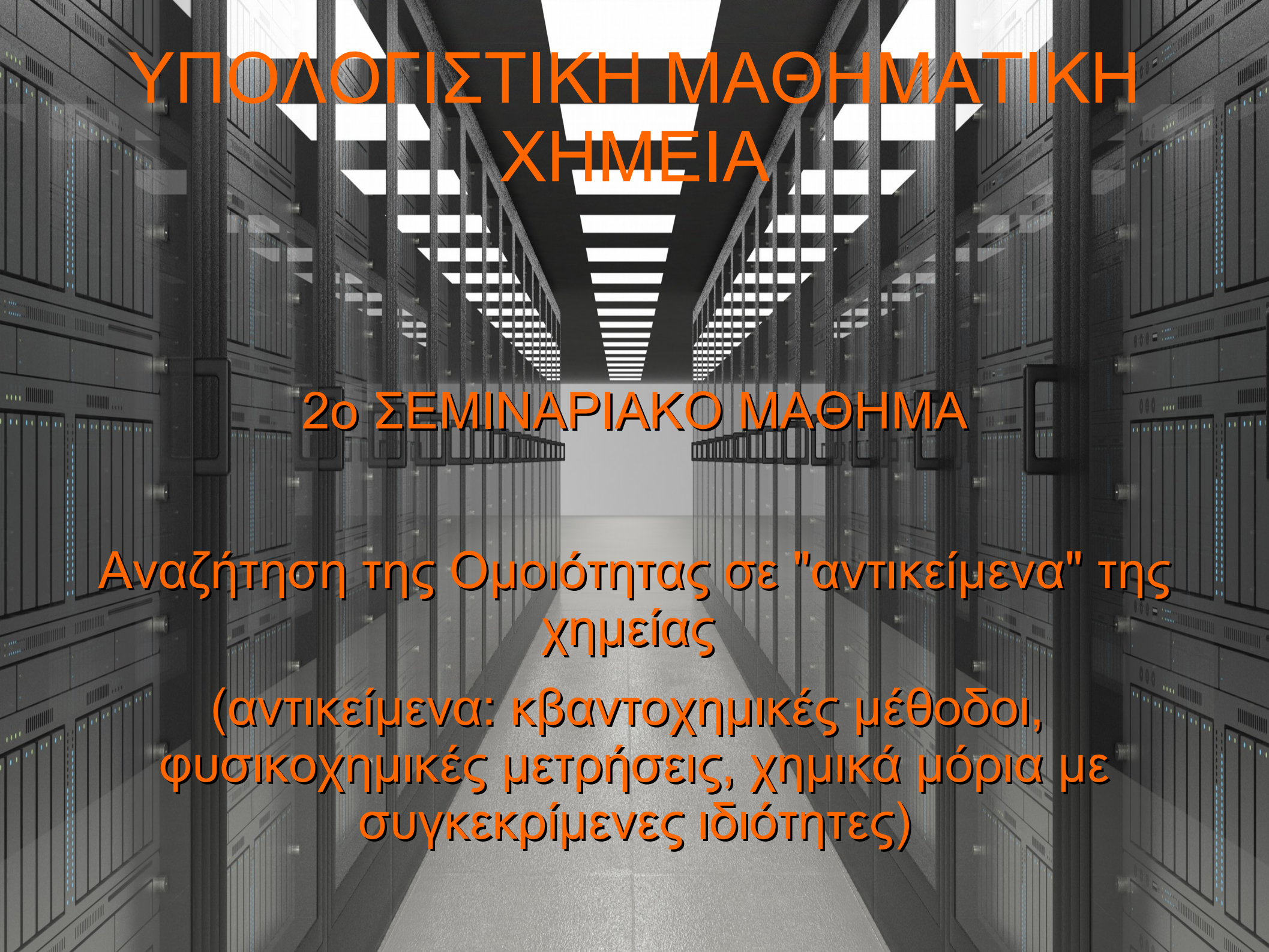


ΥΠΟΛΟΓΙΣΤΙΚΗ ΜΑΘΗΜΑΤΙΚΗ ΧΗΜΕΙΑ

2ο ΣΕΜΙΝΑΡΙΑΚΟ ΜΑΘΗΜΑ

Αναζήτηση της Ομοιότητας σε "αντικείμενα" της χημείας

(αντικείμενα: κβαντοχημικές μέθοδοι, φυσικοχημικές μετρήσεις, χημικά μόρια με συγκεκριμένες ιδιότητες)

Εισαγωγή…

Είναι γνωστό ότι στα ερωτήματα <<<είναι όμοια;>>> ή <<<με τι μοιάζει>>> η απάντηση σπάνια είναι μονοσήμαντη…

Εισαγωγή…

Είναι επίσης γνωστό, όμως, ότι υπάρχει μια αδιάκοπη προσπάθεια στο χώρο της επιστημής, προκειμένου να αντικατασταθεί ένα <<αντικείμενο>> με ένα άλλο ή μια διεργασία με μια άλλη, χωρίς ωστόσο να επηρεάζεται δραματικά η ακρίβεια και η αποτελεσματικότητα, καθώς και το κόστος.

Εισαγωγή…

Για τον σκοπό αυτό αναπτύχθηκαν ή εισήχθηκαν μαθηματικά εργαλεία ώστε μετά από μια επαναληπτική διαδικασία ελέγχου, να καταλήγουμε σε ένα <<αντικείμενο>> το οποίο ταιριάζει στο ζητούμενο pattern ή δείχνει την ίδια συμπεριφορά…

Έτσι προέκυψε ο κλάδος της

***ποσοτικής σχέσης δομής δραστικότητας/ιδιότητας (QSA(P)R)*.**

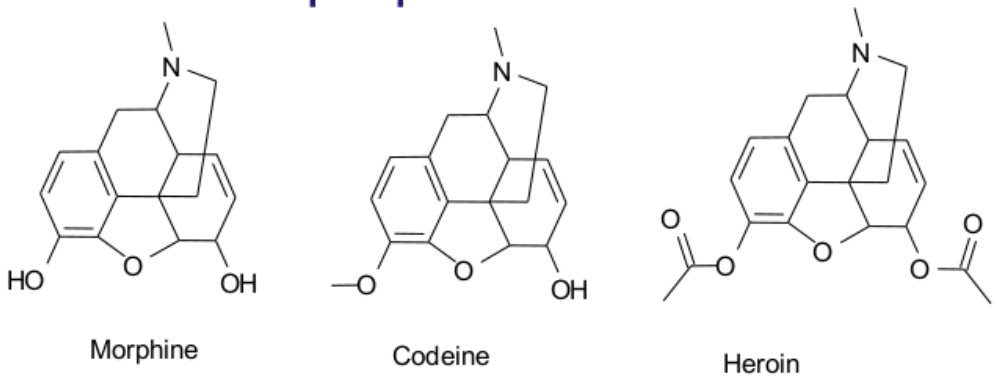
Εισαγωγή…

- **Θα προσπαθήσουμε να περιγράψουμε το μαθηματικό υπόβαθρο των διαθέσιμων μεθόδων αναζήτησης**
- **Θα συζητήσουμε την μετατροπή των δεδομένων σε πληροφορία και έπειτα σε γνώση**
- **Θα παρουσιάσουμε δεδομένα από διαφορετικούς τομείς της χημείας, π.χ:**
  - **κβαντοχημικές μέθοδοι**
  - **Φυσικοχημικές μέθοδοι**
  - **Μόρια με συγκεκριμένες ιδιότητες α) γλυκαντικές, β) ψυχομιμητικές-ψυχοκινητικές**

Θεωρία...

Why is molecular similarity important?

- Much of chemistry is based on structural analogies, and would be very difficult if this were not the case
- More formally, the *similar property principle* states that structurally similar molecules tend to have similar properties



Θεωρία...

Note that there are many exceptions to the principle but it is an excellent rule-of-thumb in the absence of more detailed knowledge

Θεωρία...

- **Molecular similarity encompasses**

  (i) bonding patterns,

  (ii) atomic positions,

  (iii) conformation,

  (iv) shape and

  (v) spatial disposition of molecular properties.

  The reverse of similarity is complementarity. In between lays molecular dissimilarity (or diversity). All notions of similarity involve perception of patterns and subsequent attempts to classify and quantify the patterns.

Θεωρία...

- **The similarity research methods can be divided into those which are based on**

  (i) discrete properties, and those which are founded on (ii) continuous properties.

  For example, a pharmacophore may be defined by three or more atoms separated by a set of distances; there may be some tolerances on the distances but the general pattern is discrete.

  A searching algorithm then finds only those structures from a database that satisfy the pharmacophore query. In a continuous space problem, the user may wish to optimize the overlap of a continuous field such as the electrostatic potential projected onto defined molecular surfaces. The result can then be expressed by a similarity coefficient.

Θεωρία...

- **Similarity searching in large chemical databases needs representations of the molecules that are both (i) effective, i.e., can differentiate between molecules that are different, and (ii) efficient, i.e., quick to calculate, in operation. There is a general conflict between these two requirements in that the most effective methods of representation tend to be the least efficient to calculate, and vice versa, so a suitable compromise needs to be made.**
- **For instance, quantum-mechanical descriptions, such as the electron probability density function described by Carbo and Calabuig[2], take too long to calculate whereas, at the other extreme, simple atom and bond counts are generally too trivial to discriminate between many molecules. In the middle lay descriptors based on 2D and 3D substructural fragments or properties**

Θεωρία...

- **The descriptors have to be encoded in order to enable similarity calculations to be carried out. The representation that is overwhelmingly used as a basis for similarity calculations in large databases is the fingerprint (the fixed-length bit-string)**

Θεωρία...

- **At the center of any similarity search is the measure that is used to quantify the degree of resemblance between the reference structure and each of the structures in the database (real or virtual) that is being screened.**
- **A similarity measure comprises three components:**
  - **the representation that is used to characterize the molecules that are being compared;**
  - **the weighting scheme that is used to assign differing degrees of importance to the various components of these representations;**
  - **and the coefficient that is used to determine the degree of relatedness between two structural representations.**

Θεωρία...

- **Some coefficients are measures of the distance, or dissimilarity between structures (and have a value of 0 for identical structures), while others measure similarity directly (and have their maximum value for identical structures).**
- **In most cases the values that can be taken by a coefficient lie in the range from 0 to 1 or can be normalized to that range: this is typically effected by means of a function based on the values of the attributes for the two structures that are being compared, with the resulting coefficients being referred to as association coefficients.**

Θεωρία...

- **The most intuitive approach is to identify the overlap between the graphs representing a pair of molecules**
  - **Such maximum common subgraph isomorphism methods are very slow**
- **Use of 2D fingerprints originally developed for substructure searching as an alternative**
  - **Binary vectors (or bit-strings) encoding chemical substructures (or fragments)**
  - **Currently, the standard way of computing molecular similarity (e.g., similarity searching, clustering and diversity analysis)**

Θεωρία...

- Weighted fingerprints associate a degree of relative importance with each bit in a fingerprint
  - Number of occurrences of a fragment in a molecule
  - Number of occurrences of a fragment in an entire database
- The former approach appears to be more useful, and can be more effective than binary fingerprints
- Much less studied to date than descriptors and coefficients

Θεωρία... Similarity coefficients

Tanimoto

- **A "similarity ratio" is given over bitmaps, where each bit of a fixed-size array represents the presence or absence of a characteristic in the plant being modelled. The definition of the ratio is the number of common bits, divided by the number of bits set (i.e. nonzero) in either sample.**

- **Presented in mathematical terms, if samples X and Y are bitmaps, X_i is the *ith* bit of X, and $\wedge,\vee$ are bitwise and, or operators respectively, then the similarity ratio**

$$T_s(X,Y) = \frac{\sum_i (X_i \wedge Y_i)}{\sum_i (X_i \vee Y_i)}$$

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. Thus, first we count the number (a) of "1s" in $\mathcal{A}$, then the number (b) of "1s" in $\mathcal{B}$ and finaly the number (c) of "1s" that are common to both. Then, the similarity $(S_{\mathcal{A},\mathcal{B}})$ between $\mathcal{A}$ and $\mathcal{B}$ is computed as:

$$S_{\mathcal{A},\mathcal{B}} = \frac{c}{a+b-c}$$

Θεωρία... Similarity coefficients

Sørensen–Dice

- **The Sørensen–Dice index, also known by other names, is a statistic used for comparing the similarity of two samples. It was independently developed by the botanists Thorvald Sørensen and Lee Raymond Dice,who published in 1948 and 1945 respectively. The Sørensen–Dice is also known as F1 score or Dice similarity coefficient (DSC).**

- **Sørensen's original formula was intended to be applied to presence/absence data, and is**

$$QS = \frac{2|X \cap Y|}{|X| + |Y|}$$

where |X| and |Y| are the numbers of elements in the two samples. Based on what is written here,

Θεωρία... Similarity coefficients

Sørensen–Dice

- When taken as a string similarity measure, the coefficient may be calculated for two strings, x and y using bigrams as follows:

$$s = \frac{2n_t}{n_x + n_y}$$

- where nt is the number of character bigrams found in both strings, nx is the number of bigrams in string x and ny is the number of bigrams in string y.

For example, to calculate the similarity between:

{night} and {nacht}

We would find the set of bigrams in each word: {ni,ig,gh,ht}, {na,ac,ch,ht}

Each set has four elements, and the intersection of these two sets has only one element: ht.

Inserting these numbers into the formula, we calculate, s = (2 · 1) / (4 + 4) = 0.25.

Θεωρία... Similarity coefficients

Tversky index

- The Tversky index, named after Amos Tversky, is an asymmetric similarity measure on sets that compares a variant to a prototype. The Tversky index can be seen as a generalization of Dice's coefficient and Tanimoto coefficient.

$$S(X,Y) = \frac{|X \cap Y|}{|X \cap Y| + \alpha|X - Y| + \beta|Y - X|}$$

- Here, X-Y denotes the relative complement of Y in X.
- Further, $\alpha,\beta \geq 0$ are parameters of the Tversky index. Setting $\alpha=\beta= 1$ produces the Tanimoto coefficient; $\alpha=\beta= 0.5$ produces Dices coefficient.
- If we consider X to be the prototype and Y to be the variant, then α corresponds to the weight of the prototype and β corresponds to the weight of the variant.
- Because of the inherent asymmetry, the Tversky index does not meet the criteria for a similarity metric. However, if symmetry is needed a variant of the original formulation has been proposed using max and min functions.

Θεωρία...

Hamming metric

- **In information theory, the Hamming distance between two strings of equal length is the number of positions at which the corresponding symbols are different. In other words, it measures the minimum number of substitutions required to change one string into the other, or the minimum number of errors that could have transformed one string into the other. In a more general context, the Hamming distance is one of several string metrics for measuring the edit distance between two sequences.**

- **A major application is in coding theory, more specifically to block codes, in which the equal-length strings are vectors over a finite field.**

Θεωρία...

