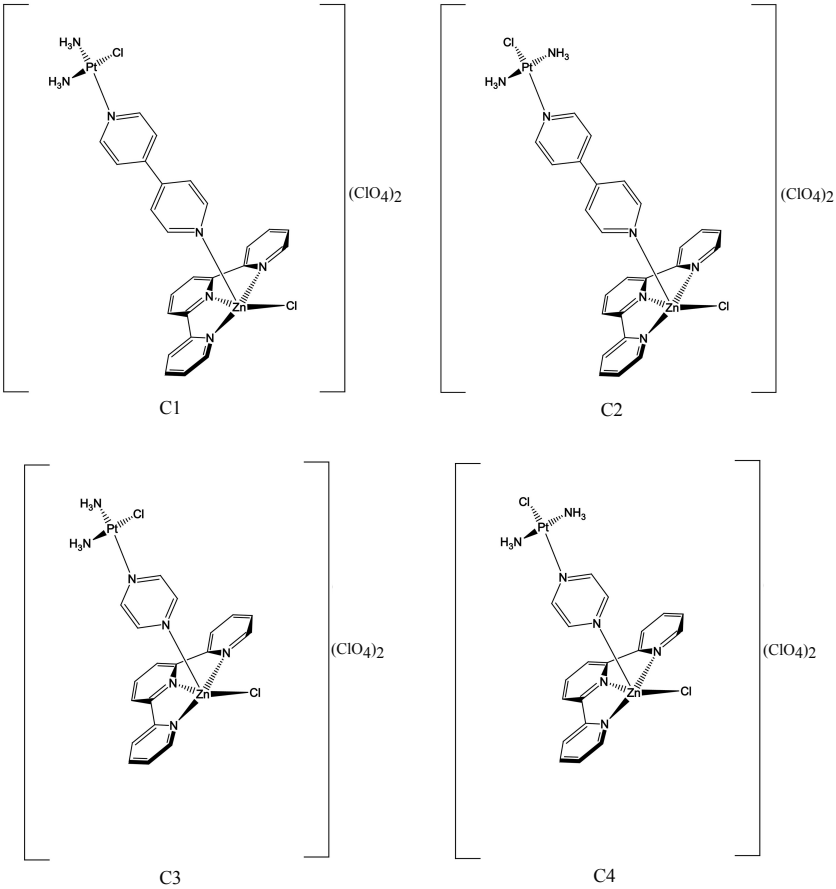

Figure 1. Heterometallic bridged Pt(II)-Zn(II) complexes.

computer-assisted retrosynthesis [16], and the prediction of the crystal structure of a molecule [17]. These applications highlight the broad scope and potential impact of molecular similarity in diverse fields. At the core of molecular similarity lies the principle of structural similarity leading to similar activity [18]. This fundamental concept underpins the aforementioned applications and serves as a guiding principle for understanding the relationships between molecular structure and function. By recognizing the structural similarities between molecules, researchers can make informed predictions about their properties and behaviors. However, the

relationship between a structural feature and the related molecular activity is complicated and not always clear-cut. As a result, there are some obstacles in this concept, including activity cliffs [19–22].

Quantification of the molecular structure is a prerequisite for calculating the similarity of two molecules. Various methods of encoding structural information have been suggested. There is currently a plethora of molecular descriptors accessible, and this number is growing. Descriptors come in a variety of forms, from straightforward ones like counting descriptors to intricate quantum-chemical molecular descriptors [23–26]. Typically, the molecular descriptors are categorized based on the dimensionality of the molecular representation required to compute a descriptor. As a result, descriptors in one, two, three, and higher dimensions exist. Therefore, the selection of an appropriate descriptor that effectively captures and encodes essential molecular information, enabling robust and meaningful comparisons between molecules is an essential step. Structural fingerprints have a special significance in similarity-related calculations [27, 28]. In their simplest form, these are numerical strings made up of zeros and ones.

Another important step in similarity calculation is choosing an adequate similarity metric. The wide recognition of the impact of selecting different similarity metrics on the variability of similarity assessment results is accompanied by a multitude of coefficients proposed in the scientific literature. These coefficients quantify the similarity of structural molecular descriptors [29–31]. They effectively capture the resemblance between two objects and play a crucial role in constructing a similarity matrix that represents pairwise resemblances. Such a matrix serves as valuable input for multivariate data analysis methods, including multidimensional scaling and cluster analysis. The Tanimoto (Jaccard-Tanimoto) index stands out as one of the most popular similarity metrics [32–34].

Since most of the similarity studies and methodologies have been devoted to and deal with similarities related to solely organic molecules, this study aims to provide new applications of similarity techniques by performing this proof-of-concept study. Namely, we show how newly introduced similarity approaches may be applied to the chemistry of metal complexes in order to distinguish and evaluate their structural and physi-

ological similarities. Such a treatment of metal complexes allows a better understanding of the structure-activity relationship and can serve as a guide for new synthetic targets.

