



MISKOLCI EGYETEM
MŰSZAKI ANYAGTUDOMÁNYI KAR
TUDOMÁNYOS DIÁKKÖRI TANÁCS



Computational study of the interaction of short DNA structures with polycyclic aromatic hydrocarbons

Author:

Barbara Söjtöry

Supervisors:

Dr. Rachid Hadjadj

Dr. Béla Fiser

Institute of Chemistry
Faculty of Materials and Chemical Engineering
University of Miskolc

Miskolc, 2022

Contents

Abstract	2
1. Introduction.....	3
2. Methods.....	6
1. Preparation of the DNA structure.....	7
2. Preparation of the ligand structures.....	7
3. The geometry of the search space	8
4. Creation of a configuration file	9
5. Docking	9
3. Results and discussion	10
4. Conclusion	15
5. Acknowledgements.....	15
6. References.....	16

Abstract

Polycyclic aromatic hydrocarbons (PAHs), usually generated during incomplete combustion of organic materials, are toxic molecules present in the environment. Their harmful effect on human health and other living organisms is a hot topic in the scientific community, knowing their mutagenic effects caused by their interaction and reaction with DNA leading directly to cancer. Although many different types of PAHs exist, 16 molecules were selected by the US Environmental Protection Agency to describe the effect of all of them. In this work, molecular docking was used in order to get a better understanding on how PAH molecules interact with DNA. The AutoDock Vina molecular docking software has been used. After noticing that all of the chosen PAHs by the US EPA are planar, 11 non-planar PAH molecules were also selected in order to see how linearity affects their reactivity. The results showed that the best fitting planar PAH molecules between the chosen human DNA spirals were the dibenz[a,h]anthracene and indeno[1,2,3-cd]pyrene, both of them having docking scores of **-8.9** kcal/mol, making them the ones having the strongest interactions with DNA and therefore, potentially the most toxic molecules. The best fitting non-planar PAH molecule is the benzo(a)perylene with the docking score of **-8.7** kcal/mol.

1. Introduction

Pollution is a major problem that the world is facing [1]. Pollutants can either be produced because of natural disasters like volcano eruptions, forest fires, or generated by human industrial and domestic activity. These can have impact on water, land and air.

Most of the time, air pollution is invisible to us. However, it can be very dangerous as it can increase the risk of getting cancer, respiratory problems, birth defects and also harm the nervous system[2]. These effects are caused by toxic molecules such as dioxin, benzene, hydrogen chloride or hydrocarbons[2].

Polycyclic aromatic hydrocarbons (PAHs) are part of a long list of hazardous compounds classified by the U.S. Environmental Protection Agency (U.S. EPA)[2]. These molecules can be produced when different organic materials, for example oil, wood, biomass or gas are burned[3]. They can cause a high impact on human health since they can cause irritation of the eyes and the breathing passages, blood and liver abnormalities, skin irritation, and much worse diseases like cancer [3]. They are also known as being the precursors of soot[4].

Researches have been made on laboratory animals where scientists have reported that certain PAHs increased occurrences of different types of cancer[5] (lungs, bladder, liver and stomach cancers). Several other researchers have looked deeper into the problem and wanted to investigate the factors of their mutagenic effect. They found out that because of their shape (similar to the DNA nucleotides), their interaction with DNA can lead to reactions leading to mutations and directly to cancer[6].

The fact that PAHs are able to bind to DNA is taken as evidence of their mutagenic potential[7]. These PAH-DNA adducts play a key role in chemical carcinogenesis[7]. Adduct formation is the result of covalent binding of their reactive metabolites[8].

The U.S. EPA identified 16 priority PAHs called parent PAHs. These species have different carcinogenic potentials[9] (**Figure 1**).

