Article

A Hairpin DNA-Based Piezoelectric E-Nose: Exploring the Performances of Heptamer Loops for the Detection of Volatile Organic Compounds

Sara Gaggiotti [1], Marcello Mascini [2,3,]*[ ], Angelo Cichelli [1], Michele Del Carlo [2] and Dario Compagnone [2,]*[ ]

[1] Department of Medical, Oral and Biotechnological Sciences, "G. d'Annunzio" University Chieti-Pescara, Via dei Vestini 31, 66100 Chieti, Italy; sgaggiotti@unite.it (S.G.); cichelli@unich.it (A.C.)

[2] Faculty of Bioscience and Technology for Food, Agriculture and Environment, University of Teramo, 64100 Teramo, Italy; mdelcarlo@unite.it

[3] Department of Analytical Chemistry, Faculty of Chemistry, University Complutense of Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain

* Correspondence: mmascini@unite.it (M.M.); dcompagnone@unite.it (D.C.)

Abstract: A hairpin DNA (hpDNA) piezoelectric gas sensors array with heptamer loops as sensing elements was designed, realized, and challenged with pure volatile organic compounds VOCs and real samples (beer). The virtual binding versus five chemical classes (alcohols, aldehydes, esters, hydrocarbons, and ketones) of the entire combinatorial library of heptamer loops (16,384 elements) was studied by molecular modelling. Six heptamer loops, having the largest variance in binding the chemical classes, were selected to build the array. The six gas sensors were realized by immobilizing onto gold nanoparticles (AuNPs) via a thiol spacer the hpDNA constituted by the heptamer loops and the same double helix stem of four base pairs (GAAG at 5' and CTTC at 3' end). The HpDNA-AuNP was used to modify the surface of 20 MHz quartz crystal microbalances (QCMs). The realized E-nose was able to clearly discriminate among 15 pure VOCs of different chemical classes, as demonstrated by hierarchical cluster analysis. The analysis of real beer samples during fermentation was also carried out. In such a challenging matrix consisting of 23 different VOCs, the hpDNA E-nose with heptamer loops was able to discriminate among different fermentation times with high success rate. Class assignment using the Bayes theorem gave an excellent 98% correct beer samples classification in cross-validation.

Keywords: electronic nose; hpDNA; SPME/GC-MS; gas sensors; beer; QCMs

1. Introduction

Gas sensors are able to detect volatile organic compounds (VOCs) through the presence of sensitive elements on their surface and may be used to improve their recognition ability, in an array format: electronic nose (E-noses). Different types of sensitive elements, inspired by the biological sense of smell, have been used in E-nose technology [1]: odorant-binding proteins (OBP), insect antennae, olfactory tissues, olfactory neurons, synthetic peptides, molecularly imprinted polymers (MIP), and DNA [2–5]. Gaggiotti et al. [6] recently reviewed the progress of DNA-based gas sensors.

Genomic DNA [7] or long sequences of DNA to realize an array of sensors based on field-effect transistors technologies (FETs) has been reported [8].

Hairpin DNA (hpDNA) has been used for the first time as a sensing element in piezoelectric gas sensors (QCMs) after immobilization onto gold nanoparticles (hpDNA-AuNPs) [9]. The hpDNA consists of a stem-loop based structure. The stem is formed when a base pairing pattern of molecules with a complementary sequence is present at the distal ends of an oligonucleotide. This generates an unpaired loop that can interact with different volatile compounds.

The use of hpDNA in the sensor array is a particularly advantageous sensor array since variability in binding organic molecules is definitely larger than that of classical metal oxides sensors. In this particular case, it is given by the different bases and by the loop size, as already reported in a previous work [9]. With respect to other organic sensor arrays, hpDNA has clear advantages in the simplicity of the realization of the array and the cost of the sensors with respect to olfactory neurons and does not present problems of sensor assembly and/or reversibility of the signal found for molecularly imprinted polymers [6]. Moreover, considering the use of peptide-based sensor arrays [10], it has already been demonstrated that the information given by hpDNA can be complementary [10] because of the binding of particular classes of compounds as terpenes [11].

In this work, we exploit the performances of hpDNA having seven bases in the loop for the realization of gas sensor arrays. A novel approach was used for the selection of the loop sequences among the resulting 16,384 elements of the library. Six different sequences with the largest differences in binding five chemical classes were used for the realization of the array. The potential of this array was tested both on pure compounds and on challenging samples as beer. The data obtained, compared to smaller loops and validated using GC-MS, demonstrate that larger single-stranded DNA (ssDNA) loops with more complex shape results in improved efficiency and effectiveness in gas sensing using the E-nose format.

2. Materials and Methods

2.1. In Silico Selection of ssDNA Heptamer Loops

The in-silico screening procedure was aimed to test the virtual binding affinities of the ssDNA heptamer library versus 50 VOCs (12 alcohols, 13 aldehydes, 17 esters, 5 hydrocarbons and 3 ketones) of different molecular weight, polar surface area, and LogP. Name, functional group, label, molecular weight, polar surface area, and hydrophobicity (LogP) of the VOCs are reported in Supplementary Material (Table S1).

All possible combinations of the four DNA bases were used to build the ssDNA heptamer library, resulting in 16,384 elements. Hyperchem 8.0.5 (Hypercube inc, Gainesville, FL, USA) was used to generate the library. Tools from OpenEye Scientific Software package (Santa Fe, NM, USA) under academic license were used at different stages of the in-silico procedure. VOCs were obtained via LEXICHEM 2.1.0 package (is a part of the Opeye software) by converting the ligands' standard IUPAC names into their corresponding structures [12]. SZYBKI 1.5.7 with default parameterization was used to optimize molecular geometries [13]. Conformational spaces for both ssDNA and VOCs were taken into account with OMEGA 2.4.6 [14–16]. Multi-conformer rigid body docking was carried out using OEDocking 3.0.0, having Chemgauss4 as a scoring function [17,18]. Structures' visualization and generation of molecular surfaces were performed using VIDA 4.1.1 [19].

All ssDNA receptors had a dedicated box (approximatively 21 nm^3); therefore, the VOCs were able to interact with the whole ssDNA surface; the time for processing each ssDNA conformer was about 2 min per processor (from generation of the initial 3D structures to final docking). The binding strength between the hpDNA loop and the VOCs was calculated through noncovalent interactions (hydrogen bond, van der Waals forces) by a rigid body docking approach, reducing by orders of magnitude the time machine to process large amounts of data. The maximum number of conformers was fixed at 10 for ssDNA and 200 for the 50 VOCs. The final output was the binding score average calculated for all the conformers. A bash script and a freeware BASIC-like scripting language (AutoIT V3) were used for automation and post processing of the data, respectively.