Metric (mathematics)

Hamming metric

- **The Hamming distance between:**

- **"karolin" and "kathrin" is 3.**
- **"karolin" and "kerstin" is 3.**
- **1011101 and 1001001 is 2.**
- **2173896 and 2233796 is 3.**

Θεωρία...

Metric (mathematics)

Manhattan or Taxicab geometry,

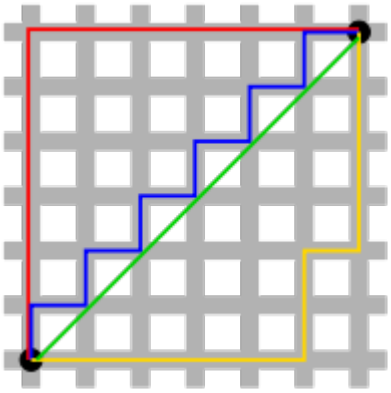
- **Considered by Hermann Minkowski in 19th-century Germany, is a form of geometry in which the usual distance function of metric or Euclidean geometry is replaced by a new metric in which the distance between two points is the sum of the absolute differences of their Cartesian coordinates.**
- **The taxicab metric is also known as rectilinear distance, L_1 distance or ℓ_1 norm (see Lp space), snake distance, city block distance, Manhattan distance or Manhattan length, with corresponding variations in the name of the geometry.**
- **The latter names allude to the grid layout of most streets on the island of Manhattan, which causes the shortest path a car could take between two intersections in the borough to have length equal to the intersections' distance in taxicab geometry.**

Θεωρία...

Metric (mathematics)

Manhattan or Taxicab geometry,

- **The latter names allude to the grid layout of most streets on the island of Manhattan, which causes the shortest path a car could take between two intersections in the borough to have length equal to the intersections' distance in taxicab geometry.**



Taxicab: τα μονοπάτια
κόκκινο, κίτρινο και μπλε έχουν
όλα μήκος 12.
Στην Ευκλείδια γεωμετρία, η πράσινη
γραμμή έχει μήκος 6 $\sqrt{2} \approx 8.49$, και είναι η μοναδική
σύντομη διαδρομή!

Θεωρία...

Metric (mathematics)

Manhattan or Taxicab geometry,

- **The taxicab distance, d_1, between two vectors p, q in an n-dimensional real vector space with fixed Cartesian coordinate system, is the sum of the lengths of the projections of the line segment between the points onto the coordinate axes. More formally,**

$$d_1(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum_{i=1}^{n} |p_i - q_i|,$$

$$\mathbf{p} = (p_1, p_2, \ldots, p_n) \text{ and } \mathbf{q} = (q_1, q_2, \ldots, q_n)$$

- **For example, in the plane, the taxicab distance between $p=(p_1, p_2)$, $q=(q_1, q_2)$ is $|p_1-q_1|+|p_2-q_2|$**

Θεωρία...

Metric (mathematics)

Mahalanobis distance,

- **The Mahalanobis distance is a measure of the distance between a point P and a distribution D, introduced by P. C. Mahalanobis in 1936. It is a multi-dimensional generalization of the idea of measuring how many standard deviations away P is from the mean of D.**
- **This distance is zero if P is at the mean of D, and grows as P moves away from the mean: along each principal component axis, it measures the number of standard deviations from P to the mean of D. If each of these axes is rescaled to have unit variance, then Mahalanobis distance corresponds to standard Euclidean distance in the transformed space. Mahalanobis distance is thus unitless and scale-invariant, and takes into account the correlations of the data set.**

Θεωρία...

Metric (mathematics)

Mahalanobis distance,

- **The Mahalanobis distance of an observation**

$$\vec{x} = (x_1, x_2, x_3, \ldots, x_N)^T$$

from a set of observations with mean

$$\vec{\mu} = (\mu_1, \mu_2, \mu_3, \ldots, \mu_N)^T$$

and covariance matrix S is defined as:

$$D_M(\vec{x}) = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})}.$$

Θεωρία...

Metric (mathematics)

Mahalanobis distance,

- **Consider the problem of estimating the probability that a test point in N-dimensional Euclidean space belongs to a set, where we are given sample points that definitely belong to that set. Our first step would be to find the average or center of mass of the sample points. Intuitively, the closer the point in question is to this center of mass, the more likely it is to belong to the set.**

- **However, we also need to know if the set is spread out over a large range or a small range, so that we can decide whether a given distance from the center is noteworthy or not. The simplistic approach is to estimate the standard deviation of the distances of the sample points from the center of mass. If the distance between the test point and the center of mass is less than one standard deviation, then we might conclude that it is highly probable that the test point belongs to the set. The further away it is, the more likely that the test point should not be classified as belonging to the set.**

Θεωρία...

Metric (mathematics)

Mahalanobis distance,

- **This intuitive approach can be made quantitative by defining the normalized distance between the test point and the set to be**

$$\frac{x - \mu}{\sigma}$$

By plugging this into the normal distribution we can derive the probability of the test point belonging to the set.

- **The drawback of the above approach was that we assumed that the sample points are distributed about the center of mass in a spherical manner. Were the distribution to be decidedly non-spherical, for instance ellipsoidal, then we would expect the probability of the test point belonging to the set to depend not only on the distance from the center of mass, but also on the direction. In those directions where the ellipsoid has a short axis the test point must be closer, while in those where the axis is long the test point can be further away from the center.**
-
- **Putting this on a mathematical basis, the ellipsoid that best represents the set's probability distribution can be estimated by building the covariance matrix of the samples. The Mahalanobis distance is the distance of the test point from the center of mass divided by the width of the ellipsoid in the direction of the test point**

Θεωρία...

Metric (mathematics)

Minkowski distance,

- Minkowski metric $d(i,j) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ (n equals to the dimensions of the space) has been implemented

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

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.

Θεωρία...

Single-linkage clustering,

- **In statistics, single-linkage clustering is one of several methods of hierarchical clustering. It is based on grouping clusters in bottom-up fashion (agglomerative clustering), at each step combining two clusters that contain the closest pair of elements not yet belonging to the same cluster as each other.**
- **A drawback of this method is that it tends to produce long thin clusters in which nearby elements of the same cluster have small distances, but elements at opposite ends of a cluster may be much farther from each other than to elements of other clusters. This may lead to difficulties in defining classes that could usefully subdivide the data.**

Θεωρία...

Clustering

Single-linkage clustering,

- **In the beginning of the agglomerative clustering process, each element is in a cluster of its own. The clusters are then sequentially combined into larger clusters, until all elements end up being in the same cluster. At each step, the two clusters separated by the shortest distance are combined. The definition of 'shortest distance' is what differentiates between the different agglomerative clustering methods.**

- **In single-linkage clustering, the distance between two clusters is determined by a single element pair, namely those two elements (one in each cluster) that are closest to each other. The shortest of these links that remains at any step causes the fusion of the two clusters whose elements are involved. The method is also known as nearest neighbour clustering. The result of the clustering can be visualized as a dendrogram, which shows the sequence of cluster fusion and the distance at which each fusion took place.**

Single-linkage clustering,

- **Mathematically, the linkage function – the distance D(X,Y) between clusters X and Y – is described by the expression**

$$D(X,Y) = \min_{x \in X, y \in Y} d(x,y),$$

where X and Y are any two sets of elements considered as clusters, and d(x,y) denotes the distance between the two elements x and y.

Θεωρία...

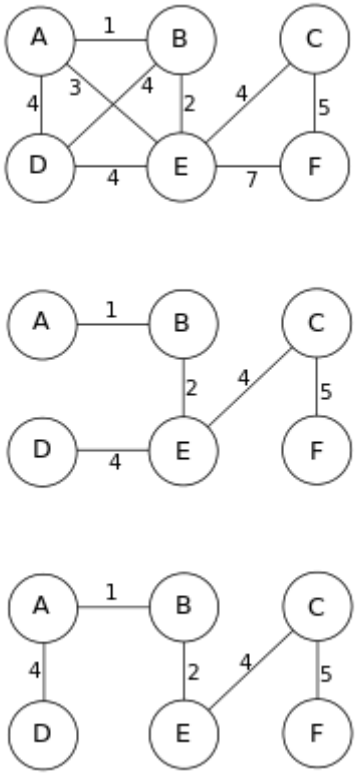
Minimum Spanning Tree,

Προκειμένου να ορίσουμε το MST θα ορίσουμε αρχικά το τι είναι γράφος: Γράφος G είναι ένα πεπερασμένο σύνολο από αντικείμενα (κορυφές, V) μαζί με ένα σύνολο από ζεύγη (ακμές, E) κορυφών του G. Το σύνολο των κορυφών είναι το V(G) και το σύνολο των ακμών είναι το E(G). Η ακμή e={u,v} ενώνει τις κορυφές u, v. Ένας γράφος είναι συνδεδεμένος όταν όλες οι κορυφές του είναι συνδεδεμένες σε ένα κοινό μονοπάτι. Ενώ ως δένδρο ορίζεται ένας μη κυκλικός συνδεδεμένος γράφος.

Θεωρία...

Clustering

Minimum Spanning Tree,



This figure shows
there may be more
than one minimum spanning tree in a graph.
In the figure, the two trees below the graph
are two possibilities of minimum
spanning tree of the given graph.

Θεωρία...

Prim's Algorithm,

Ενώ το MST είναι το δένδρο για το οποίο το άθροισμα των βαρών (δλδ τα μήκη) των ακμών είναι το ελάχιστο. Για την κατασκευή του θα χρησιμοποιήσουμε τον αλγόριθμο του Prime.

Ο αλγόριθμος αυτός κατασκευάζει μια κορυφή του δένδρου σε κάθε επανάληψη. Ξεκινά από μια οποιαδήποτε κορυφή (έστω Α), και βρίσκει την κορυφή (έστω Β) που απέχει μικρότερη απόσταση (μικρότερο κόστος) και είναι ενωμένη με την αρχική κορυφή.

Από την κορυφή Α ή Β, θα βρει την επόμενη κορυφή που απέχει την μικρότερη απόσταση, χωρίς να δημιουργείται κύκλος, και έτσι υλοποιείται... Κάποια στιγμή όλες οι κορυφές του γράφου θα έχουν συνδεθεί, και θα έχει παραχθεί το MST (δύο οι περισσότερες ακμές μπορεί να έχουν το ίδιο κόστος δλδ όμοιες κορυφές, τότε ο αλγόριθμος διαλέγει μία από τις δύο).

Prim’s Algorithm,

- **Παράδειγμα:**

https://www.youtube.com/watch?v=cplfcGZmX7I

Θεωρία...

Γενικά

Για να κατανοήσουμε την αντιστοιχία ή αναλογία της αναγνώρισης προτύπου με τα "αντικείμενα" της χημεία παραθέτουμε τον παρακάτω πίνακα

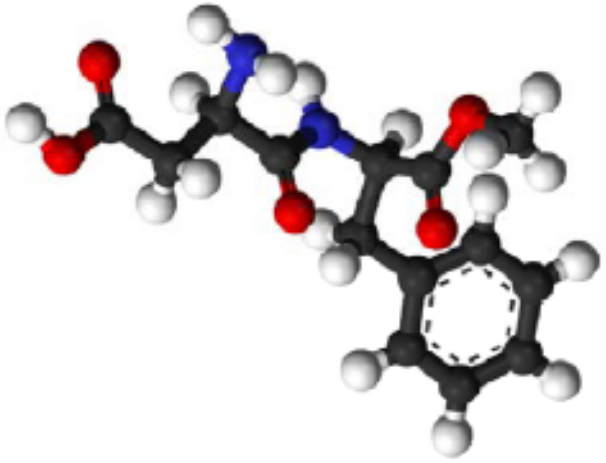
Αναζήτηση προτύπου	Χημεία
Στην αναζήτηση προτύπου έχουμε αντικείμενα	Στην έχουμε τις μεθόδους, ενώσεις, φυσικοχημικές μετρήσεις
Τα αντικείμενα έχουν χαρακτηριστικά	Με τις μεθόδους υπολογίζουμε τιμές μοριακών ιδιοτήτων, στις ενώσεις μας ενδιαφέρει η δραστικότητα
Πρότυπο είναι ένα σύνολο από χαρακτηριστικά	Μια θεωρητική περιγραφή είναι το σύνολο των τιμών ιδιοτήτων
Η ολότητα των προτύπων αποτελεί τον χώρο των προτύπων	Η ολότητα των θεωρητικών περιγραφών αποτελεί τον χώρο των θεωρητικών περιγραφών.

Παραδείγματα (QSPR)

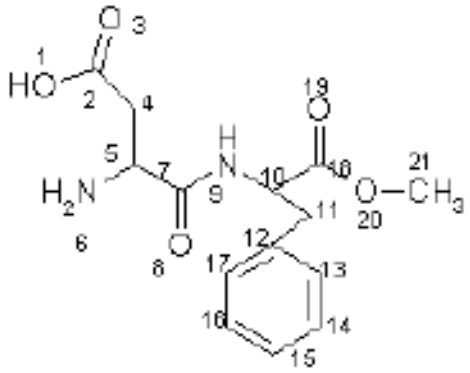
1ο Παράδειγμα από το χώρο της ομοιότητας „γλυκαντικών” ουσιών...

- Aspartame

Εικόνα 1: Τριασδιάστατη δομή της Ασπαρτάμης



Εικόνα 2: Μοριακός γράφος της Ασπαρτάμης



Παραδείγματα (QSPR)

Adjaceny matrix of Aspartame

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	0.296	0.704	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0.704	0	0.352	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0.352	0.296	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	1	0	0.817	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0.817	0.183	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	1	0	0	0.352	0.817	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0.352	0.296	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0.817	0	0.183	0.817	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0.817	0	1	0	0	0	0	0	0	1	0	0	0
11	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	1	0	0.667	0	0	0	0.667	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0.667	0	0.667	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0.667	0	0.667	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0.667	0	0.667	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.667	0	0.667	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0.667	0	0	0	0.667	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0.352	0.704	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.352	0.296	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.704	0	0.296	0.704
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.704	0

Παραδείγματα (QSPR)

Τιμές για όλα τα μόρια που συμμετείχαν στην μελέτη...