Planar PAHs

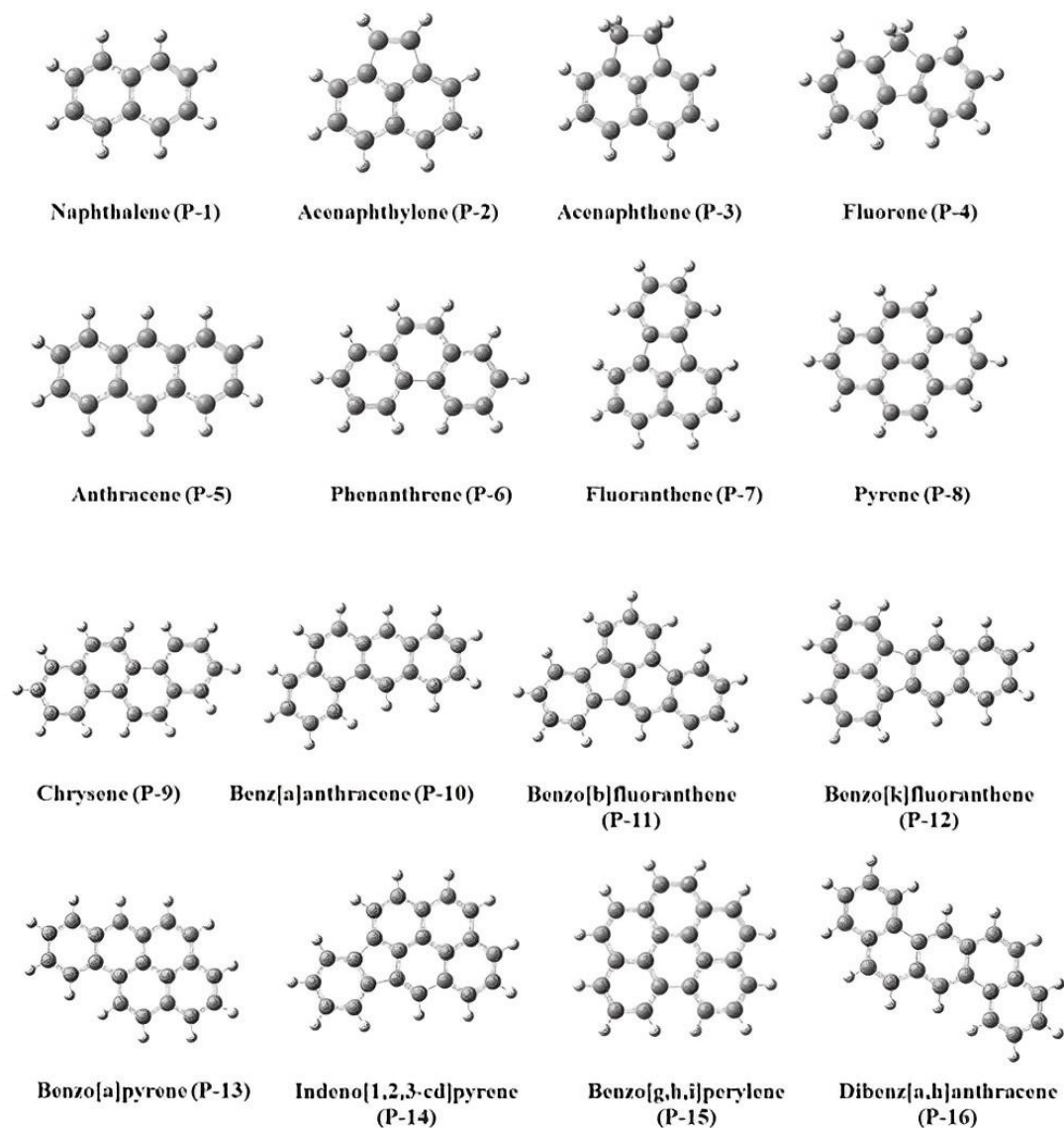


Figure 1: Chemical structures of the examined planar polycyclic aromatic hydrocarbons (PAHs). The 16 so-called parent PAHs selected by the U.S. Environmental Protection Agency (U.S. EPA). A code has been attributed to each planar PAH molecule containing a letter (P - planar) followed by a number.