2.2. HpDNA Gas Sensors Array Setup

The samples and standards analyses were carried out using an Electronic nose UTV (Tor Vergata Sensors Group, Rome, Italy), equipped with a 11 Quartz Crystal Microbalance (QCM, 20 MHz) sensor array. QCM sensor modification was carried out using gold nanoparticles (AuNPs) functionalized with different loop sizes of hpDNA; the preparation

of the sensors was similar [9]. Briefly, AuNPs of 14 nm average diameter were synthesized using the method described in the literature [20]. Covalent functionalization with hpDNA was achieved via a C6 thiol modifier group attached to the 5′ phosphate end of the hpDNA. The AuNPs colloidal suspensions were incubated at +5 °C for 12 h with 1.406 μM of hpDNA and then centrifuged at 13,000 rpm for 30 min at 4 °C. The colorless supernatant was discarded, and the solid pellet was resuspended in 1 mL of deionized water. 5 μL of the hpDNA–AuNP suspension was deposited by drop casting on each side of the QCM and let dry for few minutes. QCM sensors were completely dried under N_2 at a flow rate of 2 L/h and stored at room temperature.

The heptamer loops (AATCAGC, CCCTGTC, CCGATTT, GCGAAGG, GTCCCTA, and GTCCGTT) were from Thermo Fischer Scientific (Italy). Output was compared with sensors obtained using pentamer loops (AAGTA, CCCGA, and TAAGT) and hexamer loops (ATAATC and CATCTG) purchased from Integrated DNA Technologies (Coralville, IA, USA).

All the selected sequences were extended with the same double helix stem of four base pair DNA attaching to the 5′ end the sequence GAAG and to the 3′ end the sequence CTTC.

2.3. Standard VOCs and Beer Samples

The 15 volatile organic compounds (VOCs) were purchased from Sigma-Aldrich (Milan, Italy). The 15 VOCs (ethanol, 1-pentanol, 1-hexanol, α-terpineol, linalool, hexanal, octanal, nonanal, ethyl acetate, ethyl-2-methylbutyrate, ethyl octanoate, butane-2,3-dione, β-caryophyllene, o-cymene, terpinolene) were of analytical grade. The analysis of VOCs was carried out using 50 μL of pure standard using a glass lab bottle (100 mL).

The 15 beer samples (craft beer) used in this study were obtained from a local micro-brewery and consisted of 5 different lots of samples taken at time zero (raw beer), after 20 days (t1) and after 40 days (t2) of fermentation. 5 mL of beer were placed in 20 mL gas-tight vials and hermetically sealed with a gas-tight septum. 15 vials were prepared for each fermentation time, in order to have three replicates (three different vials) for each measurement. Thus, a total of 45 vials have been analyzed by means of E-nose and GC-MS.

E-nose analysis of beer samples was carried out using 5 mL of beer in glass lab bottles (100 mL). The beer was kept for 5 min at 40 °C before starting the measurement to enrich the headspace of volatile compounds. The headspace was then assayed by the E-nose for 5 min at 40 °C. The frequency shift (ΔF) was taken as an analytical signal. Nitrogen (N2) was used as gas carrier.

Figure 1 shows the interaction of two different sequences of hpDNA with (3R,6Z)-3,7,11-trimethyldodeca-1,6,10-trien-3-ol. The two sequences exhibited the highest and lowest binding score for the VOC.

The volatile profile of the 15 beer samples was obtained using solid-phase micro extraction (SPME) coupled with gas chromatography–mass spectrometry (GC–MS) analysis. GC–MS analysis was carried out according to literature [21] with slight modifications. A Clarus 580 Gas Chromatography (GC) coupled with a Clarus SQ 8 Mass Spectrometer (MS) (PerkinElmer, Waltham, MA, USA) was used. The samples were kept 5 min at 25 °C and then exposed to the fiber (Divinylbenzene/Carboxen/Polydimethylsiloxane, DVB/CAR/PDMS, 50/30 μm, Supelco, Bellefonte, PA, USA) for 5 min at the same temperature. The fiber was then inserted in the desorption chamber and GC analysis was carried out using the following GC oven temperature program: 40 °C for 3 min, 5 °C/min to 150 °C, 15 °C/min to 200 °C, 200 °C for 2 min. The compounds were identified by matching the obtained spectra with the NIST Mass Spectral Library 2.0 (NIST, Gaithersburg, MD, USA) and confirmed by retention index (RIndex).
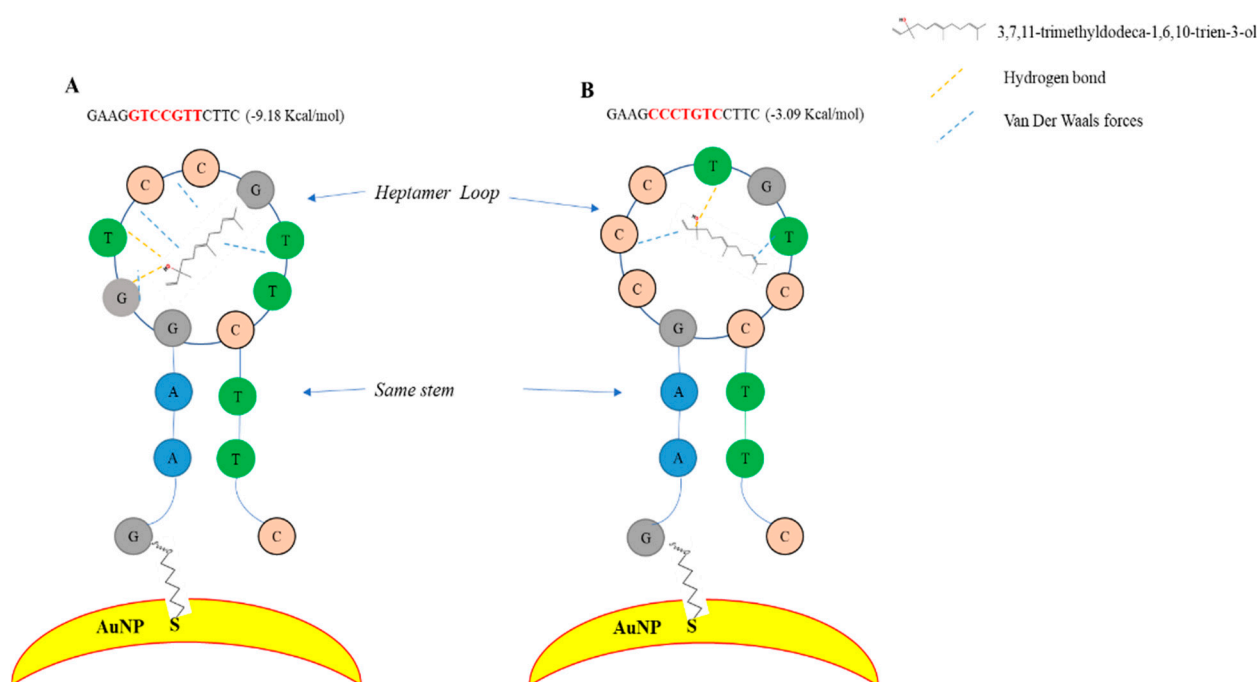
Figure 1. Sketch of the electrostatic interactions (hydrogen bonds and Van der Waals forces) between one VOC (3,7,11-trimethyldodeca-1,6,10-trien-3-ol and (**A**) the heptamer loop GTCCGTT (highest binding score = −9.18 Kcal/mol) and (**B**) heptamer loop CCCTGTC (lowest binding score = −3.09 Kcal/mol).