2 Methodology details

2.1 Molecular structure descriptors

In order to quantify molecular structure in this work we have used three types of the most popular and frequently used structure descriptors. The extended-connectivity fingerprint [35], also known as Morgan circular fingerprint, with radius 2, has been employed. This descriptor has been converted into an analogous binary vector with a length of 1024 bits. The other two descriptors are RDKit fingerprints [36] and 166 MACCS structural keys [37]. Also, these two descriptors have been converted into binary vectors to calculate the structural similarity of our molecules. It is noteworthy to say that perchlorate ions were excluded from the construction of the fingerprints. These occur in all four complexes and their exclusion does not affect the complex part of the compound, thus does not affect the structural fingerprint and structural similarity between molecules.

2.2 Physiological binary vectors

In our recent paper [13], we have presented a novel methodology for evaluating the chemical similarity of molecules. This approach requires the physicochemical properties of compounds and the calculation of binary fingerprints from them by applying an algorithm based on a binary-coded decimal system. In this way, the physicochemical description of molecules are encoded into binary vectors, which then may be evaluated and compared by applying similarity measures. Here, we adopt this approach to compute the physiological similarities between our compounds. For this purpose, we have used experimental rate constants (Table 1) of the first substitution reaction between C1-C4 complexes and guanosine-5'-monophosphate (5'-GMP), inosine-5'-monophosphate (5'-IMP), and glutathione (GSH) biomolecules presented in [6]. These values simulate the

behavior of our molecules in a living organism, and, therefore, are good descriptors of physiological activity.

Table 1. Experimental rate constants for the first substitution reaction between C1-C4 complexes and 5'-GMP, 5'-IMP, and GSH in $M^{-1} \cdot s^{-1}$.

	5'-GMP	5'-IMP	GSH
C1	$13.87 \cdot 10^2$	$10.62 \cdot 10^2$	$7.25 \cdot 10^2$
C2	$165.73 \cdot 10^2$	$86.07 \cdot 10^2$	$8.12 \cdot 10^2$
C3	$109.11 \cdot 10^2$	$70.01 \cdot 10^2$	$19.83 \cdot 10^2$
C4	$184.13 \cdot 10^2$	$135.65 \cdot 10^2$	$28.75 \cdot 10^2$

As one may see, all rate constants in Table 1 include 10^2 factor, so this constant value was expelled during the construction of a physiological fingerprint without affecting it. A physiological fingerprint for each complex was constructed using the procedure presented in [13], resulting in a binary vector of length 60 bits.

2.3 Similarity indices

Numerous similarity indices have been proposed to deal with binary structure descriptors [38, 39]. Generally, they can all be described as follows. Let x and y be binary vectors that both include variables with a value of 1 or 0. The similarity indices may be then calculated from parameters presented in Table 2,

Table 2. Parameters for describing different events while simultaneously looping over two binary vectors.

	$y_i = 1$	$y_i = 0$	
$x_i = 1$	$a+ = 1$	$b+ = 1$	$a + b$
$x_i = 0$	$c+ = 1$	$d+ = 1$	$c + d$
	$a + c$	$b + d$	p

where a, b, c, and d are the frequencies of the events $(x_i = y_i = 1)$, $(x_i = 1; y_i = 0)$, $(x_i = 0; y_i = 1)$, and $(x_i = y_i = 0)$, respectively. Finally, the $p\,(= a + b + c + d)$ is the total number of variables, i.e., the length of each binary vector.

In this work, we have used the following similarity indices:

$$RR = \frac{a}{p} \tag{1}$$

$$SM = \frac{a+d}{p} \tag{2}$$

$$Ja = \frac{3a}{3a+b+c} \tag{3}$$

$$JT = \frac{a}{a+b+c} \tag{4}$$

$$RT = \frac{a+d}{p+b+c} \tag{5}$$

$$Gle = \frac{2a}{2a+b+c} \tag{6}$$

$$SS_1 = \frac{a}{a+2b+2c} \tag{7}$$

$$CT_3 = \frac{log(1+a)}{log(1+p)} \tag{8}$$

$$CT_4 = \frac{log(1+a)}{log(1+a+b+c)} \tag{9}$$

$$Sor = \frac{a^2}{(a+b)(a+c)} \tag{10}$$

where RR, SM, Ja, JT, RT, Gle, SS_1, CT_3, CT_4, and Sor stand for Russell-Rao [40], Sokal-Michener [41], Jaccard [42], Jaccard-Tanimoto [43], Rogers-Tanimoto [43], Gleason [44], Sokal-Sneath [45], Consonni-Todeschini 3 [46], Consonni-Todeschini 4 [46], and Sorgenfrei index [47], respectively. As can be seen from their definitions, these indices differ in the weight they place on different parts of the vector. For example, Ja and Gle highlight the structural details shared by both vectors, while SS_1 highlights the parts where they differ. Such a selection of similarity measures enables a comprehensive comparison of molecules. For all these computa-

tions, as well as the calculation of molecular descriptors, a Python code with an implemented RDKit package has been devised [36].