The parent PAHs are assembled by the association of 2 to 6 rings and no alkyl side chains. Up until now, researchers consider that these molecules are representatives for all the PAHs[9]. In this work, the interaction between PAHs and DNA is studied. In general, the carcinogenic potency of PAHs increases with the number of aromatic rings [10]. After noticing that the selected molecules by the U.S. EPA are all planar, we decided to extend the scope of the current research and study another 11 non-planar PAHs as well (**Figure 2**).

Non-planar PAHs

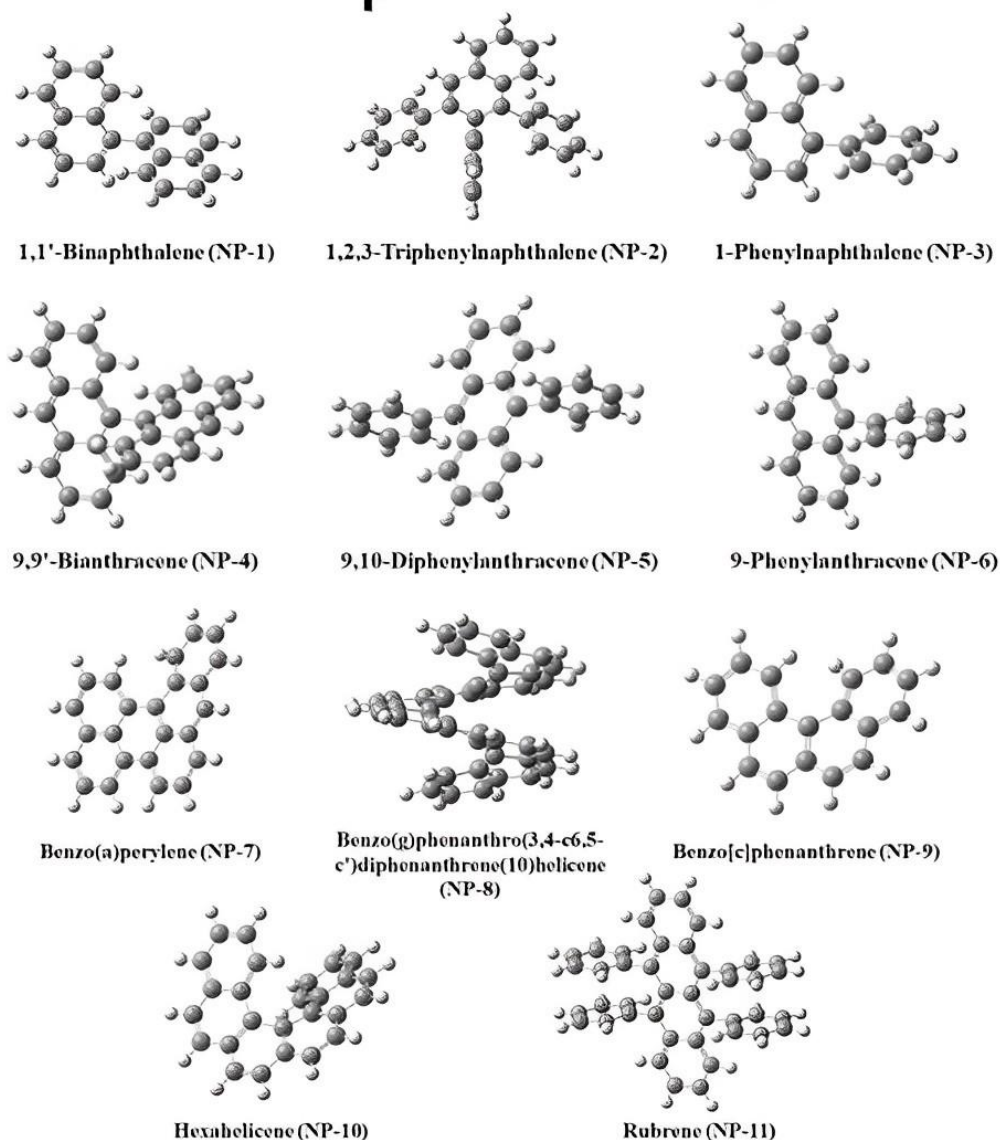


Figure 2: Chemical structures of the examined 11 non-planar polycyclic aromatic hydrocarbons (PAHs). A code have been attributed to each non-planar PAH containing two letters (NP - non-planar) followed by a number.

Several types of PAHs have been selected with degrees of torsion aiming to see the effect of the geometries in the potential interactions with DNA.

2. Methods

In this work molecular docking was applied to get a better understanding of the interaction between the PAH molecules and DNA.

Molecular docking is a method that helps to evaluate the interaction of macromolecules like proteins or enzymes with ligands. It has the ability to find the preferred position of a ligand within the macromolecule. The preferred location can be used to predict the binding affinity. The two key pieces of information in molecular docking are the geometry and energy[11].