2.4. Statistical Analysis

XLSTAT Addinsoft USA software (XLSTAT Version 2016.02.28451) was used to explore data. Statistical significance was assessed using analysis of variance (ANOVA) with the Tukey HSD (honestly significant difference) multiple comparison analysis and the Pearson correlation test. The criterion for statistical significance of differences was p-value < 0.05 for all the comparisons. Hierarchical Cluster analysis (HCA) was applied using Ward's method setting as dissimilarity parameter of the Euclidean distance.

Multivariate statistical analysis was performed using principal component analysis (PCA), HCA, and partial least square discriminant analysis (PLS-DA) using MatLab R2011b (MathWorks, Natick, MA, USA) integrated with two toolboxes for MatLab provided by Milano Chemometrics and quantitative structure activity relationship (QSAR) Research Group [22,23]. PLS-DA was run on the dataset with optimal components for the model automatically calculated by the toolbox using a "venetian blinds" approach by dividing the dataset in cross-validation (cv) groups equal to 5 and removing them one at a time from the training set used to calibrate the PLS-DA classification model. Two classes of assignment methods were used: "Max" and "Bayes". Using "Bayes", the sample can be defined as "not assigned" based on the Bayes theorem. On the other hand, using "Max", samples are always classified based on the highest probability and the class assignment could be biased [22,24]. Before statistical procedures, all datasets were auto scaled (zero mean and unitary variance). The optimal components for the model were previously calculated using the MatLab toolbox.

3. Results

3.1. In Silico Selection of ssDNA Heptamers Loops

The ability of hpDNA to work as a binding element in gas sensors for the detection of VOCs was recently demonstrated by our group [9]; in the previous work, the performance of hpDNA having a loop of 4, 5, and 6 bases was tested and it was concluded that a key parameter for increasing the affinity of hpDNA sensors versus VOCs was the size of the DNA loop within the hairpin structure. Keeping this in mind and considering that

an increased number of bases might contribute synergistically to binding the VOCs, we calculated the simulated binding properties of the entire heptamer ssDNA library (16,384 elements) and compared the data with pentamer (1024 elements) and hexamer (4096 elements) libraries. The molecular docking functions computed hpDNA-VOCs binding using noncovalent interactions; therefore, only unpaired bases (the hpDNA loops) were screened to predict the synergic cooperation between DNA bases to interact with the VOCs target atoms.

In Figure S1A, the binding scores of pentamer, hexamer, and heptamer ssDNA libraries toward alcohols, aldehydes, esters, hydrocarbons, and ketones is reported. The binding score was reported vs. chemical group average calculated on 12 alcohols, 13 aldehydes, 17 esters, 5 hydrocarbons, and 3 ketones. The score values were calculated using the Chemgauss4 scoring function; thus, lower values represent higher ssDNA–ligand affinity. As expected from previous considerations, the ssDNA heptamer library exhibited higher ssDNA–ligand affinity for all chemical classes with respect to the other libraries. The three ssDNA oligomer libraries had a common trend with best binding scores for alcohols, followed by hydrocarbons, aldehydes, esters, and ketones. This is well represented in Figure S1B, which reports the binding score trend of the ssDNA heptamer library for the five chemical classes. A similar Gaussian distribution was observed for all chemical classes; all docking runs resulted in approximately 5% of the ssDNA-VOC complexes having higher scores and 5% with worse scores, well separated from the rest of the population. The docking output dataset of heptamers revealed a tendency of ssDNA to bind VOCs more strongly with hydroxylic groups having high molecular weight, unsaturation, and without ketone group. In particular, the best binding score of the dataset was the interaction between (3R,6Z)-3,7,11-trimethyldodeca-1,6,10-trien-3-ol and GTCCGTT.

The virtual screening output was used to select a series of ssDNA heptamer loops in order to build a hpDNA gas sensors array (E-nose) with combinatorial selectivity for VOCs.

ssDNA heptamer loops were selected by looking at the largest differences among ssDNA in binding the different chemical classes rather than the absolute "best" binding scores. This type of selection was supposed to give the E-nose the best possible performance, considering the limited selectivity achieved with the single hpDNA sensor [9]. Thus, to estimate the VOCs pattern recognition of the ssDNA heptamer library, the unsupervised multivariate algorithm PCA was applied to the data matrix of the binding scores of the 16,384 ssDNA heptamer loops for the 55 VOCs (50 VOCs plus the five chemical classes average). Figure 2 reports the scores and loading plots of the first three principal components. The first component represented 77.7% of the variance, the second 6.0%, and the third 3.2%, together displaying a cumulative variance of 86.9%.

The ssDNA heptamer loops having the smallest and largest score values on PC1, PC2, and PC3 were then selected to build the hpDNA E-nose. It should be noted that some heptamer loops were discarded due to stem-loop intramolecular base pairing. As reported in Figure 2A, PC1 axis, bringing most of the variance, contributed significatively to the separation of the six ssDNA heptamers, underlining the high degree of similarity between oligonucleotides in binding VOCs. The information needed for the ssDNA selection was mostly in the PC1. In this case, PCA was particularly useful to select ssDNA with the highest differences in binding VOC with either the highest or lowest binding vs VOC. The choice to select the hpDNA loop with the highest difference in binding VOC was made to have the highest performances in pattern recognition. Therefore, the experimental signal output obtained by hpDNA loop with low selectivity can also contribute synergistically to identifying the pattern characteristics of a given food matrix. The PC2 axis also discriminated well the sequence GCGAAGG from both CCGATTT and GTCCCTA. The variance within the heptamers library was due particularly to the interaction among alcohols and the others chemical classes. Esters, aldehydes, hydrocarbons, and ketones had very similar pattern recognition performance contributing only in spreading the heptamers on PC2 and PC3 bringing together less than 10% of the total variance. On the other hand, molecular