ONOMA	MT	MB	Tm	Density	Tr(ίχνος)	*N* =1-log(1/sweet)
Aspartame	$C_{14}H_{18}N_2O_5$	294.31	246.5	0.5	1.846	3.25527
D-Fructose	$C_6H_{12}O_6$	180.16	104	1.665	1.776	1.23045
Erythritol	$C_4H_{10}O_4$	122.12	121	1.45	1.184	0.84509
Glycerol	$C_3H_8O_3$	92.09382	18	1.261	0,888	0.77815
Lactitol	$C_{12}H_{24}O_{11}$	344.31	146	1.547	3.256	0.60206
Saccharin	$C_7H_5NO_3S$	183.1845	229.25	0.828	1.047	3.47712
Sorbitol	$C_6H_{14}O_6$	182.17	95	0.68	1.776	0.77815
Sucrose	$C_{12}H_{22}O_{11}$	342.29648	186	1.587	3.256	1
Xylitol	$C_5H_{12}O_5$	152.15	94	1.52	1.48	1.00000
Capsaicin	$C_{18}H_{27}NO_3$	305.41	63.5	1.041	0.296	-1
Salt	NaCl	58.442	801	2.16	0	0

Παραδείγματα (QSPR)

Πίνακας αποστάσεων για όλες τις ενώσεις...

	1	2	3	4	5	6	7	8	9	10	11
1	0	0.94345	1.02331	1.09014	0.99177	0.50330	0.71391	0.95090	0.96394	1.13680	1.74137
2	0.94345	0	0.31454	0.50078	0.75098	0.76281	0.60206	0.73765	0.16825	0.88924	1.19819
3	1.02331	0.31454	0	0.22313	1.00819	0.74332	0.54192	1.00651	0.15318	0.84936	1.07454
4	1.09014	0.50078	0.22313	0	1.16846	0.77985	0.55303	1.17560	0.33677	0.87668	1.18817
5	0.99177	0.75098	1.00819	1.16846	0	1.17867	0.89827	0.10555	0.87290	1.0378	1.68945
6	0.50330	0.76281	0.74332	0.77985	1.17867	0	0.67149	1.13504	0.73431	1.13898	1.44016
7	0.71391	0.60206	0.54192	0.55303	0.89827	0.67149	0	0.91371	0.52708	0.77403	1.45703
8	0.95090	0.73765	1.00651	1.17560	0.10555	1.13504	0.91371	0	0.86913	1.08412	1.66491
9	0.96394	0.16825	0.15318	0.33677	0.87290	0.73431	0.52708	0.86913	0	0.83904	1.15234
10	1.13680	0.88924	0.84936	0.87668	1.03780	1.13898	0.77403	1.08412	0.83904	0	1.46495
11	1.74137	1.19819	1.07454	1.18817	1.68945	1.44016	1.45703	1.66491	1.15234	1.46495	0

1: Aspartame, 2: D-Fructose, 3: Erythritol, 4: Glycerol, 5: Lactitol, 6: Saccharin, 7: Sorbitol, 8: Sucrose, 9: Xylitol, 10: Capsaicin, 11: Salt.

Παραδείγματα (QSPR)

Πίνακας αποστάσεων για όλες τις γλυκαντικές ενώσεις...

	1	2	3	4	5	6	7	8	9
1	0	1.44590	1.48782	1.72497	1.49887	0.63158	1.18506	1.39526	1.47271
2	1.44590	0	0.41595	0.74108	0.95185	1.23377	0.86094	0.97137	0.22757
3	1.48782	0.41595	0	0.50975	1.25210	1.18724	0.75467	1.27489	0.22422
4	1.72497	0.74108	0.50975	0	1.54202	1.41741	0.79400	1.61521	0.53404
5	1.49887	0.95185	1.25210	1.54202	0	1.67092	1.18798	0.22593	1.10202
6	0.63158	1.23377	1.18724	1.41741	1.67092	0	1.15645	1.57188	1.22233
7	1.18506	0.86094	0.75467	0.79400	1.18798	1.15645	0	1.25075	0.74542
8	1.39526	0.97137	1.27489	1.61521	0.22593	1.57188	1.25075	0	1.13854
9	1.47271	0.22757	0.22422	0.53404	1.10202	1.22233	0.74542	1.13854	0

1: Aspartame, 2: D-Fructose, 3: Erythritol, 4: Glycerol, 5: Lactitol, 6: Saccharin, 7: Sorbitol, 8: Sucrose, 9: Xylitol.

Παραδείγματα (QSPR)

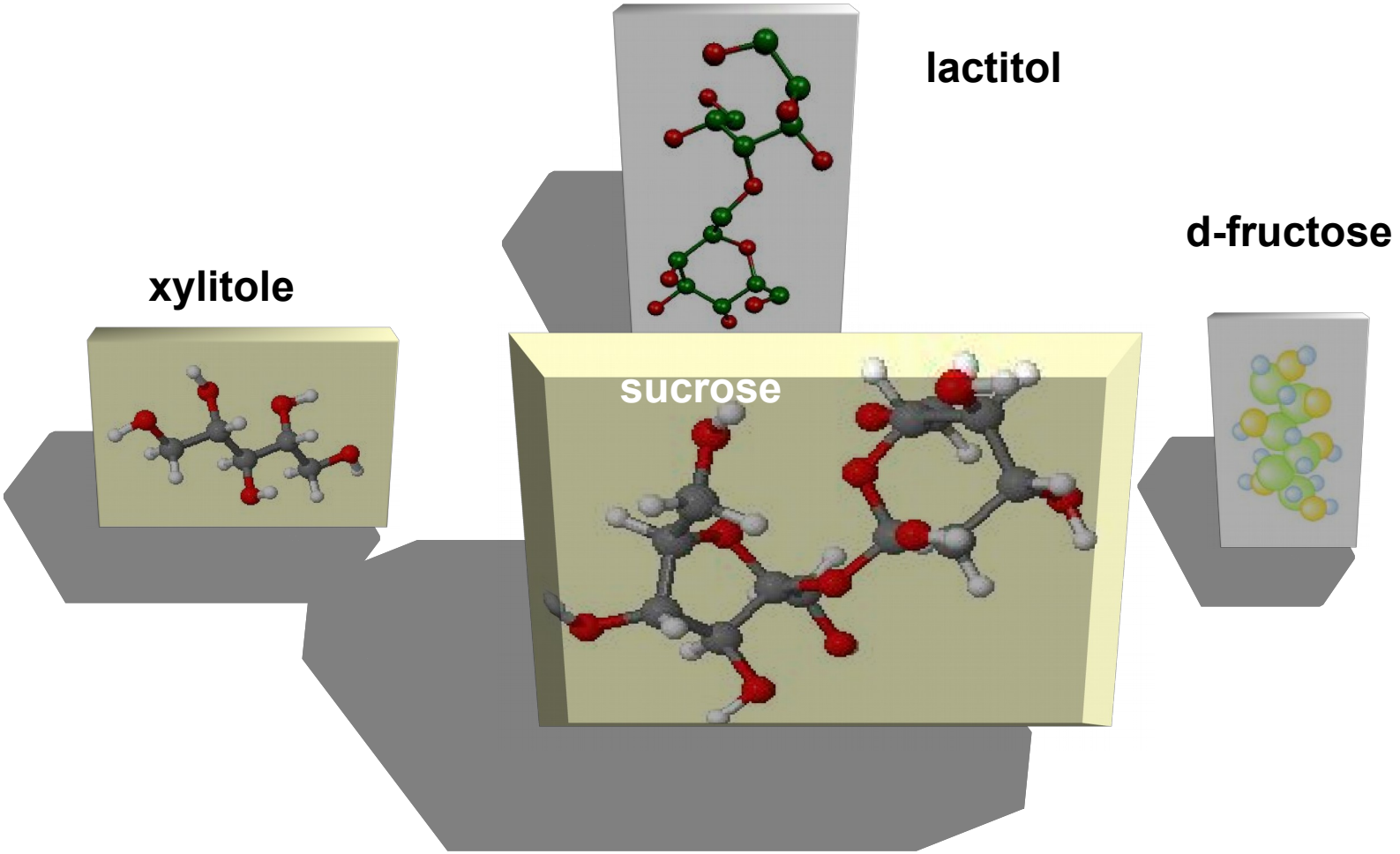
Πίνακας ομοιότητας για όλες τις γλυκαντικές ενώσεις...

	1	2	3	4	5	6	7	8	9
1	1	0.16178	0.13748	0	0.13107	0.63386	0.31300	0.19114	0.14624
2	0.16178	1	0.75887	0.57038	0.44819	0.28476	0.50089	0.43687	0.86807
3	0.13748	0.75887	1	0.70449	0.27413	0.31173	0.56250	0.26092	0.87001
4	0	0.57038	0.70449	1	0.10606	0.17830	0.53970	0.06363	0.69040
5	0.13107	0.44819	0.27413	0.10606	1	0.03133	0.31130	0.86902	0.36113
6	0.63386	0.28476	0.31173	0.17830	0.03133	1	0.32958	0.08875	0.29139
7	0.31300	0.50089	0.56250	0.53970	0.31130	0.32958	1	0.27491	0.56786
8	0.19114	0.43687	0.26092	0.06363	0.86902	0.08875	0.27491	1	0.33996
9	0.14624	0.86807	0.87001	0.69040	0.36113	0.29139	0.56786	0.33996	1

1: Aspartame, 2: D-Fructose, 3: Erythritol, 4: Glycerol, 5: Lactitol, 6: Saccharin, 7: Sorbitol, 8: Sucrose, 9: Xylitol.

Παραδείγματα (QSPR)

Αποτελέσματα



Παραδείγματα (QSPR)

Tanimoto Index vs. Euclidean Distance. What can they tell us on the Similarity of Analgesic and Euphoric Compounds?

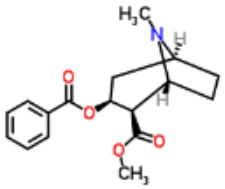
D. Xenides[1], D. Fostiropoulou and D. Vlachos
Laboratory of Mathematical Modeling and Complex Systems,
Department of Computer Science and Technology,
Faculty of Sciences and Technology,
University of Peloponnese,
GR-22100 Tripolis, Greece

Παραδείγματα (QSPR)

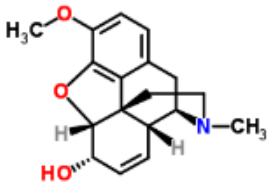
Table 1: Properties that construct the space of theoretical descriptions

	Polar Surface	Polarizability	Density (g/cm^3)	Flash Point	Boiling Point	Molecular Weight
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
Casein	314	80.7	1.4			783.91

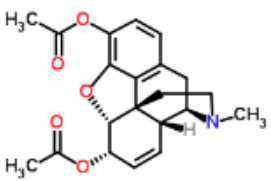
Παραδείγματα (QSPR)



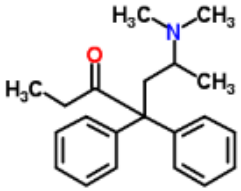
COCAINE



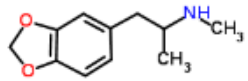
CODEINE



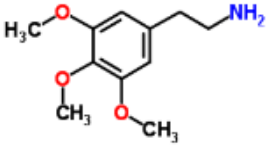
HEROINE



METHADONE

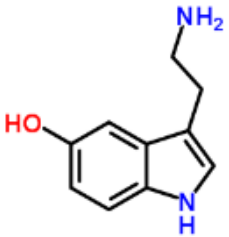


ECSTASY

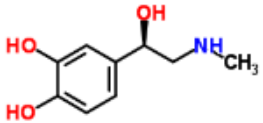


MESCALINE

CAFFEINE

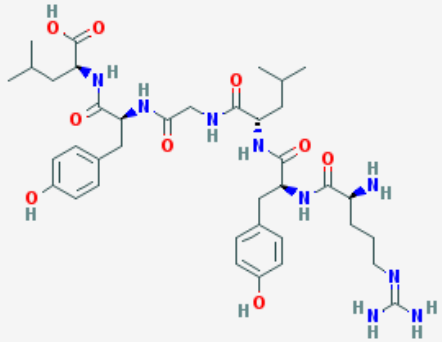


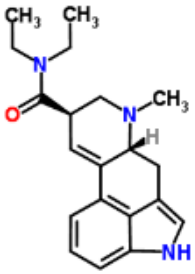
SEROTONINE



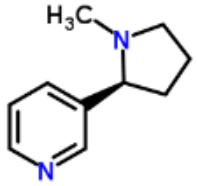
ADRENALINE

MORPHINE

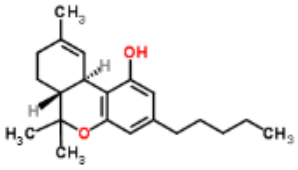




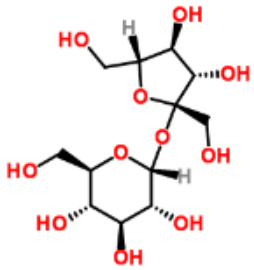
LSD



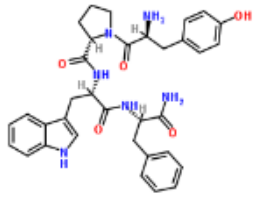
NICOTINE



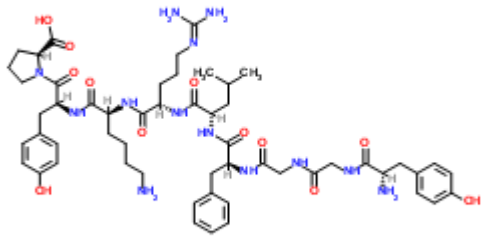
THC



SUCROSE



ENDORPHINE-1



Παραδείγματα (QSPR)