2.4 Sum of ranking differences (SRD) statistical procedure

SRD is a relatively new general-purpose statistical procedure introduced by K. Héberger [48]. This is an evaluation technique in methods, procedures, models, etc comparisons. It has been proven to be quite useful in different scenarios, like for the correct splitting of data in QSAR modeling, for the appropriate column selection in chromatography, and for experimental data analysis [49–51]. Also, the quality of the similarity index was assessed using the SRD procedure [32]. This statistical tool is available as an MS Office Excel macro at http://aki.ttk.hu/srd/. Input data should be arranged as a matrix. The results are ranked for each method according to the ranking of experimental (if available) or reference values. If the standard value is not available, like in our case, then the mean value for all methods may be used. In the paper [52] the full description of SRD computation and validation may be found. Generally, the closer the SRD value is to zero (i.e., the closer is the ranking to the golden standard), the better the method. The similarity of the evaluated methods is determined by the proximity of SRD values: the closer the values, the more similar the methods.

3 Results and Discussion

In this section, we present the results of the structural and physiological similarity of molecules depicted in Figure 1. Table 3 presents the percentages of structural similarities between our compounds. These values are calculated by using Morgan circular fingerprints as molecular structure descriptors. As one may see, similarity indices yielded values in different ranges. In Table 1S in Supporting Information correlation coefficients between similarity values are presented. The coefficients show that all values are highly correlated. This indicates that all similarity indices show the

same trends. The lowest similarity values are obtained by the RR index, and the highest by the SM measure. Also, the RR shows the lowest standard deviation, while the SS_1 index yields highly dispersed values.

Table 3. The percentage of structural similarity of complexes C1-C4 calculated using Morgan circular fingerprints. The s denotes the standard deviation of data.

	RR	SM	Ja	JT	RT	Gle	SS_1	CT_3	CT_4	Sor
C1-C2	4.49	99.41	95.83	88.46	98.83	93.88	79.31	55.54	96.97	88.13
C1-C3	3.61	97.75	82.84	61.67	95.61	76.29	44.58	52.47	88.49	58.21
C1-C4	3.42	97.36	79.55	56.45	94.86	72.16	39.33	51.69	86.49	52.08
C2-C3	3.42	97.36	79.55	56.45	94.86	72.16	39.33	51.69	86.49	52.08
C2-C4	3.61	97.75	82.84	61.67	95.61	76.29	44.58	52.47	88.49	58.21
C3-C4	4.39	99.41	95.74	88.24	98.83	93.75	78.95	55.23	96.9	87.89
s	0.44	0.89	7.01	13.97	1.72	9.39	17.66	1.59	4.53	15.69

The C1-C2 and C3-C4 pairs of molecules have been found to have the highest similarity values by all similarity measures, which is expected considering that C1 and C2, and C3 and C4 differ only on the geometry around Pt ion. Also, data presented in Table 3 shows that the similarity of C1 and C3 complexes is comparable to the similarity of C2 and C4 complexes. This is also the case with similarities between C1-C4 and C2-C3 compounds. Namely, the data shows that C1-C4 and C2-C3 similarity values are the lowest. This is due to the structural differences in the bridge between two metal atoms and the $cis-trans$ geometry around the Pt ion. Finally, these results show that Morgan circular fingerprints are capable of encoding small structural differences like $cis-trans$ geometry.

In Tables 4 and 5 structural similarity values of our compounds are presented, which are calculated using RDKit fingerprints and MACCS keys as molecular structure descriptors. Also, Tables 2S and 3S show correlation coefficients between similarity results. It is evident that the similarity results obtained by these two descriptors are highly correlated. Namely, the correlation coefficient in most cases is 1. However, Tables 4 and 5 reveal an important finding. Namely, RDKit fingerprints and MACCS keys were unable to distinguish between $cis-trans$ isomers. Namely, these descriptors yielded 100% similarity for C1-C2 and C3-C4 pairs. For this reason, results obtained by RDKit fingerprints and MACCS

keys will not be further used and discussed. We focus on the structural similarity results calculated by the Morgan circular fingerprints.

Table 4. The percentage of structural similarity of complexes C1-C4 calculated using RDKit fingerprints. The s denotes the standard deviation of data.