weight played a key role in the dataset variance. Molecules with lower molecular weight (A10 = Ethanol, A2 = (2S)-propane-1,2-diol, D9 = Acetaldehyde, E13 = Methyl methanoate, K2 = acetone) contributed mostly to the separation on the PC1 of the ssDNA heptamers. The contribution of the other components was limited. We assume that the low-variance components, which is similarly correlated to noise, do not help in selecting hpDNA loops important in the experimental part for pattern recognition, as reported by the Malinowski imbedded error [23] (Figure S2).
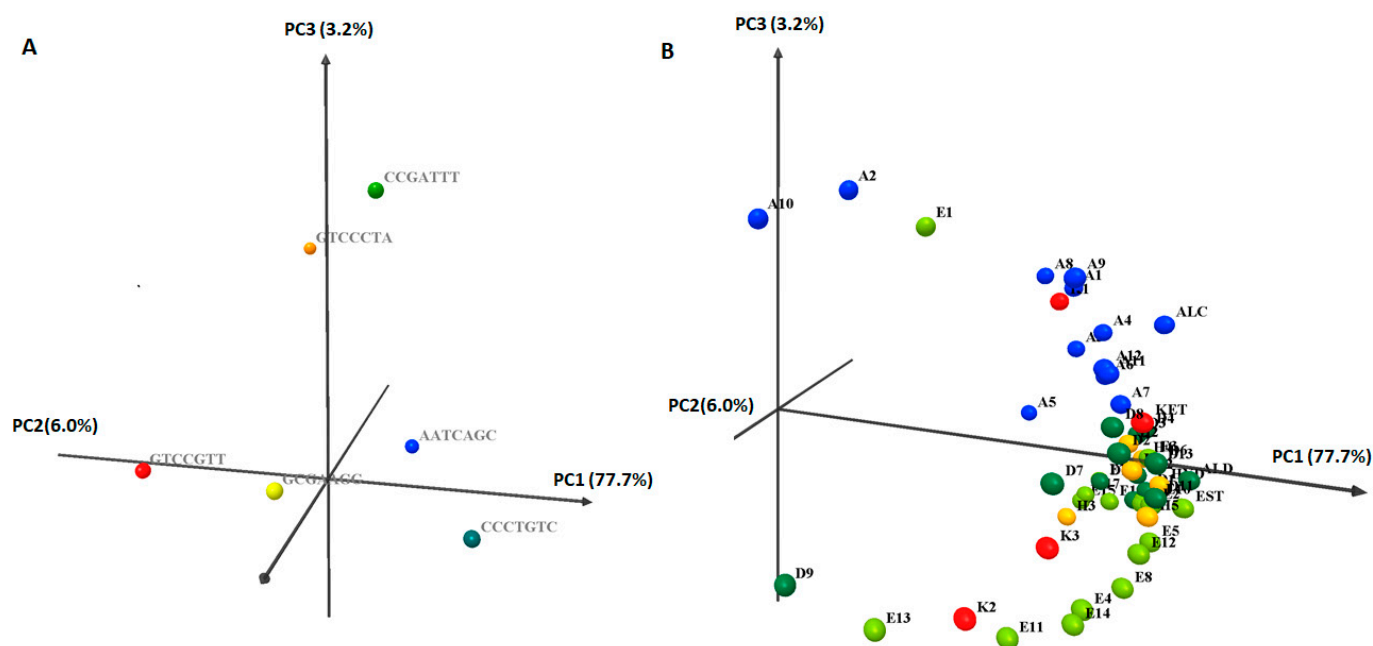

Figure 2. PCA results obtained from the matrix of 16,384 rows (the ssDNA heptamer library) and 55 columns (the 50 VOCs plus the average of the five chemical classes). (**A**) Score plot reporting only the six ssDNA heptamer sequences (having the smallest and largest score values on PC1, PC2, and PC3) used for the loop of the hpDNA gas sensor array. (**B**) Loading plot. The loadings labels are reported in Table S1; Blue points = Alcohols; Yellow points = Aldehydes; Dark green points = Esters; Light green points = Hydrocarbons; Red points = Ketones. The first three components were a variance of PC1 = 77.7%; PC2 = 6.0%; PC3 = 3.2%; total = 86.9%.

Table 1 reports the binding scores' average toward the five chemical classes and the Pearson coefficients for the six heptamers. The sequence GTCCGTT exhibited the highest affinity for all chemical classes, followed by GCGAAGG and GTCCCTA. The sequences GTCCGTT and GCGAAGG have the same trend in binding the chemical groups with a marked difference of the binding scores for alcohols-hydrocarbons and aldehydes-esters. The Pearson coefficients (Table 1B) highlight the difference among the high-binding heptamers; high correlation was in fact found for heptamers GTCCGTT and GCGAAGG, and low correlation of these two with GTCCCTA. On the other hand, the sequences AATCAGC and CCCTGTC had the lowest affinity in binding all chemical classes, except for the reasonable binding of AATCAGC toward alcohols. These two heptamers had the same trend in binding the chemical classes, as evidenced also by the correlation. The sequence CCGATTT was an average between lowest and highest binding toward the five chemical classes having a binding trend more similar to the high-binding heptamers, as highlighted by the Pearson coefficients.

Table 1. (**A**) Binding score average (Kcal/mol) of the six ssDNA heptamers used for hpDNA loops in the experimental part. The chemical group average and standard deviation was calculated over 12 alcohols, 13 aldehydes, 17 esters, 5 hydrocarbons, and 3 ketones. (**B**) The dataset of the binding score values of the 6 ssDNA aptamers toward the 50 VOCs was used to calculate the Pearson coefficients in the form of a correlation matrix. The values reported in bold were different from 0 with a significance level alpha = 0.05.

A	AATCAGC			CCCTGTC			CCGATTT			GCGAAGG			GTCCCTA			GTCCGTT		
							Binding score (Kcal/mol)											
Alcohols	−3.46	±	0.63	−2.79	±	0.42	−3.32	±	0.52	−4.94	±	1.09	−4.18	±	0.72	−6.59	±	1.63
Aldehydes	−2.72	±	0.22	−1.88	±	0.44	−3.09	±	0.52	−3.77	±	1.09	−4.36	±	0.50	−5.34	±	1.38
Esters	−2.62	±	0.31	−1.96	±	0.39	−3.16	±	0.36	−3.76	±	1.01	−4.42	±	0.35	−5.24	±	1.16
Hydrocarbons	−2.24	±	0.22	−2.06	±	0.27	−3.32	±	0.27	−4.20	±	0.60	−3.60	±	0.37	−5.99	±	0.78
Ketones	−2.23	±	0.48	−1.25	±	0.08	−2.23	±	0.14	−2.68	±	0.75	−3.77	±	0.64	−4.25	±	1.37
B	**AATCAGC**			**CCCTGTC**			**CCGATTT**			**GCGAAGG**			**GTCCCTA**			**GTCCGTT**		
AATCAGC	**1**			**0.726**			**0.357**			**0.339**			**0.381**			0.238		
CCCTGTC	**0.726**			**1**			**0.720**			**0.767**			0.168			**0.687**		
CCGATTT	**0.357**			**0.720**			**1**			**0.866**			0.147			**0.840**		
GCGAAGG	**0.339**			**0.767**			**0.866**			**1**			−0.071			**0.976**		
GTCCCTA	**0.381**			0.168			0.147			−0.071			**1**			−0.108		

The heptamer loops' selection was driven by the idea that the synergistic response in the VOCs detection of an heptamers mix with low-high VOC binding and different trend toward chemical classes should improve the discrimination ability of the (E-nose).