Euclidean

Sij	cocaine	codeine	heroine	methadone	ecstasy	mescaline	caffeine	serotonin	adrenaline	morphine	LSD	nicotine	THC	Sugar	Endorphin-1	Endorphin-2
cocaine	1.00000	0.91577	0.83784	0.83400	0.81328	0.85226	0.77088	0.82733	0.81824	0.87038	0.86372	0.73576	0.85321	0.46546	0.24519	0.34027
codeine	0.91577	1.00000	0.87875	0.79232	0.77321	0.79658	0.79966	0.82592	0.81105	0.93297	0.89853	0.69291	0.80345	0.49336	0.26531	0.36044
heroine	0.83784	0.87875	1.00000	0.71640	0.69061	0.69750	0.74234	0.73732	0.73225	0.88600	0.85784	0.58138	0.73062	0.56850	0.35715	0.45184
methadone	0.83400	0.79232	0.71640	1.00000	0.84582	0.79461	0.63603	0.70673	0.68981	0.72968	0.80302	0.75311	0.96372	0.31558	0.18571	0.27717
ecstasy	0.81328	0.77321	0.69061	0.84582	1.00000	0.82335	0.67693	0.74517	0.72448	0.72995	0.74689	0.78737	0.84903	0.32044	0.12570	0.21796
mescaline	0.85226	0.79658	0.69750	0.79461	0.82335	1.00000	0.76123	0.85651	0.84832	0.76862	0.74081	0.86144	0.80696	0.38501	0.12081	0.21521
caffeine	0.77088	0.79966	0.74234	0.63603	0.67693	0.76123	1.00000	0.88657	0.88200	0.84598	0.70811	0.67223	0.64358	0.53555	0.14982	0.24356
serotonin	0.82733	0.82592	0.73732	0.70673	0.74517	0.85651	0.88657	1.00000	0.96739	0.84017	0.74910	0.75566	0.71663	0.48768	0.15741	0.25183
adrenaline	0.81824	0.81105	0.73225	0.68981	0.72448	0.84832	0.88200	0.96739	1.00000	0.82962	0.73550	0.74078	0.70172	0.50540	0.16342	0.25758
morphine	0.87038	0.93297	0.88600	0.72968	0.72995	0.76862	0.84598	0.84017	0.82962	1.00000	0.84986	0.65851	0.74047	0.54747	0.26949	0.36446
LSD	0.86372	0.89853	0.85784	0.80302	0.74689	0.74081	0.70811	0.74910	0.73550	0.84986	1.00000	0.64585	0.81479	0.46743	0.32883	0.42255
nicotine	0.73576	0.69291	0.58138	0.75311	0.78737	0.86144	0.67223	0.75566	0.74078	0.65851	0.64585	1.00000	0.75261	0.25434	0.00000	0.09388
THC	0.85321	0.80345	0.73062	0.96372	0.84903	0.80696	0.64358	0.71663	0.70172	0.74047	0.81479	0.75261	1.00000	0.33257	0.20132	0.29311
sugar	0.46546	0.49336	0.56850	0.31558	0.32044	0.38501	0.53555	0.48768	0.50540	0.54747	0.46743	0.25434	0.33257	1.00000	0.40260	0.47168
Endorphine-1	0.24519	0.26531	0.35715	0.18571	0.12570	0.12081	0.14982	0.15741	0.16342	0.26949	0.32883	0.00000	0.20132	0.40260	1.00000	0.90449
Endorphine-2	0.34027	0.36044	0.45184	0.27717	0.21796	0.21521	0.24356	0.25183	0.25758	0.36446	0.42255	0.09388	0.29311	0.47168	0.90449	1.00000

1
2
3

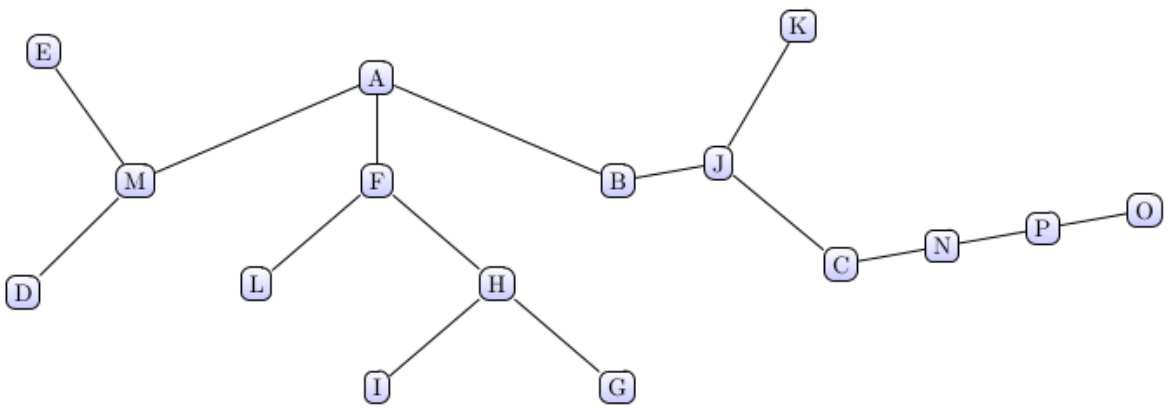
Παραδείγματα (QSPR)

Tanimoto

Sij	cocaine	codeine	heroine	methadone	ecstasy	mescaline	caffeine	serotonin	adrenaline	morphine	LSD	nicotine	THC	Sugar	Endomorphine-1	Endomorphine-2	Casein
cocaine	1.00000	0.24201	0.25322	0.32407	0.18519	0.18803	0.18657	0.16923	0.25455	0.24312	0.21256	0.21053	0.18023	0.23214	0.27950	0.30534	0.29932
codeine	0.24201	1.00000	0.90148	0.20305	0.28283	0.24599	0.12444	0.15493	0.23404	0.99454	0.27778	0.17512	0.44878	0.13171	0.21951	0.24413	0.25551
heroine	0.25322	0.90148	1.00000	0.20094	0.26267	0.22222	0.11885	0.14163	0.21154	0.89655	0.27178	0.17021	0.41518	0.13514	0.21212	0.23377	0.25000
methadone	0.32407	0.20305	0.20094	1.00000	0.27551	0.25301	0.12727	0.21875	0.28750	0.20408	0.21348	0.31250	0.18310	0.11957	0.30534	0.41667	0.34483
ecstasy	0.18519	0.28283	0.26267	0.27551	1.00000	0.59211	0.09302	0.31373	0.50000	0.27778	0.18878	0.21186	0.26712	0.12150	0.27891	0.35398	0.30075
mescaline	0.18803	0.24599	0.22222	0.25301	0.59211	1.00000	0.08108	0.39506	0.57143	0.24064	0.15217	0.18447	0.29134	0.07609	0.25564	0.33333	0.27731
caffeine	0.18657	0.12444	0.11885	0.12727	0.09302	0.08108	1.00000	0.16667	0.13333	0.12500	0.13171	0.14516	0.11515	0.12264	0.16149	0.12593	0.15436
serotonin	0.16923	0.15493	0.14163	0.21875	0.31373	0.39506	0.16667	1.00000	0.38272	0.15566	0.32353	0.33663	0.16447	0.05660	0.51261	0.28319	0.24060
adrenaline	0.25455	0.23404	0.21154	0.28750	0.50000	0.57143	0.13333	0.38272	1.00000	0.23529	0.19209	0.19802	0.22556	0.07692	0.28682	0.39362	0.36036
morphine	0.24312	0.99454	0.89655	0.20408	0.27778	0.24064	0.12500	0.15566	0.23529	1.00000	0.27881	0.17593	0.45098	0.13235	0.22041	0.24528	0.25664
LSD	0.21256	0.27778	0.27178	0.21348	0.18878	0.15217	0.13171	0.32353	0.19209	0.27881	1.00000	0.31461	0.18966	0.06566	0.40909	0.23858	0.23365
nicotine	0.21053	0.17512	0.17021	0.31250	0.21186	0.18447	0.14516	0.33663	0.19802	0.17593	0.31461	1.00000	0.17722	0.07080	0.36957	0.26230	0.23404
THC	0.18023	0.44878	0.41518	0.18310	0.26712	0.29134	0.11515	0.16447	0.22556	0.45098	0.18966	0.17722	1.00000	0.12414	0.16080	0.18072	0.18681
Sugar	0.23214	0.13171	0.13514	0.11957	0.12150	0.07609	0.12264	0.05660	0.07692	0.13235	0.06566	0.07080	0.12414	1.00000	0.07097	0.08264	0.11029
Endomo-1	0.27950	0.21951	0.21212	0.30534	0.27891	0.25564	0.16149	0.51261	0.28682	0.22041	0.40909	0.36957	0.16080	0.07097	1.00000	0.70086	0.58696
Endmor-2	0.30534	0.24413	0.23377	0.41667	0.35398	0.33333	0.12593	0.28319	0.39362	0.24528	0.23858	0.26230	0.18072	0.08264	0.70086	1.00000	0.78641
Casein	0.29932	0.25551	0.25000	0.34483	0.30075	0.27731	0.15436	0.24060	0.36036	0.25664	0.23365	0.23404	0.18681	0.11029	0.58696	0.78641	1.00000

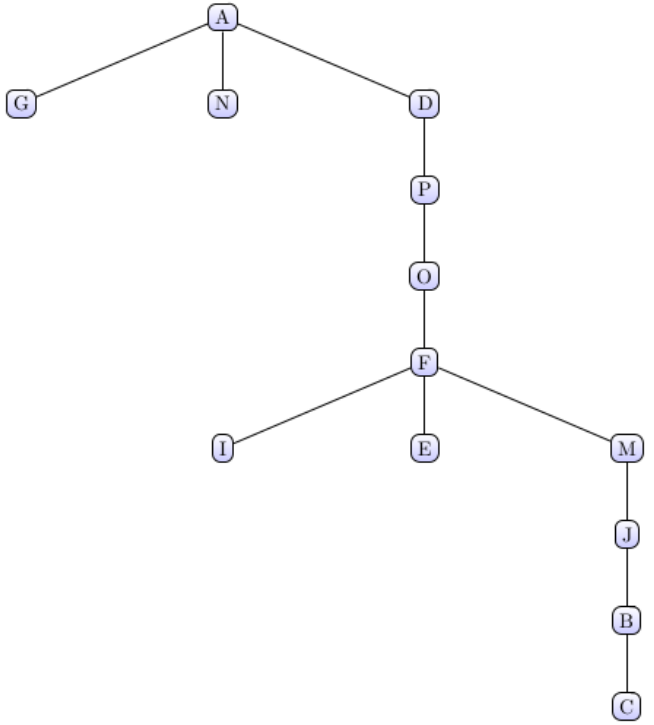
1
2
3

Παραδείγματα (QSPR)



Euclidean MST

cocaine
codeine
heroine
methadone
ecstasy
mescaline
caffeine
serotonin
adrenaline
morphine
LSD
nicotine
THC
Sugar
Endomo-1
Endmor-2



Tanimoto MST

Παραδείγματα (QSPR)

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Ταξινόμηση των πυραζολο[3,4,-1]πυριμιδινών ως πιθανών αντιπολλαπλασιαστικών και προαποπτοτικών παραγόντων έναντι των A431 και 8701-BC κυττάρων. Μελέτη βασιζόμενη στο χώρο των θεωρητικών περιγραφών των φυσικοχημικών ιδιοτήτων τους.

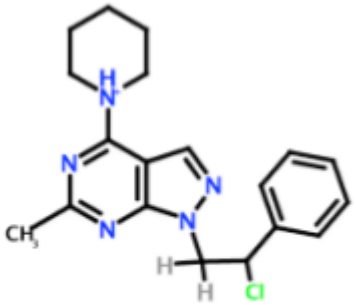
ΜΗΤΡΟΠΟΥΛΟΥ ΕΛΕΝΗ

Παραδείγματα (QSPR)

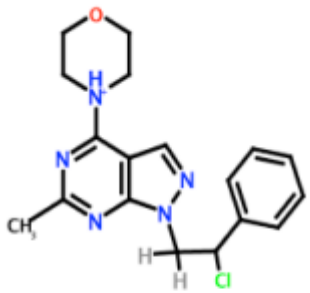
Στην παρούσα μελέτη εξετάζονται και συγκρίνονται χημικές ενώσεις οι οποίες ανήκουν στα παράγωγα των πυραζολο[3,4,-1]πυριμιδινών κι οι οποίες αποτρέπουν την ανάπτυξη καρκινικών κυττάρων παρεμβαίνοντας στη φωσφορυλίωση (διαδικασία κατα την οποία μία ή περισσότερες φωσφορικές ομάδες εισάγονται σ'ένα μόριο) του γονίδιου Src και δρουν ως προαποπτοτικοί παράγοντες με την παρεμπόδιση του αντιαποπτοτικού γονιδίου BCL2. Βιολογικές μελέτες έχουν δείξει οτι τέτοιες ενώσεις έχουν μία ισχυρή αντιπολλαπλασιαστική επίδραση στα ανθρώπινα επιδερμοειδή καρκινικά A431 κύτταρα. Αυτό σε συνδυασμό με το οτι τα καρκινικά 8701-BC κύτταρα του μαστού υπερκεφράζουν το γονίδιο Src έχει οδηγήσει στο να αξιολογηθούν οι αντιπολλαπλασιαστικές ιδιότητες των χημικών ενώσεων έναντι στα 8701-BC κύτταρα. Βιολογικά δεδομένα έχουν δείξει οτι οι παραπάνω χημικές ενώσεις δρουν ανασταλτικά στα 8701-BC κύτταρα με αξιοσημείωτη δύναμη. Συνεπώς, σύμφωνα με όλες αυτές τις βιολογικές ανακαλύψεις, τα παράγωγα των πυραζολο[3,4,-1]πυριμιδίνων θα μπορούσαν να είναι ένα χρήσιμο εργαλείο για την αναστολή της ανάπτυξης των καρκινικών κυττάρων.

Σκοπός αυτής της μελέτης είναι να συγκριθούν μεταξύ τους αυτά τα παράγωγα, να βρεθούν ομοιότητες ή διαφορές, να ταξινομηθούν ως προς την ικανοτητά τους να δρουν ανασταλτικά στην ανάπτυξη των καρκινικών κυττάρων όπως επίσης και να αξιολογηθεί εάν κάποιο από αυτά είναι πιο ισχυρό από τα υπόλοιπα.