To carry out docking calculations, there are many softwares available, in this work AutoDock Vina will be used. This software generates the possible best fitting macromolecule-ligand complexes and ranks them based on their calculated energy levels[12]. The one with the lowest energy level will be the most stable.

As a first step, to study the potential harmful effect of PAHs on human health, a human DNA structure has to be selected along with the PAHs. To acquire these files databases like the RCSB Protein Data Base (rcsb.org) [13] and Pubchem (pubchem.ncbi.nlm.nih.gov) [14] are available.

Before starting the docking process, a number of preliminary measures must be carried out (**Figure 3**).

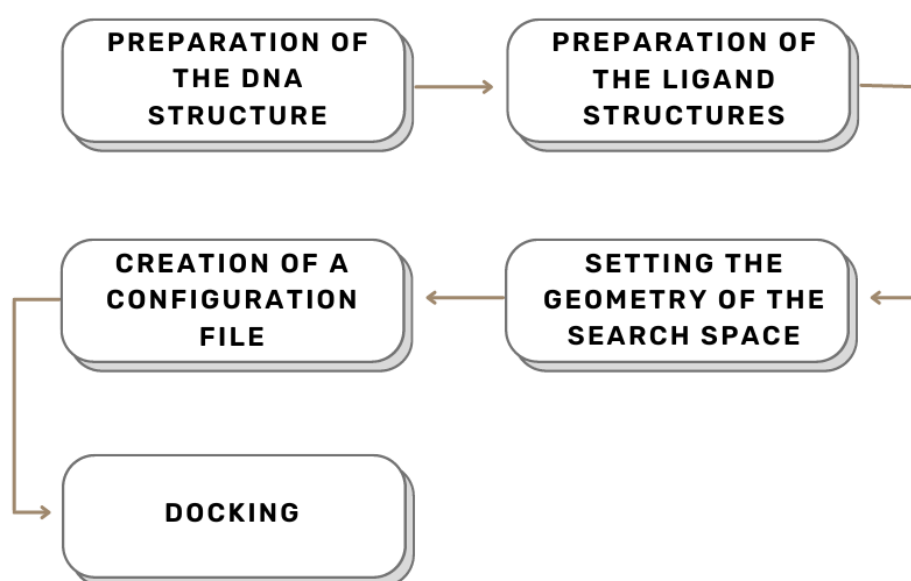


Figure 3: Workflow of the pre-docking steps.

First of all, the DNA and ligand structures have to be prepared. Then, the parameters of the search space in the DNA structure have to be defined. Since the docking is done by the computer via series of commands, a creation of a configuration file is recommended.

1. Preparation of the DNA structure

The DNA structure, which was used during the calculations (**Figure 4**), is available at the RCSB Protein Data Base, under the PDB ID of 1HF0[15].