	RR	SM	Ja	JT	RT	Gle	SS_1	CT_3	CT_4	Sor
C1-C2	71	100	100	100	100	100	100	95.51	100	100
C1-C3	66.5	87.06	93.91	83.71	77.09	91.13	71.99	94.65	97.6	83.12
C1-C4	66.5	87.06	93.91	83.71	77.09	91.13	71.99	94.65	97.6	83.12
C2-C3	66.5	87.06	93.91	83.71	77.09	91.13	71.99	94.65	97.6	83.12
C2-C4	66.5	87.06	93.91	83.71	77.09	91.13	71.99	94.65	97.6	83.12
C3-C4	74.95	100	100	100	100	100	100	96.22	100	100
s	3.26	6.10	2.87	7.68	10.80	4.18	13.20	0.61	1.13	7.96

Table 5. The percentage of structural similarity of complexes C1-C4 calculated using MACCS keys. The s denotes the standard deviation of data.

	RR	SM	Ja	JT	RT	Gle	SS_1	CT_3	CT_4	Sor
C1-C2	28.14	100	100	100	100	100	100	75.55	100	100
C1-C3	27.54	97.01	96.5	90.2	94.19	94.85	82.14	75.14	97.44	90.04
C1-C4	27.54	97.01	96.5	90.2	94.19	94.85	82.14	75.14	97.44	90.04
C2-C3	27.54	97.01	96.5	90.2	94.19	94.85	82.14	75.14	97.44	90.04
C2-C4	27.54	97.01	96.5	90.2	94.19	94.85	82.14	75.14	97.44	90.04
C3-C4	29.94	100	100	100	100	100	100	76.73	100	100
s	0.88	1.41	1.65	4.62	2.74	2.43	8.42	0.58	1.21	4.70

By using physiological binary vectors and the novel methodology introduced in [13], we have calculated the physiological similarity of our compounds. The results are presented in Table 6. At first glance, one may see that the percentages of physiological similarity obtained by ten similarity measures are lower compared to values of structural similarity computed from Morgan circular fingerprints. For example, the JT index shows that C1-C2 are almost 89% structurally similar, while the same index yields only 12.5% of physiological similarity for the same pair of molecules. Such discrepancy may be attributed to the employment of different methodologies of molecular information embedding for similarity calculation. Namely, the physiological binary vectors are much shorter (only 60 bits long) and denser compared to long (1024 bits) sparse Morgan

circular fingerprints. Also, the standard deviation of physiological similarity results is smaller in comparison to the deviation of structural similarity values. Moreover, the correlation coefficients from Table 4S in Supporting Information show that all similarity indices yielded highly correlated data.

Table 6. The percentage of physiological similarity of complexes C1-C4 calculated using physiological fingerprints derived from the rate constants of the reactions between heterometallic complexes and biomolecules. The s denotes the standard deviation of data.

	RR	SM	Ja	JT	RT	Gle	SS_1	CT_3	CT_4	Sor
C1-C2	6.67	53.33	30	12.5	36.36	22.22	6.67	39.15	46.03	4.95
C1-C3	8.33	63.33	40.54	18.52	46.34	31.25	10.2	43.59	53.77	9.8
C1-C4	13.33	61.67	51.06	25.81	44.58	41.03	14.81	53.45	63.4	17.11
C2-C3	11.67	66.67	51.22	25.93	50	41.18	14.89	50.58	62.4	17.19
C2-C4	16.67	65	58.82	32.26	48.15	48.78	19.23	58.33	69.19	23.92
C3-C4	13.33	65	53.33	27.59	48.15	43.24	16	53.45	64.6	19.39
s	3.33	4.38	9.52	6.46	4.46	8.73	4.09	6.48	7.71	6.26
SRD	2	6	0	0	6	0	0	2	2	0

Data for all similarity indices show that C2 and C4 molecules are physiologically the most similar complexes. Also, C3 and C4 exhibit high physiological similarity. However, the results show that C1 and C2 are the least similar molecules. This is interesting since C1 and C2 structurally differ only on $cis - trans$ geometry, which is the case also for C3 and C4, however, physiologically they are very different.

To evaluate physiological similarity results we have performed the SRD analysis. As already mentioned, this statistical procedure enables us to compare similarity results computed by different coefficients. To rank indices the average value for all ten similarity measures has been used as an "ideal" standard for each compound since the reference value is not available. In Table 6 are given SRD values for all ten similarity indices. For the Ja, JT, Gle, SS_1, and Sor indices the SRD is 0. This shows that these measures give statistically valuable results in comparing two physiological binary vectors since their SRD value is practically equal to the golden standard. Moreover, this finding reveals that these indices operate with very similar performance. Another group of similar indices are RR, CT_3, and CT_4 measures. Their calculated SRD value is 2. This is very close to

0, which shows that these indices are very good in physiological similarity estimation, i.e., in comparing two physiological binary vectors. Slightly less favorable results are those obtained by SM and RT. For these indices, the SRD value is 6.