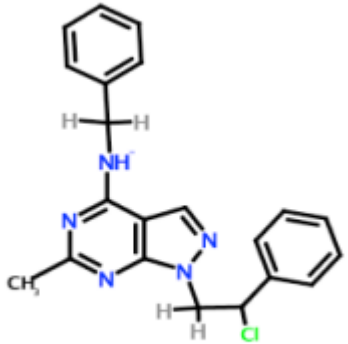
Παραδείγματα (QSPR)



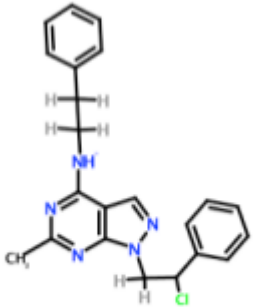
Εικόνα 2-1: Ένωση 1f



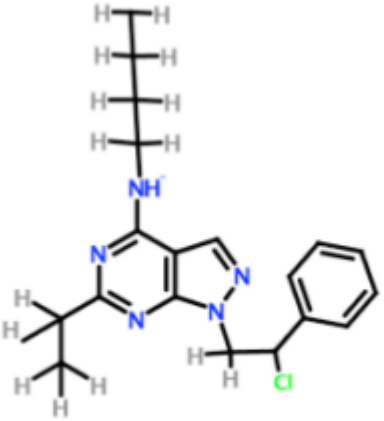
Εικόνα 2-2: Ένωση 1g



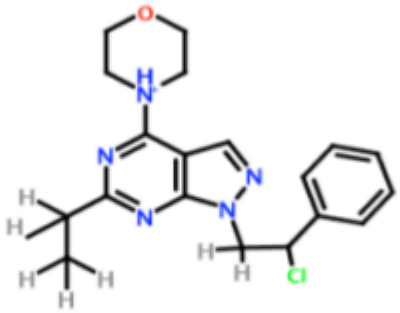
Εικόνα 2-3: Ένωση 1j



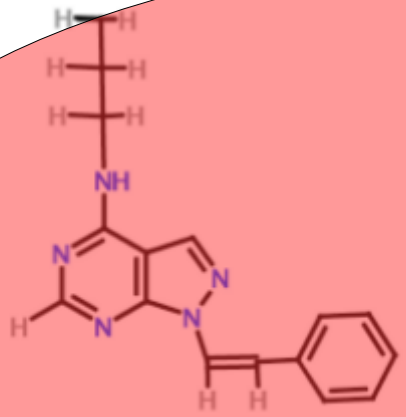
Εικόνα 2-4: Ένωση 1k



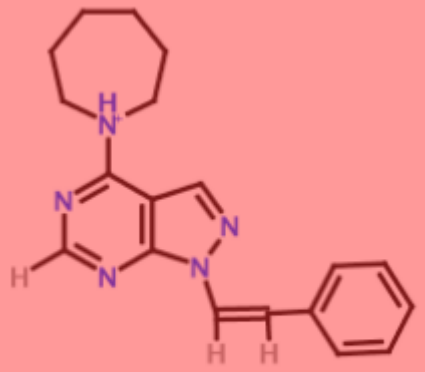
Εικόνα 2-5: Εικόνα 2b



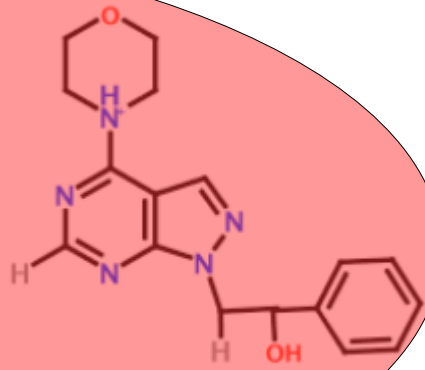
Εικόνα 2-6: Εικόνα 2e



Εικόνα 2-7: Ένωση 5a



Εικόνα 2-8: Ένωση 5i



Εικόνα 2-9: Ένωση 6e

Παραδείγματα (QSPR)

ONOMA	MB	MP(·C)	Ki(μM)	IC50(μM)	
				8701-BC	A431
1f	328	94-95	2.4±0.7	67.5±1.2	38.4±0.8
1g	330	116-117	6.5±1.0	64.2±1.8	24.3±0.4
1j	349	142-143	3.7±0.9	53.2±1.2	24.8±0.7
1k	363	73-74	0.7±0.2	31.2±0.5	18.2±0.9
2b	329	100-101	0.6±0.2	38.8±0.5	40.4±0.4
2e	344	130-131	3.6±0.9	79.8±3.0	42.3±0.7

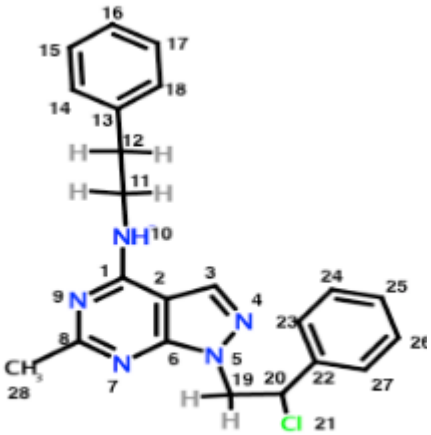
Πίνακας 3-1: Περιέχει τις τιμές από όλες τις φυσικοχημικές ιδιότητες για τις πρώτες 6 ενώσεις

ONOMA	MB	MP(·C)	A431	Tr(Ίχνος)
1f	328	94-95	38.4±0.8	1.011
1g	330	116-117	24.3±0.4	1.307
1j	349	142-143	24.8±0.7	1.011
1k	363	73-74	18.2±0.9	1.819
2b	329	100-101	40.4±0.4	0.915
2e	344	130-131	42.3±0.7	1.307
5a	251	189-190	61.3±1.0	0.915
5i	292	190-191	60.9±2.1	0.915
6e	298	179-180	84.2±3.7	1.507

Πίνακας 3-2: Περιέχει τις τιμές από κάποιες φυσικοχημικές ιδιότητες για όλες τις ενώσεις που αναφέρθηκαν

Παραδείγματα (QSPR)

Από τον μοριακό γράφο της ένωσης 1k προκύπτει ο παρακάτω πίνακας γειτνίασης:



Εικόνα 3-4: Μοριακός Γράφος 1k

	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
1	0	1.000	0	0	0	0	0	0	0.409	0.817	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1.000	0	1.000	0	0	0.500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1.000	0	0.409	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0.409	0.183	0.668	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0.668	0.183	0.817	0	0	0	0	0	0	0	0	0	0	0	0	0.817	0	0	0	0	0	0	0	0	0
6	0	0.500	0	0	0.817	0	0.817	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0.817	0.183	0.409	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0.409	0	0.817	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.000
9	0.409	0	0	0	0	0	0	0.817	0.183	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0.817	0	0	0	0	0	0	0	0	0.183	0.817	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0.817	0	1.000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	1.000	0	1.000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	1.000	0	1.000	0	0	0	0.500	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	1.000	0	0.500	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0.817	0	0	0	0.500	0	1.000	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.000	0	0.500	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.500	0	1.000	0.000	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0.500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0.817	0	0	0	0	0	0	0	0	0	0	0	1.000	0	0	1.000	0	0	0	1.000	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.000	0	0.904	1.000	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.904	0.904	0.000	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.500	0	0	0.500	0	0	0	1.000	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.500	0	1.000	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.000	0	0.500	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.500	0	1.000	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.000	0	0.500	0
27	0	0	0	0	0	0	0	0.000	0	0	0	0	0	0	0	0	0	0	0	0	0	1.000	0	0	0	0.500	0	0
28	0	0	0	0	0	0	0	1.000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Πίνακας 3-6: Πίνακας Γειτνίασης 1k

Το ίχνος του πίνακα: Tr(1k) = 1.819

Παραδείγματα (QSPR)

Συνεπώς, οι τελικοί πίνακες μετά την προσθήκη και του "ίχνους" γίνονται :

ONOMA	MB	MP(°C)	Ki(μM)	IC50(μM)		Tr(Ίχνος)
				8701-BC	A431	
1f	328	94-95	2.4±0.7	67.5±1.2	38.4±0.8	1.011
1g	330	116-117	6.5±1.0	64.2±1.8	24.3±0.4	1.307
1j	349	142-143	3.7±0.9	53.2±1.2	24.8±0.7	1.011
1k	363	73-74	0.7±0.2	31.2±0.5	18.2±0.9	1.819
2b	329	100-101	0.6±0.2	38.8±0.5	40.4±0.4	0.915
2e	344	130-131	3.6±0.9	79.8±3.0	42.3±0.7	1.307

Πίνακας 3-12: Περιλαμβάνει τις 6 πρώτες ενώσεις με όλες τις ιδιότητες

ONOMA	MB	MP(°C)	A431	Tr(Ίχνος)
1f	328	94-95	38.4±0.8	1.011
1g	330	116-117	24.3±0.4	1.307
1j	349	142-143	24.8±0.7	1.011
1k	363	73-74	18.2±0.9	1.819
2b	329	100-101	40.4±0.4	0.915
2e	344	130-131	42.3±0.7	1.307
5a	251	189-190	61.3±1.0	0.915
5i	292	190-191	60.9±2.1	0.915
6e	298	179-180	84.2±3.7	1.507

Πίνακας 3-13: Περιλαμβάνει όλες τις ενώσεις χωρίς τις ιδιότητες που δεν ήταν δυνατό να οριστούν

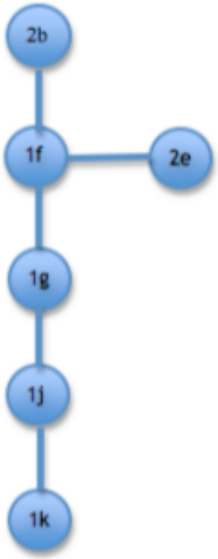
Παραδείγματα (QSPR)

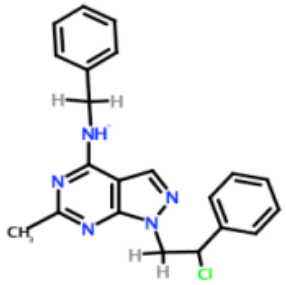
Εφαρμόζοντας τη Μετρική σχέση (τύπος (4)) στον Πίνακα 3-12 , ο οποίος περιλαμβάνει τις 6 πρώτες ενώσεις κι όλες τις ιδιότητες που έχουν αναφερθεί παραπάνω υπολογίζεται ο Πίνακας Διαφορών Dij, που ακολουθεί ακριβώς από κάτω:

Ενώσεις	1f	1g	1j	1k	2b	2e
1f	0	0.83643	1.13887	1.79859	0.68434	0.84852
1g	0.83643	0	0.90595	1.75772	1.40072	1.05053
1j	1.13887	0.90595	0	1.58029	1.22113	0.99427
1k	1.79859	1.75772	1.58029	0	1.72338	1.88134
2b	0.68434	1.40072	1.22113	1.72338	0	1.16109
2e	0.84852	1.05053	0.99427	1.88134	1.16109	0

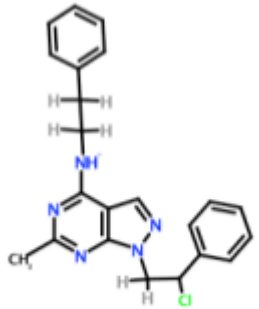
Πίνακας 3-14: Πίνακας Διαφορών για τις 6 ενώσεις με όλες τις ιδιότητες

Και στη συνέχεια ακολουθεί το ελάχιστο επικαλύπτον δέντρο το οποίο προκύπτει:





Εικόνα 2-3: Ένωση 1j



Εικόνα 2-4: Ένωση 1k

Παραδείγματα (QSPR)

4350 *J. Med. Chem.* **2002,** *45,* 4350−4358

Do Structurally Similar Molecules Have Similar Biological Activity?

Yvonne C. Martin,*,† James L. Kofron,‡ and Linda M. Traphagen‡

Global Pharmaceutical Research and Development, Abbott Laboratories, Abbott Park, Illinois 60064-6100

Received April 12, 2002

To design diverse combinatorial libraries or to select diverse compounds to augment a screening collection, computational chemists frequently reject compounds that are ≥0.85 similar to one already chosen for the combinatorial library or in the screening set. Using Daylight fingerprints, this report shows that for IC_{50} values determined as a follow-up to 115 high-throughput screening assays, there is only a 30% chance that a compound that is ⩾ 0.85 (Tanimoto) similar to an active is itself active. Although this enrichment is greater than that found with random screening and docking to three-dimensional structures, this low fraction of actives within similar compounds occurs not only because of deficiencies in the Daylight fingerprints and Tanimoto similarity calculations but also because similar compounds do not necessarily interact with the target macromolecule in similar ways. The current study emphasizes the statistical or probabilistic nature of library design and that perfect results cannot be expected.

Παραδείγματα Κβαντοχημικές Μέθοδοι

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

CHRISTOS CHRISTODOULEAS,[1] DEMETRIOS XENIDES,[1,2] THEODORE E. SIMOS[1]
[1]*Laboratory of Computational Sciences, Department of Computer Science and Technology, Faculty of Sciences and Technology, University of Peloponnese, GR-22 100 Tripolis, Greece*
[2]*Division of Theoretical Chemistry, Department of General, Inorganic, and Theoretical Chemistry, University of Innsbruck, A-6020, Innrain 52a, Austria*

Received 8 September 2008; Revised 31 March 2009; Accepted 14 April 2009
DOI 10.1002/jcc.21328

Abstract: A test set of 10 molecules (open and ring forms of ozone and sulfur dioxide as well as water and hydrogen sulfide and their respective fluoro- and chloro-substituted analogs) of specific atmospheric interest has been formed as to assess the performance of various density functional theory methods in (hyper)polarizability calculations against well-established *ab initio* methods. The choice of these molecules was further based on (i) the profound change in the physics between isomeric systems, e.g., open (C_{2v}) and ring (D_{3h}) forms of ozone, (ii) the relation between isomeric forms, e.g., open and ring form of sulfur dioxide (both of C_{2v} symmetry), and (iii) the effect of the substitution, e.g., in fluoro- and chloro-substituted water analogs. The analysis is aided by arguments chosen from the information theory, graph theory, and pattern recognition fields of Mathematics: In brief, a multidimensional space is formed by the methods which are playing the role of vectors with the independent components of the electric properties to act as the coordinates of these vectors, hence the relation between different vectors (e.g., methods) can be quantified by a proximity measure. Results are in agreement with previous studies revealing the acceptable and consistent behavior of the mPW1PW91, B3P86, and PBE0 methods. It is worth noting the remarkable good performance of the double hybrid functionals (namely: B2PLYP and mPW2PLYP) which are for the first time used in calculations of electric response properties.

 J Comput Chem 28: 000–000, 2009

Παραδείγματα Κβαντοχημικές Μέθοδοι

Table 1. Distance Matrix for the Open Form of O_3 After all the Filtering Processes.

	SCF	MP2	MP3	DQMP4	SDQMP4	MP4	CCSD	CCSD(T)	B3LYP	B3P86	B3PW91	mPW1PW91	mPW1PBE	PBE0	B2PLYP	mPW2PLYP
SCF	0	2.27405	1.42177	1.52012	2.06645	2.44170	1.61392	2.01265	1.96320	1.56753	1.65845	1.54930	1.56147	1.62530	2.01307	1.87432
MP2	2.27405	0	0.97490	1.21718	2.00224	1.83931	1.05409	1.04312	1.29063	1.17954	1.19691	1.21539	1.21828	1.21364	1.03158	1.02842
MP3	1.42177	0.97490	0	0.67943	1.83060	1.99462	0.91262	1.26315	1.39932	1.06375	1.14258	1.07469	1.08543	1.12874	1.28330	1.17790
DQMP4	1.52012	1.21718	0.67943	0	1.27113	1.59697	0.74554	1.16086	1.19781	0.99541	1.03739	0.98829	0.99454	1.01769	1.18453	1.08981
SDQMP4	2.06645	2.00224	1.83060	1.27113	0	0.74702	1.12355	1.24630	1.15410	1.24290	1.21619	1.24010	1.23655	1.20973	1.29060	1.26763
MP4	2.44170	1.83931	1.99462	1.59697	0.74702	0	1.14204	0.95629	0.99106	1.18969	1.14236	1.22208	1.21551	1.15977	1.01327	1.06647
CCSD	1.61392	1.05409	0.91262	0.74554	1.12355	1.14204	0	0.48344	0.58451	0.34633	0.38588	0.38262	0.38701	0.37953	0.51859	0.40957
CCSD(T)	2.01265	1.04312	1.26315	1.16086	1.24630	0.95629	0.48344	0	0.53534	0.52345	0.50843	0.59362	0.59056	0.53946	0.25374	0.29350
B3LYP	1.96320	1.29063	1.39932	1.19781	1.15410	0.99106	0.58451	0.53534	0	0.49986	0.37558	0.46519	0.45118	0.38350	0.40220	0.38392
B3B86	1.56753	1.17954	1.06375	0.99541	1.24290	1.18969	0.34633	0.52345	0.49986	0	0.14450	0.13865	0.14604	0.14317	0.45696	0.32599
B3PW91	1.65845	1.19691	1.14258	1.03739	1.21619	1.14236	0.38588	0.50843	0.37558	0.14450	0	0.11898	0.10925	0.05490	0.40069	0.27777
mPW1PW91	1.54930	1.21539	1.07469	0.98829	1.24010	1.22208	0.38262	0.59362	0.46519	0.13865	0.11898	0	0.1684	0.09369	0.50410	0.37012
mPW1PBE	1.56147	1.21828	1.08543	0.99454	1.23655	1.21551	0.38701	0.59056	0.45118	0.14604	0.10925	0.01684	0	0.08378	0.49788	0.36554
PBE0	1.62530	1.21364	1.12874	1.01769	1.20973	1.15977	0.37953	0.53946	0.38350	0.14317	0.05490	0.09369	0.08378	0	0.43850	0.31256
B2PLYP	2.01307	1.03158	1.28330	1.18453	1.29060	1.01327	0.51859	0.25374	0.40220	0.45696	0.40069	0.50410	0.49788	0.43850	0	0.14436
mPW2PLYP	1.87432	1.02842	1.17790	1.08981	1.26763	1.06647	0.40957	0.29350	0.38392	0.32599	0.27777	0.37012	0.36554	0.31256	0.14436	0

Παραδείγματα Κβαντοχημικές Μέθοδοι

Table 2. Values of the Independent Compontents of Dipole Polarizability, First and Second Hyperpolarizability Obtained at the CCSD(T) Level of Theory and with the Worst(First in Line) and Best DFT Methods for the Ozone, Sulfur Dioxide and their Isomers.