3.2. Sensor Array Test with Pure Compounds

The sensor array was initially challenged with pure VOCs. The hpDNA were covalently bound to the AuNPs to improve the sensitivity versus monolayer modified QCMs, as evaluated in another work [25]. Different molecules belonging to the five chemical classes were selected: alcohols (ethanol, 1-pentanol, and 1-hexanol, α-terpineol and linalool), aldehydes (hexanal, octanal, and nonanal), esters (ethyl-2-methylbutyrate, ethyl acetate, and ethyl octanoate), ketones (butane-2,3-dione), and hydrocarbons (β-caryophyllene, o-cymene, terpinolene). Table S2 reports the variation of frequency obtained by each sensor for the different pure compounds measured in the steady state (5 min).

The heptameric hpDNA sensors gave different responses. The highest signal was achieved with the loop CCGATTT that had good interaction with all VOCs, while the lowest signal was achieved with GTCCGTT. At first glance, the other loops (AATCAGC, CCCTGTC, GCGAAGG, and GTCCCTA) were more homogeneous in terms of intensity of the response.

The potential discrimination ability of the sensor array was evaluated using HCA algorithm. Figure 3 reports the dendrogram based on the steady state signal of the dataset obtained with heptamer loops. The different chemical classes analyzed were separated very well: esters, aldehydes and ketone (green line), hydrocarbons (purple line), and alcohols (orange line). These data are a good demonstration that the realized hpDNA E-nose based on heptamer loops have good ability to discriminate between different chemical classes without overlap (i.e., 1-pentanol and 1-hexanol).
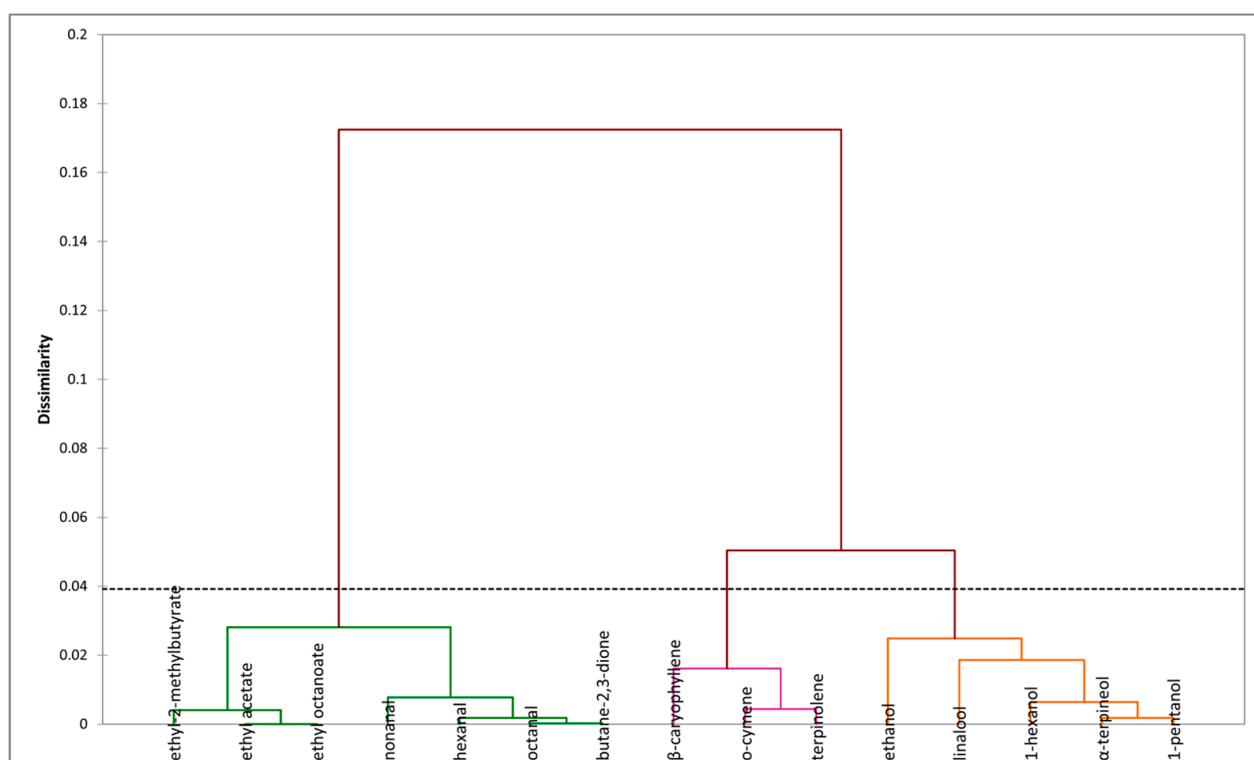

Figure 3. AHC results obtained using the response of the hpDNA gas sensors with heptamer loops by analyzing the 15 pure VOCs (5 alcohols, 3 aldehydes, 3 esters, 1 ketone, and 3 hydrocarbons).

3.3. Evaluation of the Headspace of Real Samples (Beer)

3.3.1. SPME/GC-MS Analysis

Untargeted analysis using arrays of sensors or spectroscopic techniques is generally carried out directly on samples with the aim either to classify samples or to quantify the compounds using a reference analytical method. Thus, since our purpose was to demonstrate the real applicability of the sensor array in the headspace of a challenging sample as beer, the head-space content of beer samples with SPME/GC-MS was evaluated. Twenty-three different volatile compounds were found. In Supplementary Material (Figure S3 and Table S3), typical chromatograms, for comparison of the aroma profiles, and the relative amounts of the 23 VOCs in beer samples at different fermentation times are reported.

The SPME/GC-MS analysis revealed a similar relative amount of esters in all samples (i.e., isopentyl acetate, ethyl octanoate, ethyl decanoate), while the terpenes profile decreased with fermentation time (t0 > t1 > t2). This was in agreement with a recent work by Giannetti et al. [21] on the VOCs profile of a wide range of international beers (industrial and craft products). Esters and terpenes represented the main fractions of beer flavor, and, during storage, the concentration of these compounds may decrease, bringing to a reduction of the fruity taste of the product, particularly in the industrial products. Low amounts of esters and terpenes contribute positively to the formation of the beer aroma; however, when the amount exceed the odor threshold, distinctive off-odors may occur, leading to an unacceptable taste. Beer at different fermentation times was also rich in alcohols that contribute to a strong and pungent smell and taste. One furan was also identified, certainly derived by heating of the hop.