To validate the SRD procedure and its conclusions, another statistical technique has been carried out. Namely, the comparison of ranks with random numbers (CRRN) analysis has been employed. The CRRN is a randomization test that gives a distribution of the SRD values with randomized ranks. If the coefficient is statistically not distinguishable from randomly assigned ranks, it will overlap with random numbers. Another more favorable scenario is when the coefficient is statistically distinguishable from randomly assigned ranks and does not overlap with the distribution of random numbers. To perform CRNN analysis, the SRD values should be scaled, i.e., put on the same scale as random numbers. The results of CRRN for our similarity indices are depicted in Figure 2.

As one may see, similarity indices are grouped into three groups. The red and the blue groups of similarity indices are far from randomly assigned ranks, and practically there is no overlap with their distribution. This result confirms the SRD analysis for these indices and their use in physiological similarity assessment. Namely, it can be concluded that the probability that these variables are random is negligible. However, for the third group of indices (green), i.e., for the SM and RT, there is a slight overlap with the random SRD distribution. For this reason, results obtained by these indices should be taken with caution. Finally, we can conclude that the physiological similarity results obtained by most of the indices are statistically valuable, and the performance of Ja, JT, Gle, SS_1, and Sor indices is of very high quality since their SRD value is 0 and their results may be used for further similarity analysis.

Since structural and physiological similarities are obtained by employing different molecular information embedding techniques, to enable mutual comparison of these results, values should be scaled. To obtain relative similarity values, we have used equation 11. In this way, each value is positioned relatively compared to the maximum value within the set.

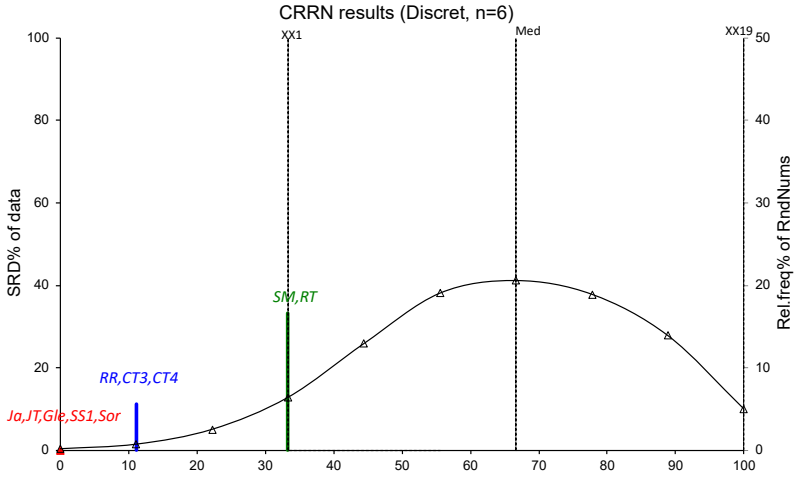

Figure 2. The Comparison of Ranks with Random Numbers (CRRN) analysis of physiological similarity results. X and left Y axes: The percentage of the scaled SRD for similarity coefficients (scaled between 0 and 100, i.e., put on the same scale as the random numbers). The scaled SRD for Ja, JT, Gle, SS_1, and Sor values is practically 0 (red), for RR, CT_3, and CT_4 data is 11.11% (blue), and for SM and RT is 33.33% (green). Right Y axis: The frequencies of random SRD are plotted (the black curve corresponds to random SRD distribution).

$$x' = \frac{x}{max(x)} \tag{11}$$

Now, we can compare the structural and physiological similarities and quantitatively determine the agreement of these two aspects of molecular similarity. In Table 7 are presented the results of the coherence between structural and physiological similarities of the corresponding pair of molecules for all ten similarity measures. The ideal coherence between structural and physiological similarity is 100%, which was not obtained for any pair of molecules. This finding is actually in accordance with the experimental observations, where structurally similar molecules show deviations in their activities. For example, all indices have found that for the C1-C2 pair of molecules, there is the highest disagreement between structural and physiological similarity. Namely, they are structurally very

similar but differ in physiological response. The *JT* for this pair of compounds shows agreement between structural and physiological aspects of approximately 39%. On the other hand, the C2-C4 pair of molecules shows lower structural similarity compared to their high physiological similarity.

Table 7. The percentage of coherence between relative physiological and relative structural (obtained from Morgan fingerprints) similarity of our compounds.