Molecule	Method	μ_z	α_{xx}	α_{yy}	α_{zz}	β_{zxx}	β_{zyy}	β_{zzz}	γ_{xxxx}	γ_{yyyy}	γ_{zzzz}	γ_{xxyy}	γ_{yyzz}	γ_{xxzz}
$O_3(C_{2v})$	CCSD(T)	−0.2184	31.46	12.01	14.53	1.3	−4.1	−19.4	4026	1099	2065	982	540	1059
	mPW1PW91	−0.2610	31.34	11.72	13.91	12.2	−3.3	−16.47	2892	1112	1982	1001	512	967
	B2PLYP	−0.2367	29.60	11.93	14.42	−3.1	−4.5	19.1	4304	1129	2270	997	559	990
$SO_2(C_{2v}$,open)	CCSD(T)	−0.6388	34.44	19.79	23.25	12.6	8.0	29.7	3428	2395	4349	1012	1071	1297
	B3LYP	−0.6686	33.80	19.78	23.12	13.1	8.3	30.6	4112	2660	5105	1137	1230	1535
	mPW2PLYP	−0.6552	34.02	19.76	23.17	12.3	8.2	31.2	3712	2497	4724	1047	1141	1387
$SO_2(C_{2v}$,ring)	CCSD(T)	−0.5890	27.73	24.45	34.63	19.5	76.4	88.4	2563	9168	4554	2418	3261	1435
	B3LYP	−0.6323	27.95	24.93	34.60	21.6	91.7	100.5	2988	12549	5513	3139	4195	1726
	mPW2PLYP	−06285	27.78	24.57	34.63	20.5	83.4	94.3	2788	10598	4934	2696	3604	1589
				α_{yy}	α_{zz}			β_{zzz}		γ_{yyyy}	γ_{zzzz}		γ_{yyzz}	
$O_3(D_{3h})$	CCSD(T)	–	–	11.96	18.77	–	–	6.5	–	1467	1256	–	693	
	B3LYP	–	–	11.79	18.95	–	–	9.2	–	1603	1390	–	799	
	PBE0	–	–	11.62	18.72	–	–	7.4	–	1460	1271	–	719	

Παραδείγματα Κβαντοχημικές Μέθοδοι

Table 3. Values of the Independent Compontents of Dipole Polarizability, First and Second Hyperpolarizability Obtained at the CCSD(T) Level of Theory and with the Worst (First in Line) and Best DFT Methods for Water, Hydrogen Sulfide and their Fluoro- and Chloro-substituted Analogues.

Molecule	Method	μ_z	α_{xx}	α_{yy}	α_{zz}	β_{zxx}	β_{zyy}	β_{zzz}	γ_{xxxx}	γ_{yyyy}	γ_{zzzz}	γ_{xxyy}	γ_{yyzz}	γ_{xxzz}
H_2O	CCSD(T)	0.7186	10.04	9.60	9.74	−9.9	−6.0	−14.3	868	2787	1503	679	740	448
	B3PW91	0.7282	10.06	9.61	9.80	−10.5	−6.0	−14.9	1038	3388	1806	829	883	530
	mPW2PLYP	0.7326	10.09	9.60	9.80	−10.02	−5.9	−14.6	974	3068	1656	753	820	496
F_2O	CCSD(T)	−0.1189	20.26	10.14	13.60	−2.2	2.2	−0.5	1366	785	920	313	358	343
	mPW1PBE	−0.1289	20.28	9.84	13.41	0.9	1.3	−0.2	1314	791	970	347	376	341
	B2PLYP	−0.1291	20.36	10.06	13.60	−0.7	1.4	−0.6	1349	812	982	348	378	349
Cl_2O	CCSD(T)	0.2180	53.08	29.38	32.06	−10.1	−27.5	−44.7	13638	6756	4862	2935	2000	2238
	B3LYP	0.2198	55.05	29.45	32.03	−11.7	−30.2	−47.9	18704	7932	5493	3584	2300	2614
	mPW2PLYP	0.2274	53.63	29.08	31.82	−13.7	−28.6	−46.3	15150	7173	5052	3175	2093	2363
H_2S	CCSD(T)	0.3854	24.64	25.57	24.66	−9.6	−6.4	−1.2	3991	14626	5295	3477	3621	1732
	B3P86	0.4103	24.66	25.40	24.68	−11.2	−5.8	−2.8	4336	15572	5632	3769	3813	1873
	mPW2PLYP	0.4018	24.75	25.51	24.72	−10.7	−6.0	−2.1	4339	15918	5703	3883	3973	1905
F_2S	CCSD(T)	−0.4169	26.40	22.09	23.81	30.2	59.3	68.5	2092	9153	3594	1594	2322	1050
	B3LYP	−0.4130	26.41	22.45	23.95	31.8	68.9	73.8	2546	12115	4387	2079	3057	1280
	mPW2PLYP	−0.4286	26.26	21.97	23.74	29.9	61.6	69.0	2255	10027	3883	1746	2547	1136
Cl_2S	CCSD(T)	−0.1572	68.91	38.92	48.58	0.9	20.2	26.9	14638	8859	9549	3794	4376	3680
	mPW1PW91	−0.1473	68.40	37.87	47.67	1.3	17.2	24.8	15044	8833	9579	4019	4517	3737
	B2PLYP	−0.1595	69.20	38.56	48.45	1.5	19.1	26.8	15297	9158	9987	4101	4669	3830

Παραδείγματα Κβαντοχημικές Μέθοδοι

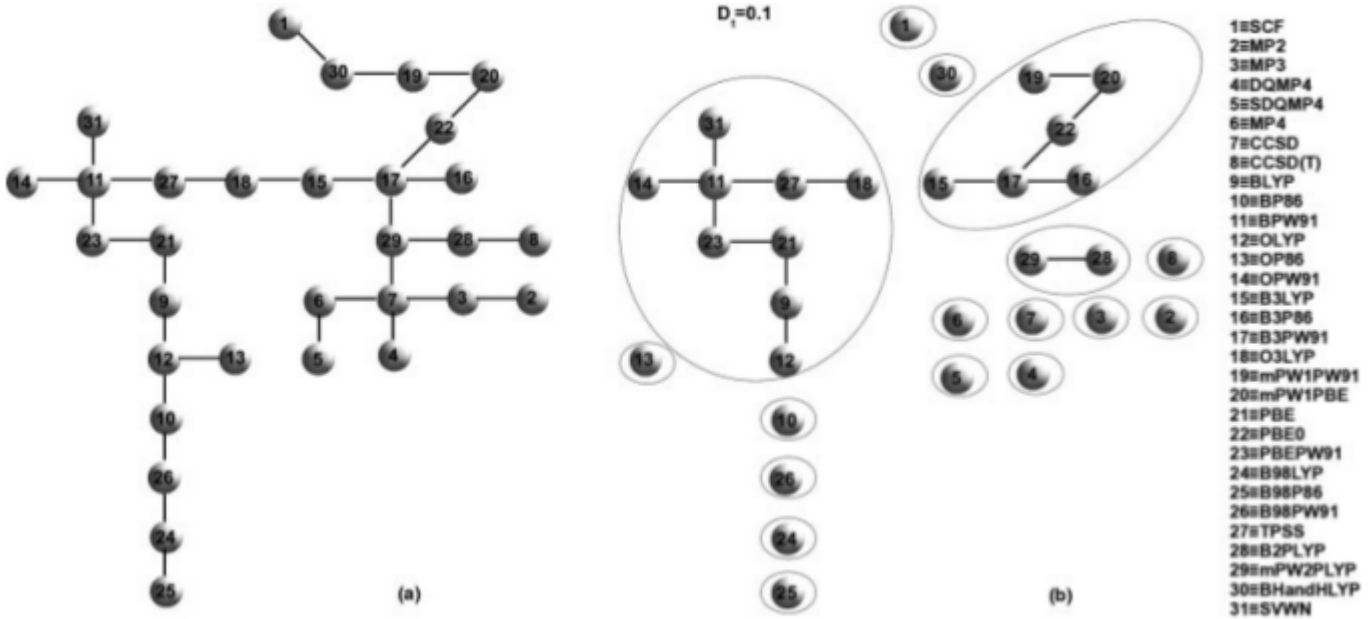


Figure 1. Minimum spanning tree (a) and clustering (b) for a $D_t = 0.1$ for the 31 methods [case study: O_3, (C_{2v})].

Παραδείγματα Κβαντοχημικές Μέθοδοι

A critical analysis of the performance of new generation functionals on the calculation of the (hyper) polarizabilities of clusters of varying stoichiometry: Test case the Si_mGe_n ($m + n = 7$, $n = 0–7$) clusters

Demetrios Xenides [a,b,*], Panaghiotis Karamanis [c], Claude Pouchan [c]

[a] Department of Computer Science and Technology, School of Natural Sciences, University of Pelaponnese, GR-22100, Tripolis, Greece
[b] Division of Theoretical Chemistry, Department of General, Inorganic, and Theoretical Chemistry, University of Innsbruck, A-6020, Innrain 52a, Austria
[c] Groupe de Chimie Théorique et Réactivité, ECP, IPREM UMR 5254, Université de Pau et de Pays de l' Adour, Hélioparc Pau Pyrénées 2 avenue du Président Angot, 64053 PAU Cedex 09, France

ARTICLE INFO

Article history:
Received 24 April 2010
In final form 17 August 2010
Available online 21 August 2010

ABSTRACT

The continuous efforts on the improvement of the Density Functional Theory (DFT) resulted to a plethora of new functionals. A choice of the promising ones belonging to the long-range corrected, hybrid meta-GGA, and the double-hybrid families along with the HF and MP2 *ab initio* methods have been introduced in electric response properties calculations of Si_mGe_n ($m + n = 7$, $n = 0–7$) clusters. An information theory based analysis of the obtained results enables us to assess the methods relative performance. The findings suggest that the methods are grouped, in respect to their overall performance, as: group **A** = MP2, B2PLYP, mPW2PLYP, group **B** = BLYP, B3LYP, group **C** = M06, CAM-B3LYP, LC-BLYP, HF and **D** = M06, with distinct and large differences between them.

Παραδείγματα Κβαντοχημικές Μέθοδοι

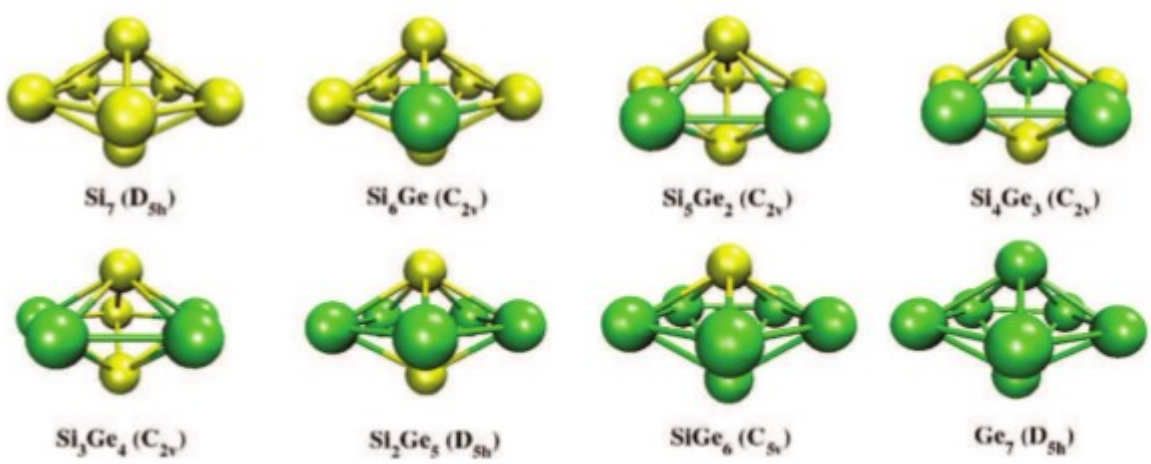


Figure 1. Schematical representation of the Si_mGe_n ($m + n = 7$, $n = 0–7$) clusters.

Παραδείγματα Κβαντοχημικές Μέθοδοι

Table 1
Ab initio and DFT isotropic and anisotropic values of electric (hyper) polarizabilities (hyperpolarizabilities have been divided by 10^3) for all clusters (in *italics* are the values when geometries have been optimized at the same level as the one used in the calculation of the properties).