In order to ascertain the VOCs significantly varying during fermentation, an ANOVA algorithm with a Tukey HSD (honestly significant difference) multiple comparison analysis was used. According to the data reported in Table S4, eight VOCs, one furan (3-(4-methyl-3-pentenyl)-furan), one terpene (humulene), four esters (isopentyl acetate, 1-pentyl isovalerate, ethyl octanoate and 3-methylbutyl isobutyrate), and two alcohols (2-methylpropan-1-ol and 2-phenylethan-1-ol) were significant in discriminating the three fermentation times.

3.3.2. HpDNA Gas Sensors Array Response vs Beer Samples

In previous papers, we demonstrated that gas sensors based on hpDNA arrays constituted by mixed pentamer and hexamer loops are useful for the analysis of VOCs (including esters and terpenes) present in food and plant matrices [9,11]. Thus, beer sample analysis was carried out both with a pentamer-hexamer loop array and newly realized heptamer loops hpDNA. The frequency shifts obtained are reported in Table S5 in the Supplementary Material. No particular drift of the signal was observed, and reproducibility of the measurement was satisfactory (RSD 15–25%). Humidity values, given by the sensor included in the e-nose, were constant (19–20%) during measurement of the samples.

The response of each hpDNA sensor (average calculated over three replicates) to the 15 beer samples was then correlated to the 23 VOCs detected and quantified by means of SPME/GC-MS. Correlation coefficients, reported in Table 2, may vary between −1 and 1, indicating negative/positive correlation (or poor correlation for values close to zero). The heptamer loops hpDNA correlated significantly (values in bold different from 0 with a significance level alpha < 0.05) with 6 of the 8 VOCs having a significant statistical difference among the beer fermentation times. The pentamer-hexamer loops correlated with 4 of the 8 VOCs.

Table 2. The average calculated over three replicates to the 15 beer samples at the three fermentation times of gas sensors (hpDNA with pentamer, hexamer, and heptamer loops) and VOCs identified by SPME/GC-MS analysis was used to calculate the Pearson coefficients in the form of a correlation matrix. The VOCs were ordered following statistical significance in classifying the beer samples at three fermentation times (Table S4). The Pearson coefficients computed between the hpDNA sensors were reported at the end of the correlation matrix. A significance level alpha of 0.05 was used to identify values different from 0 (values in bold).

Significant	Variables	AAGTA	CCCGA	TAAGT	ATAATC	CATCTG	AATCAGC	CCCTGTC	CCGATTT	GCGAAGG	GTCCCTA	GTCCGTT
Yes	3-(4-Methyl-3-pentenyl)-furan	**0.66**	−0.28	−0.29	−0.38	**−0.65**	−0.42	**−0.52**	**0.69**	0.10	−0.09	−0.05
Yes	Isopentyl acetate	0.26	0.44	0.41	0.26	−0.09	0.08	0.21	0.48	0.48	**0.52**	0.40
Yes	1-Pentyl isovalerate	0.41	−0.22	−0.10	−0.29	−0.47	−0.37	−0.46	0.48	0.02	−0.12	−0.10
Yes	2-Methylpropan-1-ol	**0.70**	**−0.54**	**−0.61**	**−0.61**	−0.50	**−0.58**	**−0.62**	0.57	−0.31	−0.31	−0.43
Yes	Humulene	0.49	**−0.57**	**−0.52**	**−0.52**	−0.30	−0.50	**−0.56**	0.30	−0.24	−0.51	−0.24
Yes	2-Phenylethan-1-ol	0.46	0.42	0.32	0.33	−0.24	0.28	0.14	**0.71**	**0.68**	0.38	**0.57**
Yes	Ethyl octanoate	−0.09	**0.71**	**0.61**	**0.69**	0.36	**0.55**	**0.55**	0.11	**0.58**	**0.62**	**0.58**
Yes	3-Methylbutyl isobutyrate	−0.32	−0.22	−0.22	0.02	0.40	0.10	−0.07	−0.49	−0.30	−0.39	−0.18
No	Carbon dioxide	0.30	0.11	0.18	0.02	0.14	−0.17	0.01	0.40	0.23	0.23	0.11
No	Ethyl butanoate	0.05	**−0.63**	−0.42	−0.35	−0.10	−0.35	−0.51	−0.23	−0.37	**−0.52**	−0.28
No	Styrene	−0.12	0.09	0.12	0.10	−0.02	0.24	0.14	−0.18	0.26	−0.10	0.25
No	Ethyl caproate	−0.05	0.36	0.42	0.28	0.28	0.20	0.35	0.08	0.11	0.28	0.13
No	2-Methylbutan-1-ol	0.29	**−0.58**	**−0.53**	−0.43	−0.43	−0.36	**−0.61**	−0.02	−0.39	−0.41	−0.42
No	2-Methylbutyl propionate	0.09	**−0.55**	−0.43	−0.36	0.04	−0.30	−0.35	−0.03	−0.34	**−0.56**	−0.27
No	Ethyl Acetate	0.41	−0.02	−0.14	−0.08	−0.29	0.10	0.06	**0.64**	0.20	0.07	0.10
No	Ethyl decanoate	−0.22	0.43	0.39	0.43	0.51	0.31	**0.52**	−0.07	0.31	0.49	0.34
No	(-)-β-Pinene	−0.48	−0.02	−0.01	0.33	0.26	0.30	0.06	−0.41	0.16	0.14	0.27
No	3-Methylbutyl 2-methylbutanoate	0.29	−0.35	−0.28	−0.23	−0.03	−0.09	−0.27	0.17	−0.10	−0.50	−0.12
No	2-Methylbutyl isobutyrate	−0.06	0.47	0.33	0.40	0.06	0.31	0.35	0.03	0.43	**0.57**	0.47
No	3-Methylbutan-1-ol	0.25	0.10	0.30	0.09	−0.14	0.07	0.01	0.30	0.01	−0.13	0.06
No	Isobutyl isobutyrate	−0.29	−0.12	−0.29	−0.12	0.21	−0.18	−0.15	**−0.58**	−0.22	−0.02	−0.14
No	Methylhydrazine	0.19	−0.26	−0.13	−0.33	0.06	−0.41	−0.24	0.05	−0.22	−0.34	−0.18
No	Ethanol	0.35	0.27	0.08	0.12	−0.28	0.36	0.33	**0.59**	0.30	−0.04	0.31
	AAGTA	**1.00**	−0.34	−0.49	**−0.52**	**−0.65**	−0.35	−0.43	**0.82**	0.02	−0.30	−0.11
	CCCGA	−0.34	**1.00**	**0.86**	**0.86**	0.48	**0.70**	**0.84**	−0.11	**0.62**	**0.80**	**0.70**
	TAAGT	−0.49	**0.86**	**1.00**	**0.85**	0.47	**0.60**	**0.72**	−0.18	**0.59**	**0.71**	**0.65**
	ATAATC	**−0.52**	**0.86**	**0.85**	**1.00**	0.50	**0.85**	**0.69**	−0.29	**0.59**	**0.72**	**0.69**
	CATCTG	**−0.65**	0.48	0.47	0.50	**1.00**	0.29	**0.63**	**−0.59**	0.06	0.39	0.16
	AATCAGC	−0.35	**0.70**	**0.60**	**0.85**	0.29	**1.00**	**0.69**	−0.13	0.39	0.43	0.48
	CCCTGTC	−0.43	**0.84**	**0.72**	**0.69**	**0.63**	**0.69**	**1.00**	−0.12	0.37	**0.60**	0.47
	CCGATTT	**0.82**	−0.11	−0.18	−0.29	**−0.59**	−0.13	−0.12	**1.00**	0.24	−0.11	0.08
	GCGAAGG	0.02	**0.62**	**0.59**	**0.59**	0.06	0.39	0.37	0.24	**1.00**	**0.73**	**0.96**
	GTCCCTA	−0.30	**0.80**	**0.71**	**0.72**	0.39	0.43	**0.60**	−0.11	**0.73**	**1.00**	**0.75**
	GTCCGTT	−0.11	**0.70**	**0.65**	**0.69**	0.16	0.48	0.47	0.08	**0.96**	**0.75**	**1.00**