	RR	*SM*	*Ja*	*JT*	*RT*	*Gle*	SS_1	CT_3	CT_4	*Sor*
C1-C2	40.0	80.0	51.0	38.7	72.7	45.6	34.7	67.1	66.5	20.7
C1-C3	69.6	96.7	82.5	87.7	95.9	82.8	96.8	80.3	86.5	74.9
C1-C4	96.2	94.6	96.2	83.8	93.2	92.8	72.6	98.6	97.6	87.6
C2-C3	93.8	97.9	95.9	83.4	96.0	92.4	72.2	93.6	99.0	87.2
C2-C4	80.4	99.2	86.4	69.7	99.6	81.3	56.2	94.5	91.3	66.1
C3-C4	82.2	97.5	90.8	85.8	96.3	88.8	83.7	92.2	93.4	81.3

To get better insight into these results, values obtained by the *JT* index are depicted in Figure 3. As can be seen, the highest coherence is observed for C1-C3 and C3-C4 pairs of molecules. It is interesting to note that C1-C4, C2-C3, and C2-C4 sets of molecules show that they are physiologically more similar than structurally. Such results indicate that the bridge, i.e., its structure and length, which connects two metal ions is not of high importance in expressing the physiological activity of these molecules. In Figure 4 are depicted density plots of relative structural similarity values and relative physiological similarity values both calculated by the *JT* index. As one may see, structural similarity values vary in a narrower range compared to physiological similarity values. Generally, there is a high percentage of overlap between these two distributions.

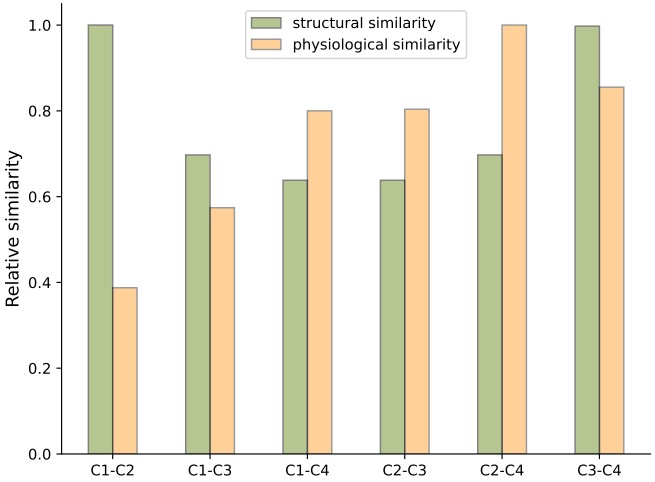

Figure 3. The coherence between relative physiological (obtained from physiological binary vectors) and relative structural (obtained from Morgan fingerprints) similarity of our compounds calculated by JT index.

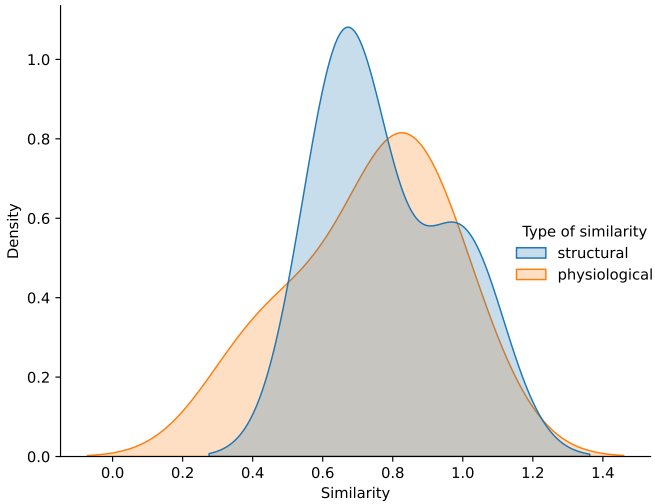

Figure 4. The density plots of the relative structural similarity and relative physiological similarity, both calculated by JT index.

4 Conclusion

In this work, four bridged heterometallic complexes (C1-C4) have been investigated. Namely, their structural and physiological similarities have been examined by using structural fingerprints and physiological binary vectors, respectively. It has been found that the Morgan circular fingerprints are convenient for similarity calculations in the chemistry of metal complexes. However, RDKit fingerprints and MACCS keys are not suitable since they do not recognize $cis - trans$ isomerism. The physiological similarity of our compounds has been assessed by binary vectors derived from rate constants. This approach enabled us to quantitatively determine physiological similarity. The obtained results have been validated by the SRD analysis. Moreover, the coherence between structural and physiological similarity has been quantitatively determined. Namely, it has been found that the C1-C2 pair of molecules has the highest discrepancy between these two aspects of molecular similarity, while the highest agreement is observed for C1-C3 and C3-C4 pairs. Also, it has been found that the bridge connecting two metals has a low influence on the physiological activity of our compounds.

***Acknowledgment*:** The authors acknowledge financial support from the State University of Novi Pazar and the Ministry of Science, Technological Development, and Innovation of the Republic of Serbia (Grant No. 451-03-47/2023-01/200252).

References

[1] M. M. Lorion, K. Maindan, A. R. Kapdi, L. Ackermann, Heteromultimetallic catalysis for sustainable organic syntheses, *Chem. Soc. Rev.* **46** (2017) 7399–7420.