Molecule	Property	HF	MP2	BLYP	LC-BLYP	B3LYP	CAM-B3LYP	M06	B2PLYP	mPW2PW91
$Si_7(D_{5h})^{(i)}$	$\bar{\alpha}$	208.41	213.71	210.90	203.54	208.43	206.46	202.59	210.24	210.03
		211.08	*218.13*	*217.89*	*198.53*	*211.79*	*206.13*	*200.79*	*213.90*	*213.10*
	$\Delta\alpha$	67.81	84.23	80.31	70.12	77.28	74.05	72.79	79.05	78.32
		67.40	*86.65*	*84.01*	*66.22*	*78.74*	*73.12*	*72.36*	*80.83*	*79.69*
	$\bar{\gamma}$	95.68	121.67	141.91	87.74	120.53	104.47	104.97	119.74	118.68
		98.12	*125.40*	*148.67*	*85.25*	*123.82*	*104.99*	*104.01*	*122.95*	*120.62*
$Si_6Ge(C_{2v})$	$\bar{\alpha}$	212.79	218.01	214.75	207.35	212.35	210.36	209.34	214.28	214.09
	$\Delta\alpha$	71.49	86.84	83.23	73.49	80.35	77.28	79.06	81.98	81.32
	$\bar{\gamma}$	99.22	127.06	146.96	89.44	124.46	107.30	126.86	124.20	123.50
$Si_5Ge_2(C_{2v})$	$\bar{\alpha}$	217.08	222.25	218.57	211.09	216.22	214.21	216.24	218.25	218.08
	$\Delta\alpha$	74.89	89.37	86.11	76.66	83.33	80.36	84.86	84.81	84.19
	$\bar{\gamma}$	102.93	132.59	151.84	91.02	128.36	110.05	137.19	128.72	128.63
$Si_4Ge_3(C_{2v})$	$\bar{\alpha}$	221.41	226.50	222.30	214.81	220.04	218.01	223.12	222. 22	222.06
		224.78	*232.56*	*231.13*	*208.04*	*224.42*	*217.56*	*222.93*	*227.12*	*226.24*
	$\Delta\alpha$	78.64	92.21	88.87	79.91	86.31	83.48	91.30	87.77	87.20
		78.34	*96.47*	*93.69*	*74.09*	*88.32*	*82.14*	*92.23*	*90.62*	*89.47*
	$\bar{\gamma}$	105.86	137.30	154.52	91.78	130.65	111.72	145.11	132.00	132.43
		109.61	*142.96*	*164.06*	*87.96*	*135.13*	*112.01*	*145.84*	*136.80*	*136.62*
$Si_3Ge_4(C_{2v})$	$\bar{\alpha}$	225.61	230.69	226.08	218.47	223.85	221.79	230.28	226.14	226.00
	$\Delta\alpha$	81.86	94.70	91.69	82.93	89.20	86.45	97.10	90.52	89.98
	$\bar{\gamma}$	109.99	143.42	159.76	94.10	135.04	115.14	150.83	137.15	138.07
$Si_2Ge_5(D_{5h})^{(ii)}$	$\bar{\alpha}$	229.62	234.51	229.83	222.12	227.65	225.56	237.50	230.06	229.95
	$\Delta\alpha$	85.42	97.31	94.45	85.92	92.05	89.38	103.09	93.28	92.78
	$\bar{\gamma}$	114.58	149.91	164.25	96.35	138.81	118.09	156.00	141.74	146.10
$SiGe_6(C_{5v})$	$\bar{\alpha}$	234.23	239.04	234.19	226.21	231.97	229.82	244.54	234.40	234.28
	$\Delta\alpha$	84.19	96.03	93.79	85.07	91.19	88.50	99.23	92.16	91.65
	$\bar{\gamma}$	120.02	156.15	175.23	100.01	148.06	123.23	161.03	147.78	148.95
$Ge_7(D_{5h})^{(iii)}$	$\bar{\alpha}$	238.21	242.89	238.01	229.68	235.70	233.46	249.85	238.15	238.00
	$\Delta\alpha$	82.95	96.20	93.48	84.08	90.66	87.78	96.48	91.55	91.01
	$\bar{\gamma}$	126.12	165.09	178.56	103.31	150.98	127.94	169.52	154.53	156.80

(i) MP4 values are: $\bar{\alpha} = 213.09$, $\Delta\alpha = 83.66$ and $\bar{\gamma} = 120.88$, (ii) MP4 values are: $\bar{\alpha} = 234.45$, $\Delta\alpha = 97.08$ and $\bar{\gamma} = 150.14$, and (iii) MP4 values are: $\bar{\alpha} = 242.57$, $\Delta\alpha = 94.67$ and $\bar{\gamma} = 167.35$, respectively.

Παραδείγματα Κβαντοχημικές Μέθοδοι

Table 2
Distances between the MP2 and all the implemented methods (in the rows under Si_7 and Si_4Ge_3 and in *italics* are the calculated distances when geometries have been optimized at the same level as the one used in the calculation of the properties).

Molecule	Method	HF	MP2	BLYP	LC-BLYP	B3LYP	CAM- B3LYP	M06	B2PLYP	mPW2PLYP
Si_7	MP2	1.29	0	0.77	1.54	0.52	0.92	1.36	0.36	0.39
		1.05	*0*	*0.71*	*1.63*	*0.35*	*0.89*	*1.33*	*0.25*	*0.35*
Si_6Ge	MP2	1.68	0	0.97	2.18	0.81	1.37	1.56	0.54	0.58
Si_5Ge_2	MP2	1.66	0	0.94	2.26	0.86	1.43	1.30	0.57	0.59
Si_4Ge_3	MP2	1.60	0	0.92	2.30	0.92	1.48	1.19	0.60	0.61
		1.32	*0*	*0.72*	*2.41*	*0.65*	*1.42*	*1.16*	*0.43*	*0.48*
Si_3Ge_4	MP2	1.54	0	0.88	2.32	0.92	1.50	1.09	0.60	0.60
Si_2Ge_5	MP2	1.14	0	0.63	1.80	0.67	0.42	0.94	0.43	0.38
$SiGe_6$	MP2	1.35	0	0.82	1.94	0.84	1.28	0.78	0.59	0.61
Ge_7	MP2	1.03	0	0.48	1.69	0.58	1.06	1.14	0.38	0.35

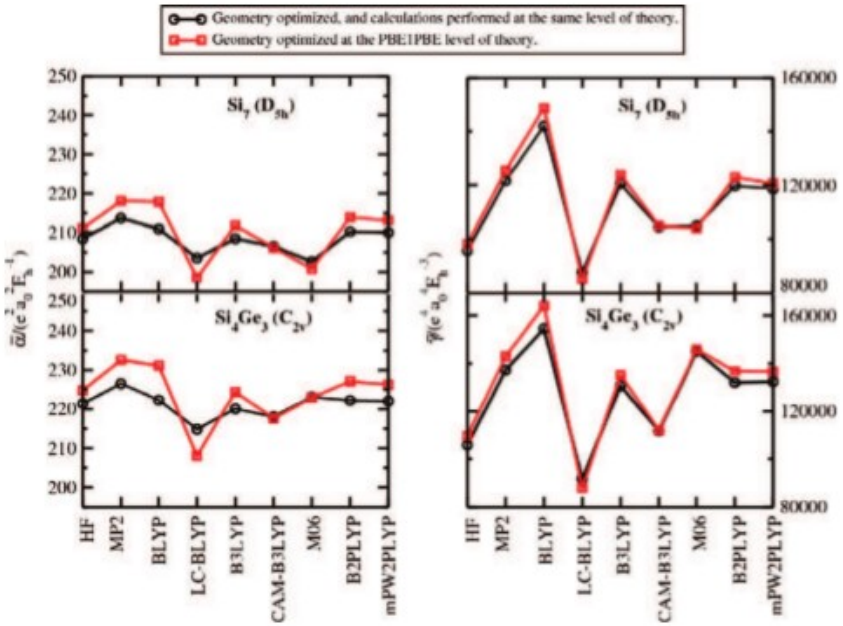


Figure 4. Structural effects on the (hyper) polarizabilities of Si_7 (up) and Si_4Ge_3 (down), respectively.

Παραδείγματα Κβαντοχημικές Μέθοδοι

To summarize the findings of the present study it can be said that the variations in the stoichiometry of the Si_mGe_n ($m + n = 7$, $n = -7$) clusters results in a smooth, monotonic and almost linear increase on both the polarizability and second hyperpolarizability as a function of the number of Ge atoms. The observed (hyper) polarizability enhancement is due to the more (hyper) polarizable character of Ge. On the other hand, the performed information theory and metric space approach analysis of the independent components, of the (hyper) polarizabilities, yield to the following classification of the implemented methods in respect to their overall performance: group **A** = {MP2, B2PLYP, mPW2PLYP}, group **B** = {BLYP, B3LYP}, group **C** = {CAM-B3LYP, LC-BLYP, HF} and group **D** = {M06}, with distinct and large differences between these groups. This grouping verifies the critical effect of the incorporation of the LC correction on conventional DFT functionals in the prediction of cluster (hyper) polarizabilities, while, it reveals that for clusters at least of this size the long-range corrected functionals we checked in this study are closer to the HF approximation than to the MP2(or 4) perturbation theory.

Παραδείγματα Κβαντοχημικές Μέθοδοι

Πίνακας 1.Τιμές των ιδιοτήτων με ab initio και Density Functional Theory (DFT) μεθόδων για το Si_2C_2.

	α_{xx}	α_{yy}	α_{zz}	α	$\Delta\alpha$	γ_{xxxx}	γ_{yyyy}	γ_{zzzz}	γ_{xxyy}	γ_{yyzz}	γ_{xxzz}	γ
SCF	113.66	59.76	63.66	79.03	52.05	56052	14321	16055	11019	4469	10026	27491
MP2	123.64	59.59	65.70	82.98	61.23	72547	16559	26614	14539	6267	15474	37656
MP3	119.00	58.37	63.88	80.42	58.08	62600	15184	20326	12599	5171	12417	31697
DQMP4	118.56	58.08	63.95	80.20	57.76	62774	14801	21043	12349	5217	12534	31763
SDQMP4	119.10	57.98	64.46	80.51	58.15	65049	14758	23636	12544	5600	13483	33339
MP4	123.70	58.71	65.64	82.68	61.82	72800	16407	28200	14365	6499	15896	38185
CCSD	118.98	58.27	64.32	80.52	57.92	63024	15263	22657	12592	5604	13110	32711
CCSD(T)	121.80	58.68	65.07	81.85	60.18	67570	16336	25586	13760	6207	14658	35748
BLYP	112.15	58.41	62.56	77.71	51.80	92040	19647	29457	18298	7076	-40782	22066
BP86	111.33	58.11	61.99	77.14	51.40	84181	15632	32054	16862	7820	16413	42812
BPW91	110.51	57.52	61.33	76.46	51.19	80285	17207	24567	16024	6074	15642	39508
B3LYP	108.27	57.61	61.18	75.69	48.97	73332	16529	22550	14659	5692	14308	36346
B3P86	107.03	57.04	60.40	74.82	48.40	65598	14845	19533	13089	5018	12574	32268
B3PW91	107.32	57.09	60.43	74.94	48.64	67529	15256	20223	13531	5173	12986	33277
OLYP	111.67	58.13	61.91	77.24	51.75	92007	18849	29191	17965	7027	17725	45096
OP86	111.03	57.84	61.51	76.79	51.46	27016	42284	-46853	3714	-12361	-36722	-13659
OPW91	110.07	57.32	60.69	76.03	51.15	79446	16283	24135	15636	5912	14937	38567
O3LYP	109.42	57.71	61.23	76.12	50.04	79948	17078	24790	15727	6118	15338	39236
B98LYP	128.20	67.95	72.69	89.62	58.03	207229	41465	70246	41659	16344	43081	104222
B98P86	126.76	67.39	71.74	88.63	57.32	196719	37020	59134	37077	13927	51937	99751
B98PW91	125.47	66.59	70.76	87.61	56.90	172783	35225	56071	35183	13370	35016	86243
PBE1PBE	107.11	57.36	60.59	75.02	48.22	66750	15179	20050	13364	5192	12810	32942
PBEPBE	111.32	58.08	61.98	77.13	51.40	84576	18064	26659	16859	6526	16623	41863
PBEPW91	111.27	58.05	61.96	77.09	51.38	84429	18058	26726	16829	6552	16620	41843
mPW1PW91	106.72	57.07	60.31	74.70	48.11	65577	15036	19874	13134	5101	12634	32445
mPW1PBE	106.77	57.11	60.33	74.73	48.13	65771	15109	19890	13231	5168	12688	32589
B2PLYP	119.35	58.31	64.16	80.61	58.34	92856	19394	34916	18370	7739	20003	47878
mPW2PLYP	118.68	58.25	64.03	80.32	57.76	91445	19372	34520	18148	7785	19649	47300

Παραδείγματα Κβαντοχημικές Μέθοδοι

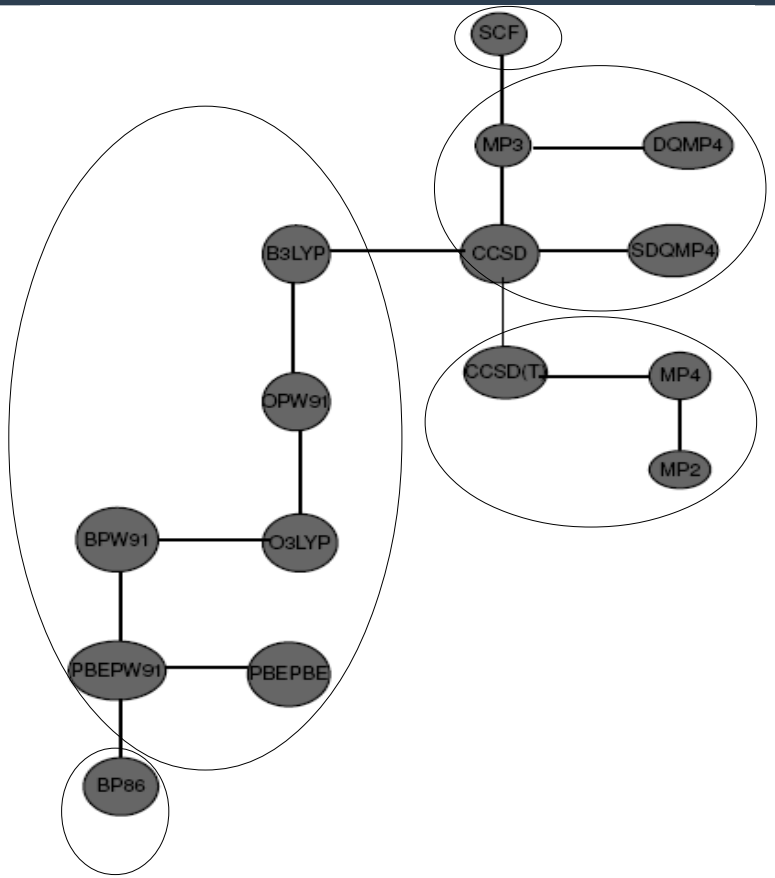
Πίνακας 2.Τιμές των αποστάσεων με ab initio και Density Functional Theory (DFT) μεθόδων για το Si_2C_2.