In particular, the heptamer loops correlated with the two esters (isopentyl acetate and ethyl octanoate) having the higher relative concentration among the significant VOCs. Pentamer–hexamer loops correlated only with ethyl octanoate. Correlations were found with the alcohol 2-methylpropan-1-ol for almost all the hpDNA loops, but only three heptamers loops correlated with the other significant alcohol 2-phenylethan-1-ol. The VOC that was most changed with beer fermentation time was 3-(4-Methyl-3-pentenyl)-furan. It showed correlation with one pentamer, one hexamer, and two heptamer loops.

It should be noticed that the correlation among heptamer loops had the same trend found in VOCs binding simulation (Table 1B). The loop GTCCGTT with the highest affinity for all chemical classes showed correlation with both GCGAAGG and GTCCCTA showing, as well, high binding affinity for VOCs.

On the other hand, the sequences AATCAGC and CCCTGTC having the lowest affinity in binding all chemical classes were correlated to each other, but no correlation was found with the other heptamer loops. The sequence CCGATTT, having average binding score toward the five chemical classes, did not correlate with any of the other heptamer loops. These experimental results highlighted a synergic complementarity of the heptamers loops in binding VOCs and confirmed a convergence between the experimental and simulated data trend; they are particularly relevant since are obtained directly in sample headspace constituted by a mixture of VOCs. Finally, the different responses of these hpDNA gas sensors to the VOCs confirmed that the binding affinities did not depend on the presence of the stem that was the same for all hpDNA.

The limited selectivity of an hpDNA sensor can be exploited optimally when used in an array for pattern recognition, applying the principle of combinatorial selectivity to maximize the signal outputs in order to identify the pattern characteristics of a given food matrix by using supervised multivariate analysis with performances well beyond that of a single selective sensor.

The supervised multivariate discriminant PLS-DA algorithm was applied to evaluate the real efficacy of the hpDNA sensors array to discriminate among samples during the beer fermentation process. As in any supervised classification technique, the classes must be chosen a priori; thus, the three fermentation times were chosen as the classes for the beer samples. The two datasets (Table S5) of the two hpDNA sensors arrays (with pentamer-hexamer loops and with heptamer loops) were then processed with the PLS-DA algorithm using two class assignment methods: "Max" and "Bayes". Table 3 reports a numerical evaluation of the classification properties using the specificity, sensitivity and precision of the three groups in training and cross-validation. The real-predicted samples reported using the confusion matrix format for the cross-validation method; the 'Venetian blinds' technique (of cv groups equal to 5) is also used.

A very good discrimination among beer fermentation times for both hpDNA gas sensors arrays was achieved. Using "Max" as the class assignment method, 93% of beer samples for the pentamer-hexamer loop array and 100% for the heptamer loops array were correctly classified in cross-validation. To increase the reliability of the hpDNA arrays' discriminant performance, the results were also evaluated by using "Bayes" as a class assignment method. This method used the Bayes theorem to classify samples, reinforcing the likelihood to have an unbiased classification. In this case, the heptamer loops' array kept excellent prediction with 98% of correct beer samples classification in cross-validation, more accurately than the pentamer-hexamer loops array that lost its ability to discriminate by correctly classifying only 80% of the beer samples.

Table 3. Results of PLS-DA classification in fitting and cross-validation by using the hpDNA gas sensors array with pentamer-hexamer loops and the hpDNA gas sensors array with heptamer loops. The X matrix used was the response of the two hpDNA sensors array to the 3 replicates of the 15 beer samples (total 45 rows, data in Table S5). The Y block was a categorical matrix of 45 rows and 3 columns related to the three different fermentation times of the beer samples (t0, t1 and t2). The 'Venetian blinds' procedure (groups equal to 5) was used in cross-validation. The confusion matrix format was used to report the total accuracy real/predicted samples in cross-validation using both "Max" and "Bayes" as class assignment methods.

	hpDNA Sensors Array with Pentamer-Hexamer Loops					hpDNA Sensors Array with Heptamer Loops			
	Sensitivity	Specificity	Precision			Sensitivity	Specificity	Precision	
class	Training				class	Training			
t0	0.87	1	1		t0	1	1	1	
t1	1	1	1		t1	1	1	1	
t2	1	0.93	0.88		t2	1	1	1	
	Cross validation					Cross validation			
t0	0.87	0.97	0.93		t0	1	1	1	
t1	1	1	1		t1	1	1	1	
t2	0.93	0.93	0.88		t2	1	1	1	
	Class assignment method "max"								
real/predicted	t0	t1	t2		real/predicted	t0	t1	t2	
t0	13	0	2		t0	15	0	0	
t1	0	15	0		t1	0	15	0	
t2	1	0	14	Total	t2	0	0	15	Total
	87%	100%	93%	93%		100%	100%	100%	100%
	Class assignment method "bayes"								
real/predicted	t0	t1	t2	not assigned	real/predicted	t0	t1	t2	not assigned
t0	11	0	0	4	t0	15	0	0	0
t1	0	12	0	3	t1	0	14	0	1
t2	0	0	13	2	t2	0	0	15	0
	73%	80%	87%	Total 80%		100%	93%	100%	Total 98%

4. Conclusions

A molecular modelling method was proved to be a convenient approach to choose ssDNA sensing elements amounting to thousands of possibilities, using as key parameter for the ssDNA selection the farthest variance in binding the chemical classes. Due to the small combinations generated by only four different DNA, the size of the hpDNA loop, as previously hypothesized, should play an important role. The increase in loop size was expected to enhance the performances of the synergic cooperation between gas sensors in the array. The correlation in VOCs detection among hpDNA gas sensors with heptamer loops followed the same trend found in VOCs binding simulation, confirming a convergence between the experimental and simulated data trend. Analysis of pure compounds and treatment of the data with HCA clearly demonstrated the ability of the array to discriminate among chemical classes of VOCs. More interestingly, the hpDNA sensors array was able to discriminate among beer samples at different times of fermentation; this is particularly relevant, considering the presence of 23 different VOCs in the sample headspace. The performance of the heptamer loop hpDNA E-nose was evidently improved with respect to a pentamer-hexamer array already reported for other applications.