[2] L. Bayer, B. S. Birenheide, F. Krämer, S. Lebedkin, F. Breher, Heterobimetallic gold/ruthenium complexes synthesized via post-functionalization and applied in dual photoredox gold catalysis, *Chem. Eur. J.* **28** (2022) #e20220185.

[3] T. V. Soldatović, E. Selimović, N. Milivojević, M. Jovanović, B. Šmit, Novel heteronuclear Pt(II)-L-Zn(II) complexes: synthesis, in-

teractions with biomolecules, cytotoxic properties. Two metals give promising antitumor activity?, *Appl. Organomet. Chem.* **34** (2020) #e5864.

[4] A. Halilagić, E. Selimović, J. S. Katanić Stanković, N. Srećković, K. Virijević, M. N. Živanović, B. Šmit, T. V. Soldatović, Novel heterometallic Zn(II)-L-Cu(II) complexes: studies of the nucleophilic substitution reactions, antimicrobial, redox and cytotoxic activity, *J. Coord. Chem.* **75** (2022) 472–492.

[5] T. V. Soldatović, *Application of the Principle of Hard and Soft Acids and Bases to Mechanisms of Bioinorganic Reactions*, Livre de Lyon Publisher, Lyon, 2019.

[6] T. V. Soldatović, B. Šmit, E. M. Mrkalić, S. Lj. Matić, R. M. Jelić, M. Ćendić Serafinović, N. Gligorijević, M. Čavić, S. Arandjelović, S. Grgurić-Šipka, Exploring heterometallic bridged Pt(II)-Zn(II) complexes as potential antitumor agents, *J. Inorg. Biochem.* **240** (2023) 112100.

[7] S. Mandal, M. Moudgil, S. Mandal, Rational drug design, *Eur. J. Pharmacol.* **625** (2009) 90–100.

[8] G. Maggiora, M. Vogt, D. Stumpfe, J. Bajorath, Molecular similarity in medicinal chemistry, *J. Med. Chem.* **57** (2014) 3186–3204.

[9] A. Bender, R. C. Glen, Molecular similarity: a key technique in molecular informatics, *Org. Biomol. Chem.* **2** (2004) 3204–3218.

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

[11] C. Cai, J. Gong, X. Liu, D. Gao, H. Li, Molecular similarity: methods and performance, *Chin. J. Chem.* **31** (2013) 1123–1132.

[12] S. Gurunathan, T. Yogalakshmi, K. Balasubramanian, Topological characterization of statistically clustered networks for molecular similarity analysis, *J. Math. Chem.* **61** (2023) 859–876.

[13] I. Redžepović, B. Furtula, Chemical similarity of molecules with physiological response, *Mol. Divers.* **27** (2023) 1603–1612.

[14] M. D. Krasowski, A. F. Pizon, M. G. Siam, S. Giannoutsos, M. Iyer, S. Ekins, Using molecular similarity to highlight the challenges of routine immunoassay-based drug of abuse/toxicology screening in emergency medicine, *BMC Emerg. Med.* **9** (2009) #5.

[15] R. L. Martin, T. F. Willems, L.-C. Lin, J. Kim, J. A. Swisher, B. Smit, M. Haranczyk, Similarity-driven discovery of zeolite materials for adsorption-based separations, *ChemPhysChem* **13** (2012) 3595–3597.

[16] C. W. Coley, L. Rogers, W. H. Green, K. F. Jensen, Computer-assisted retrosynthesis based on molecular similarity, *ACS Cent. Sci.* **3** (2017) 1237–1245.

[17] Y. Liu, Y. Cao, W. Lai, T. Yu, Y. Ma, Z. Ge, A strategy for predicting the crystal structure of energetic N-oxides based on molecular similarity and electrostatic matching, *CrystEngComm* **23** (2021) 714–723.

[18] M. A. Johnson, G. M. Maggiora, *Concepts and Applications of Molecular Similarity*, Wiley, New York, 1990.

[19] G. M. Maggiora, On outliers and activity cliffs - why QSAR often disappoints, *J. Chem. Inf. Model.* **46** (2006) 1535.

[20] R. Guha, J. H Van Drie, Structure-activity landscape index: identifying and quantifying activity cliffs, *J. Chem. Inf. Model.* **48** (2008) 646–658.

[21] D. Stumpfe, J. Bajorath, Exploring activity cliffs in medicinal chemistry: miniperspective, *J. Chem. Inf. Model.* **55** (2012) 2932–2942.

[22] J. L. Medina-Franco, Activity cliffs: facts or artifacts? *Chem. Biol. Drug Des.* **81** (2013) 553–556.

[23] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley, Weinheim, 2000.

[24] L. Xue, J. Bajorath, Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual Screening, *J. Comb. Chem. High Throughput Screen.* **3** (2000) 363–372.

[25] C. H. Yap, PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints, *J. Comput. Chem.* **32** (2011) 1466–1474.