	SCF	MP2	MP3	DQMP4	SDQMP4	MP4	CCSD	CCSD(T)	BP86	BPW91	B3LYP	OPW91	O3LYP	PBEPBE	PBEPW91
SCF	0	1.74618	0.93368	0.99675	1.21499	1.86623	1.09922	1.53073	2.37147	2.09480	1.72219	1.99304	2.02451	2.32905	2.33150
MP2	1.74618	0	1.14682	1.22883	1.07971	0.38930	1.01526	0.49056	1.51304	1.54699	1.60970	1.66732	1.54348	1.45099	1.45716
MP3	0.93368	1.14682	0	0.17322	0.38036	1.11300	0.24292	0.76746	1.74611	1.44405	1.17971	1.37196	1.41583	1.68801	1.68821
DQMP4	0.99675	1.22883	0.17322	0	0.28446	1.15085	0.24604	0.82523	1.73349	1.45612	1.18132	1.36064	1.42948	1.72482	1.72377
SDQMP4	1.21499	1.07971	0.38036	0.28446	0	0.94132	0.21271	0.66030	1.56945	1.39091	1.18519	1.31948	1.37239	1.62290	1.62120
MP4	1.86623	0.38930	1.11300	1.15085	0.94132	0	0.93169	0.37740	1.35584	1.40463	1.49678	1.51659	1.41919	1.34694	1.34945
CCSD	1.09922	1.01526	0.24292	0.24604	0.21271	0.93169	0	0.59368	1.61038	1.38358	1.16263	1.33450	1.35914	1.59895	1.59812
CCSD(T)	1.53073	0.49056	0.76746	0.82523	0.66030	0.37740	0.59368	0	1.41960	1.32274	1.30143	1.39547	1.32062	1.36179	1.36337
BP86	2.37147	1.51304	1.74611	1.73349	1.56945	1.35584	1.61038	1.41960	0	0.89386	1.14451	0.94855	0.87466	0.82847	0.82204
BPW91	2.09480	1.54699	1.44405	1.45612	1.39091	1.40463	1.38358	1.32274	0.89386	0	0.48985	0.32338	0.13365	0.47439	0.46656
B3LYP	1.72219	1.60970	1.17971	1.18132	1.18519	1.49678	1.16263	1.30143	1.14451	0.48985	0	0.37256	0.41969	0.90441	0.89860
OPW91	1.99304	1.66732	1.37196	1.36064	1.31948	1.51659	1.33450	1.39547	0.94855	0.32338	0.37256	0	0.30616	0.77208	0.76461
O3LYP	2.02451	1.54348	1.41583	1.42948	1.37239	1.41919	1.35914	1.32062	0.87466	0.13365	0.41969	0.30616	0	0.51014	0.50304
PBEPBE	2.32905	1.45099	1.68801	1.72482	1.62290	1.34694	1.59895	1.36179	0.82847	0.47439	0.90441	0.77208	0.51014	0	0.01776
PBEPW91	2.33150	1.45716	1.68821	1.72377	1.62120	1.34945	1.59812	1.36337	0.82204	0.46656	0.89860	0.76461	0.50304	0.01776	0

Παραδείγματα Κβαντοχημικές Μέθοδοι

1	3	0.9337	1.SCF
3	4	0.1732	2.MP2
3	7	0.2429	3.MP3
7	5	0.2127	4.DQMP4
7	8	0.5937	5.SDQMP4
8	6	0.3774	6.MP4
6	2	0.3893	7.CCSD
7	11	1.1626	8.CCSD(T)
11	12	0.3725	9.BP86
12	13	0.3062	10.BPW91
13	10	0.1336	11.B3LYP
10	15	0.4666	12.OPW91
15	14	0.0176	13.O3LYP
15	9	0.8222	14PBEPBE
			15.PBEPW91

D_T=0.5: {1},{3,4,5,7},{2,6,8}{10,11,12,13,14}{9}



Γράφημα 1. MST που αντιστοιχεί στον Πίνακα 2.

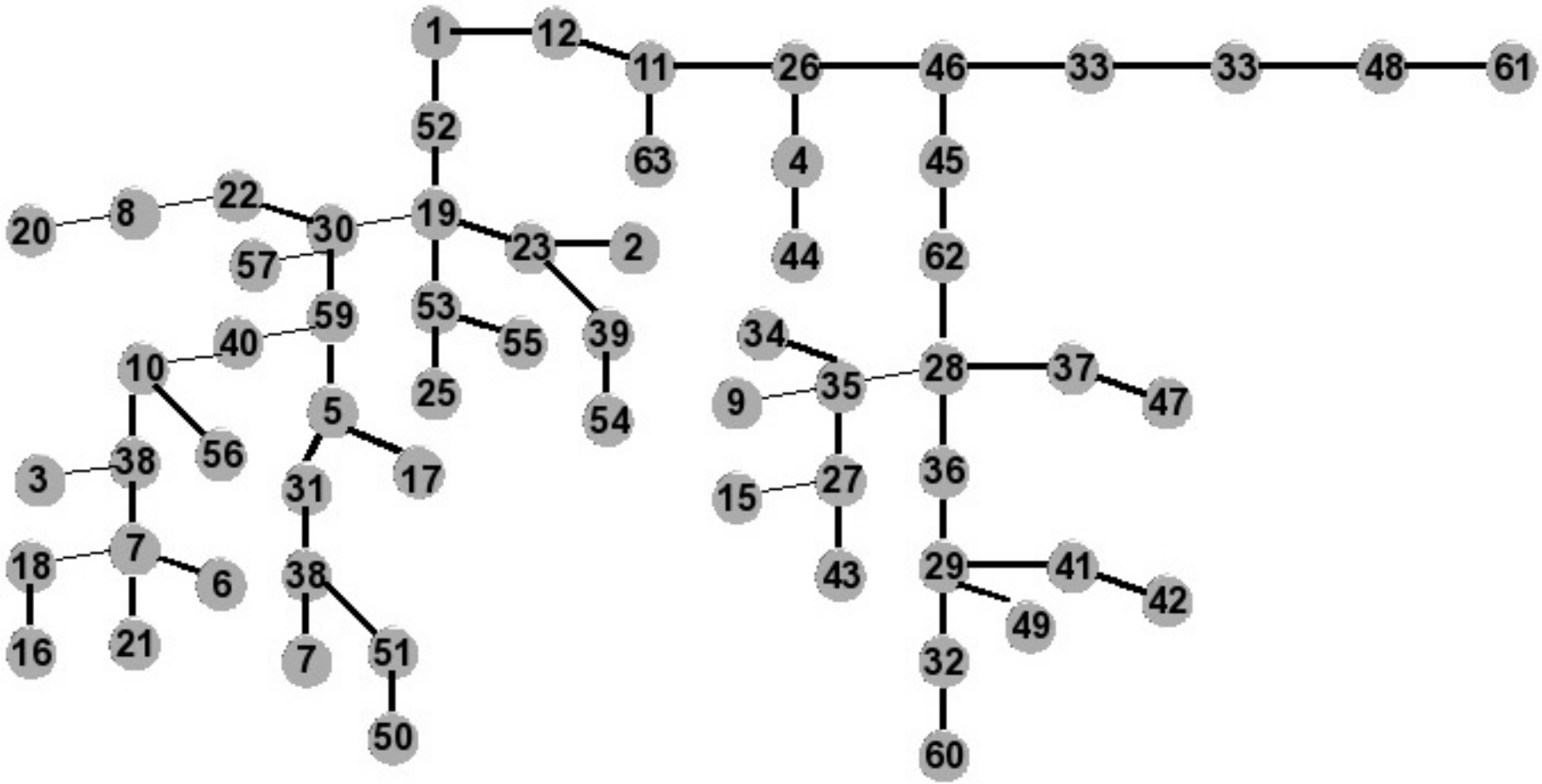
Παραδείγματα Φυσικοχημικές Μέθοδοι XRF

Α/Α Δείγματος	Δείγμα από Πίνακα ΙΙ (XRF)	Κόμβος Α	Κόμβος Β	Απόσταση μεταξύ κόμβων
1	A1	1	12	0.25
2	A2	1	58	0.31
3	A3	58	13	0.22
4	A4	1	52	0.38
5	A5	52	19	0.19
6	A6	19	30	0.23
7	A7	19	23	0.27
8	A8	23	2	0.32
9	A9	30	57	0.33
10	A10	23	39	0.33
11	A11	19	53	0.35
12	A12	53	25	0.32
13	A13	39	54	0.35
14	A14	30	59	0.36
15	A15	59	40	0.39
16	A16	59	5	0.42
17	A17	5	31	0.37
18	A18	12	11	0.43
19	A19	11	26	0.34
20	A20	26	46	0.36
21	A21	46	33	0.21
22	A22	33	48	0.25
23	B1	46	45	0.27
24	B2	11	63	0.39
25	B3	45	62	0.40
26	B4	62	28	0.38
27	B5	28	35	0.40
28	B6	35	9	0.37
29	B7	28	37	0.40
30	B8	35	27	0.40
31	B9	27	15	0.37
32	B10	35	34	0.40
33	B11	26	4	0.41
34	B12	62	24	0.42
35	B13	37	47	0.42
36	B14	27	43	0.42
37	B15	31	64	0.44
38	B16	64	14	0.34
39	B17	30	22	0.45
40	B18	40	10	0.46
41	B19	10	38	0.42
42	B20	28	36	0.46
43	C1	36	29	0.40
44	C2	29	49	0.37

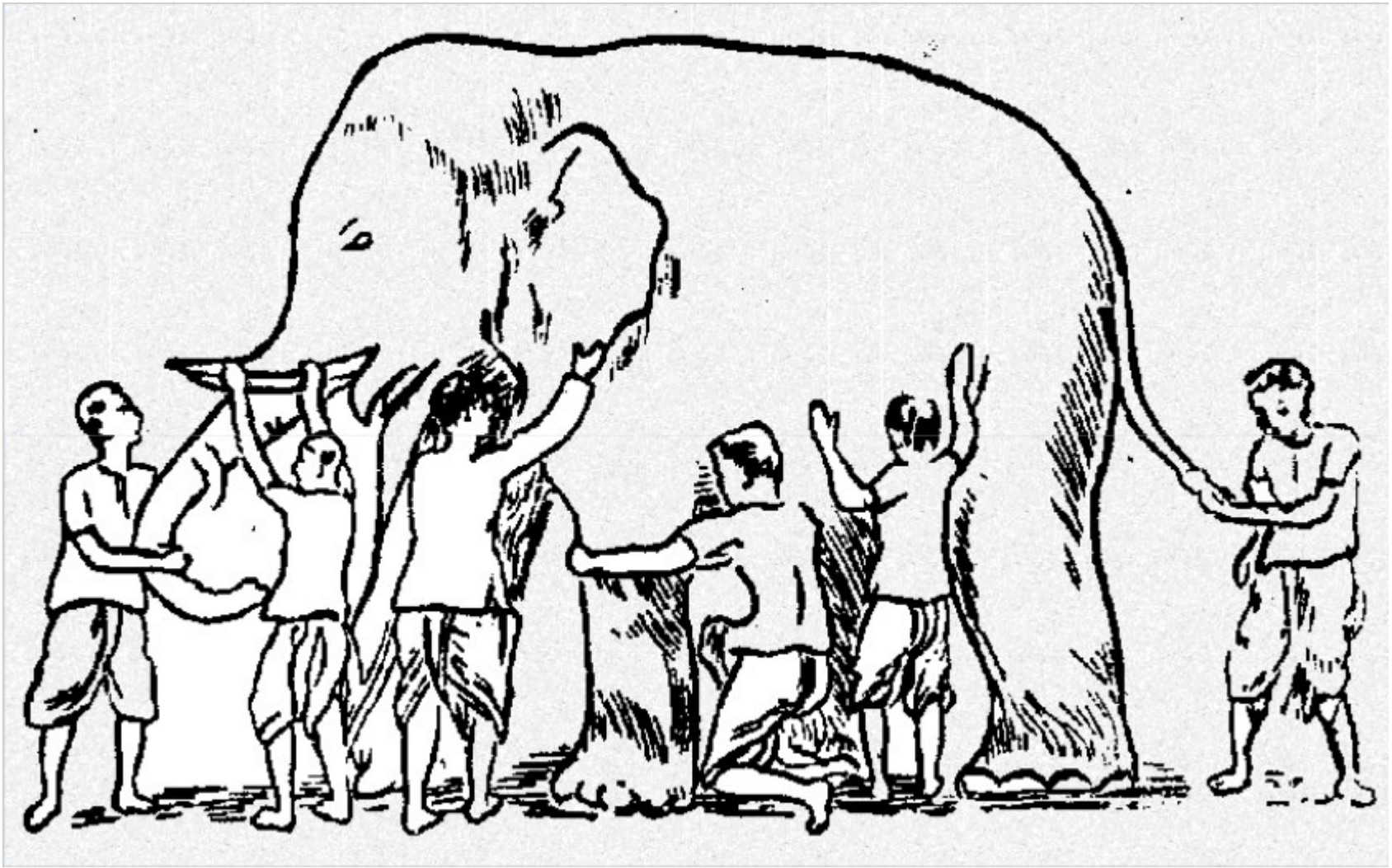
Παραδείγματα Φυσικοχημικές Μέθοδοι XRF

45	C3	29	41	0.38
46	C4	10	56	0.46
47	C5	38	7	0.47
48	C6	7	18	0.35
49	C7	7	6	0.40
50	C8	7	21	0.45
51	C9	38	3	0.48
52	C10	18	16	0.48
53	C11	5	17	0.51
54	C12	53	55	0.56
55	C13	22	8	0.56
56	C14	29	32	0.59
57	C15	32	60	0.52
58	C16	64	51	0.62
59	C17	48	61	0.67
60	C18	4	44	0.67
61	C19	41	42	0.67
62	C20	51	50	0.68
63	C21	8	20	0.76
64	C22			

Παραδείγματα Φυσικοχημικές Μέθοδοι XRF



Take home message...



Take home message...

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind
The First approached the Elephant,
And happening to fall
Against his broad and sturdy side,
At once began to bawl:
"God bless me! but the Elephant
Is very like a wall!"
The Second, feeling of the tusk,
Cried, "Ho! what have we here
So very round and smooth and sharp?
To me 'tis mighty clear
This wonder of an Elephant
Is very like a spear!"
The Third approached the animal,
And happening to take
The squirming trunk within his hands,
Thus boldly up and spake:
"I see," quoth he, "the Elephant
Is very like a snake!"

The Fourth reached out an eager hand,
And felt about the knee.
"What most this wondrous beast is like
Is mighty plain," quoth he;
" 'Tis clear enough the Elephant
Is very like a tree!"
The Fifth, who chanced to touch the ear
Said: "E'en the blindest man
Can tell what this resembles most;
Deny the fact who can
This marvel of an Elephant
Is very like a fan!"
The Sixth no sooner had begun
About the beast to grope,
Than, seizing on the swinging tail
That fell within his scope,
"I see," quoth he, "the Elephant
Is very like a rope!"
And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!