The data obtained confirms that hpDNA is a valuable binding element for the realization of gas sensor arrays for potential application in food area and biological relevant samples consisting of a complex mixture of VOCs. The proper selection of loop size may be crucial for the development of a new generation of DNA gas sensing devices.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article or supplementary material.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Boeker, P. On 'electronic nose'methodology. *Sens. Actuators B Chem.* **2014**, *204*, 2–17. [CrossRef]
2. Barbosa, A.J.; Oliveira, A.R.; Roque, A.C. Protein-and peptide-based biosensors in artificial olfaction. *Trends Biotechnol.* **2018**, *36*, 1244–1258. [CrossRef] [PubMed]
3. Liu, C.; Wyszynski, B.; Yatabe, R.; Hayashi, K.; Toko, K. Molecularly imprinted sol-gel-based QCM sensor arrays for the detection and recognition of volatile aldehydes. *Sensors* **2017**, *17*, 382. [CrossRef] [PubMed]
4. Mascini, M.; Gaggiotti, S.; Della Pelle, F.; Di Natale, C.; Qakala, S.; Iwuoha, E.; Pittia, P.; Compagnone, D. Peptide modified ZnO nanoparticles as gas sensors array for volatile organic compounds (VOCs). *Front. Chem.* **2018**, *6*, 105. [CrossRef] [PubMed]
5. Mascini, M.; Pizzoni, D.; Perez, G.; Chiarappa, E.; Di Natale, C.; Pittia, P.; Compagnone, D. Tailoring gas sensor arrays via the design of short peptides sequences as binding elements. *Biosens. Bioelectron.* **2017**, *93*, 161–169. [CrossRef] [PubMed]
6. Gaggiotti, S.; Della Pelle, F.; Mascini, M.; Cichelli, A.; Compagnone, D. Peptides, DNA and MIPs in Gas Sensing. From the Realization of the Sensors to Sample Analysis. *Sensors* **2020**, *20*, 4433. [CrossRef] [PubMed]
7. Shi, W.; Yu, X.; Zheng, Y.; Yu, J. DNA based chemical sensor for the detection of nitrogen dioxide enabled by organic field-effect transistor. *Sens. Actuators B Chem.* **2016**, *222*, 1003–1011. [CrossRef]
8. Kybert, N.J.; Lerner, M.B.; Yodh, J.S.; Preti, G.; Johnson, A.C. Differentiation of complex vapor mixtures using versatile DNA–carbon nanotube chemical sensor arrays. *ACS Nano* **2013**, *7*, 2800–2807. [CrossRef] [PubMed]
9. Mascini, M.; Gaggiotti, S.; Della Pelle, F.; Wang, J.; Pingarrón, J.M.; Compagnone, D. Hairpin DNA-AuNPs as molecular binding elements for the detection of volatile organic compounds. *Biosens. Bioelectron.* **2019**, *123*, 124–130. [CrossRef]
10. Gaggiotti, S.; Hurot, C.; Weerakkody, J.S.; Mathey, R.; Buhot, A.; Mascini, M.; Hou, Y.; Compagnone, D. Development of an optoelectronic nose based on surface plasmon resonance imaging with peptide and hairpin DNA for sensing volatile organic compounds. *Sens. Actuators B Chem.* **2020**, *303*, 127188. [CrossRef]
11. Gaggiotti, S.; Palmieri, S.; Della Pelle, F.; Sergi, M.; Cichelli, A.; Mascini, M.; Compagnone, D. Piezoelectric peptide-hpDNA based electronic nose for the detection of terpenes; Evaluation of the aroma profile in different *Cannabis sativa* L.(hemp) samples. *Sens. Actuators B Chem.* **2020**, *308*, 127697. [CrossRef]
12. LEXICHEM. OpenEye Scientific Software, Santa Fe, NM, Version 2.1.0. Available online: http://www.eyesopen.com (accessed on 19 May 2021).
13. SZYBKI. OpenEye Scientific Software, Santa Fe, NM, Version 1.5.7. Available online: http://www.eyesopen.com (accessed on 19 May 2021).
14. OMEGA. OpenEye Scientific Software, Santa Fe, NM, Version 2.4.6. Available online: http://www.eyesopen.com (accessed on 19 May 2021).
15. Hawkins, P.C.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inform. Model.* **2010**, *50*, 572–584. [CrossRef] [PubMed]
16. Hawkins, P.C.; Nicholls, A. Conformer generation with OMEGA: Learning from the data set and the analysis of failures. *J. Chem. Inform. Model.* **2012**, *52*, 2919–2936. [CrossRef] [PubMed]
17. OEDocking. OpenEye Scientific Software, Santa Fe, NM, Version 3.0.0. Available online: http://www.eyesopen.com (accessed on 19 May 2021).
18. Kelley, B.P.; Brown, S.P.; Warren, G.L.; Muchmore, S.W. POSIT: Flexible shape-guided docking for pose prediction. *J. Chem. Inform. Model.* **2015**, *55*, 1771–1780. [CrossRef] [PubMed]
19. VIDA. OpenEye Scientific Software, Santa Fe, NM, Version 4.1.1. Available online: http://www.eyesopen.com (accessed on 19 May 2021).

20. Frens, G. Controlled nucleation for the regulation of the particle size in monodisperse gold suspensions. *Nat. Phys. Sci.* **1973**, *241*, 20–22. [CrossRef]
21. Giannetti, V.; Mariani, M.B.; Torrelli, P.; Marini, F. Flavour component analysis by HS-SPME/GC–MS and chemometric modeling to characterize Pilsner-style Lager craft beers. *Microchem. J.* **2019**, *149*, 103991. [CrossRef]
22. Ballabio, D.; Consonni, V. Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Anal. Methods* **2013**, *5*, 3790–3798. [CrossRef]
23. Ballabio, D. A MATLAB toolbox for Principal Component Analysis and unsupervised exploration of data structure. *Chem. Intell. Lab. Syst.* **2015**, *149*, 1–9. [CrossRef]
24. Pérez, N.F.; Ferré, J.; Boqué, R. Calculation of the reliability of classification in discriminant partial least-squares binary classification. *Chem. Intell. Lab. Syst.* **2009**, *95*, 122–128. [CrossRef]
25. Compagnone, D.; Fusella, G.; Del Carlo, M.; Pittia, P.; Martinelli, E.; Tortora, L.; Paolesse, R.; Di Natale, C. Gold nanoparticles-peptide based gas sensor arrays for the detection of foodaromas. *Biosens. Bioelectron.* **2013**, *42*, 618–625. [CrossRef] [PubMed]