[26] A. Fernandez-Torras, A. Comajuncosa-Creus, M. Duran-Frigola, P. Aloy, Connecting chemistry and biology through molecular descriptors, *Curr. Opin. Chem. Biol.* **66** (2022) #102090.

[27] N. M. O'Boyle, R. A. Sayle, Comparing structural fingerprints using a literature-based similarity benchmark, *J. Cheminf.* **8** (2016) #36.

[28] I. Muegge, P. Mukherjee, An overview of molecular fingerprint similarity search in virtual screening, *Expert Opin. Drug Discov.* **11** (2016) 137–148.

[29] R. Todeschini, V. Consonni, H. Xiang, J. Holliday, M. Buscema, P. Willett, Similarity coefficients for binary chemoinformatics data: overview and extended comparison using simulated and real data sets, *J. Chem. Inf. Model.* **52** (2012) 2884–2901.

[30] P. Willett, Combination of similarity rankings using data fusion, *J. Chem. Inf. Model.* **53** (2013) 1–10.

[31] R. A. Miranda-Quintana, D. Bajusz, A. Rácz, K. Héberger, Differential consistency analysis: which similarity measures can be applied in drug discovery?, *Mol. Inf.* **40** (2021) #2060017.

[32] D. Bajusz, A. Rácz, K. Héberger, Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?, *J. Cheminf.* **7** (2015) #20.

[33] A. Rácz, D. Bajusz, K. Héberger, Life beyond the Tanimoto coefficient: similarity measures for interaction fingerprints, *J. Cheminf.* **10** (2018) #48.

[34] T. B. Dunn, E. López-López, T. D. Kim, J. L. Medina-Franco, R. A. Miranda-Quintana, Exploring activity landscapes with extended similarity: is Tanimoto enough?, *Mol. Inf.* **42** (2023) #2300056.

[35] D. Rogers, M. Hahn, Extended-connectivity fingerprints, *J. Chem. Inf. Model.* **50** (2010) 742–754.

[36] RDKit: Open-Source Cheminformatics Software, `https://www.rdkit.org/`

[37] J. L. Durant, B. A. Leland, D. R. Henry, J. G. Nourse, Reoptimization of MDL keys for use in drug discovery, *J. Chem. Inf. Model.* **42** (2002) 1273–1280.

[38] Z. Hubalek, Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation, *Biol. Rev.* **57** (1982) 669–689.

[39] V. Batagelj, M. Bren, Comparing resemblance measures, *J. Classif.* **12** (1995) 73–90.

[40] P. F. Russell, T. R. Rao, On habitat and association of species of Anopheline larvae in south-eastern Madras, *J. Malaria Institute India* **3** (1940) 153–178.

[41] R. R. Sokal, C. D. Michener, A statistical method for evaluating systematic relationships, *Univ. Kansas Sci. Bull.* **38** (1958) 1409–1438.

[42] P. Jaccard, The distribution of the flora of the Alpine zone, *New Phytol.* **11** (1912) 37–50.

[43] D. J. Rogers, T. T. Tanimoto, A computer program for classifying plants, *Science* **132** (1960) 1115–1118.

[44] H. Gleason, Some applications of the quadrat method, *Bull. Torrey Botan. Club* **47** (1920) 21–33.

[45] R. R. Sokal, P. H. A. Sneath, *Principles of Numerical Taxonomy*, Freeman, San Francisco, 1963.

[46] V. Consonni, R. Todeschini, New similarity coefficients for binary data, *MATCH Commun. Math. Comput. Chem.* **68** (2012) 581–592.

[47] T. Sorgenfrei, Molluscan assemblages from the marine middle Miocene of South Jutland and their environments, *Danmarks Geo. Undersøgelse Serie 2* **79** (1958) 403–408.

[48] K. Héberger, Sum of ranking differences compares methods or models fairly, *Trends Anal. Chem.* **29** (2010) 101–109.

[49] A. Rácz, D. Bajusz, K. Héberger, Consistency of QSAR models: Correct split of training and test sets, ranking of models and performance parameters, *SAR QSAR Env.. Res.* **26** (2015) 683–700.

[50] C. West, M. A. Khalikova, E. Lesellier, K. Héberger, Sum of ranking differences to rank stationary phases used in packed column supercritical fluid chromatography, *J. Chromatogr. A* **1409** (2015) 241–250.

[51] G. Vastag, S. Apostolov, N. Perišić-Janjić, B. Matijević, Multivariate analysis of chromatographic retention data and lipophilicity of phenylacetamide derivatives, *Anal. Chim. Acta* **767** (2013) 44–49.

[52] K. Héberger, K. Kollár-Hunek, Sum of ranking differences for method discrimination and its validation: comparison of ranks with random numbers, *J. Chemom.* **25** (2011) 151–158.