This geometry was obtained using the X-ray diffraction method[16].

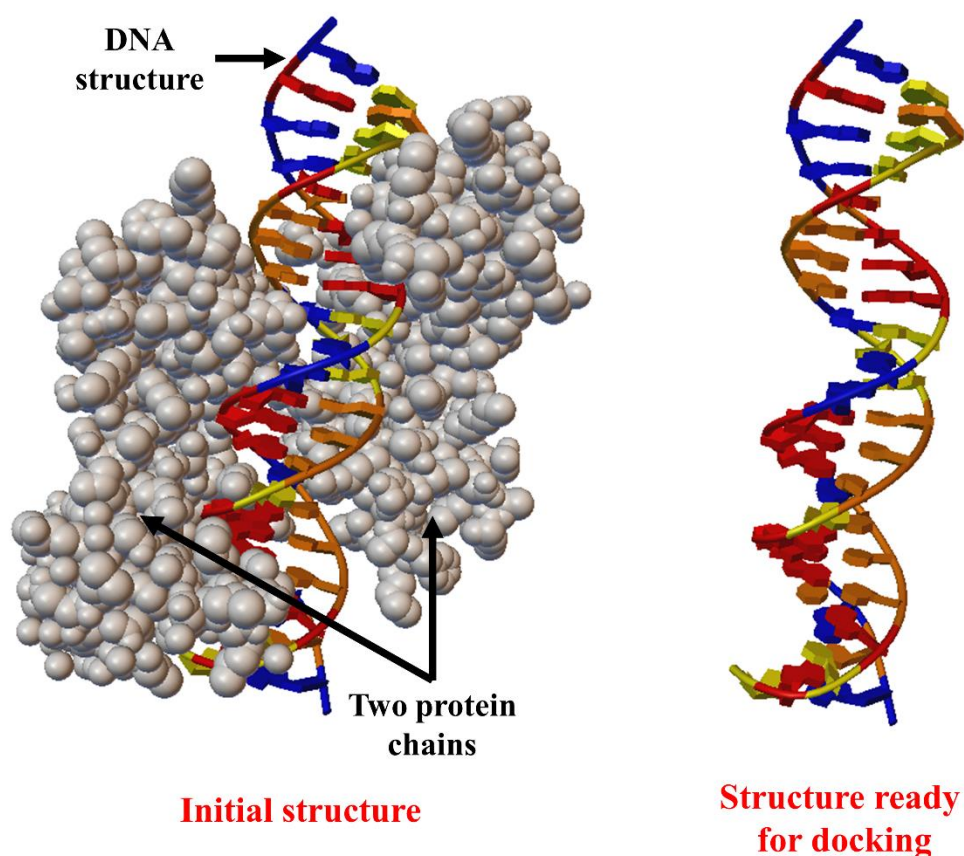


Figure 4: The DNA structure with the two initially bound protein chains (left) and the structure after the clean-up, used for docking (right).

The DNA structure is surrounded by two protein chains. These protein chains are unnecessary for the docking since we are focusing on the interactions with the DNA structure only, and therefore they have been removed, same as the water molecules. Thereafter, polar hydrogens were added and the file was saved as pdbqt file.

2. Preparation of the ligand structures

Following the preparation of the DNA, the ligand structures have to be prepared.

The 16 PAH planar molecules (**Figure 1**) have been selected from a previous study[17]. The same generated structures will be considered as ligands in our work. Regarding the non-planar

structures, the geometry of the molecules have been directly downloaded from the PubChem database[14]. The files with the geometries of the ligands are saved as pdbqt files.

3. The geometry of the search space

The geometry of the search space has to be defined. The grid box delimits a cubic volume where the docking search will be carried out. Unlike enzymes or proteins, DNA do not have a specific active site where to focus the docking search. The grid box has been shaped in a way it encompass most of the DNA chains. This will direct the software in a way that it does the search in most of the DNA. (Figure 5).

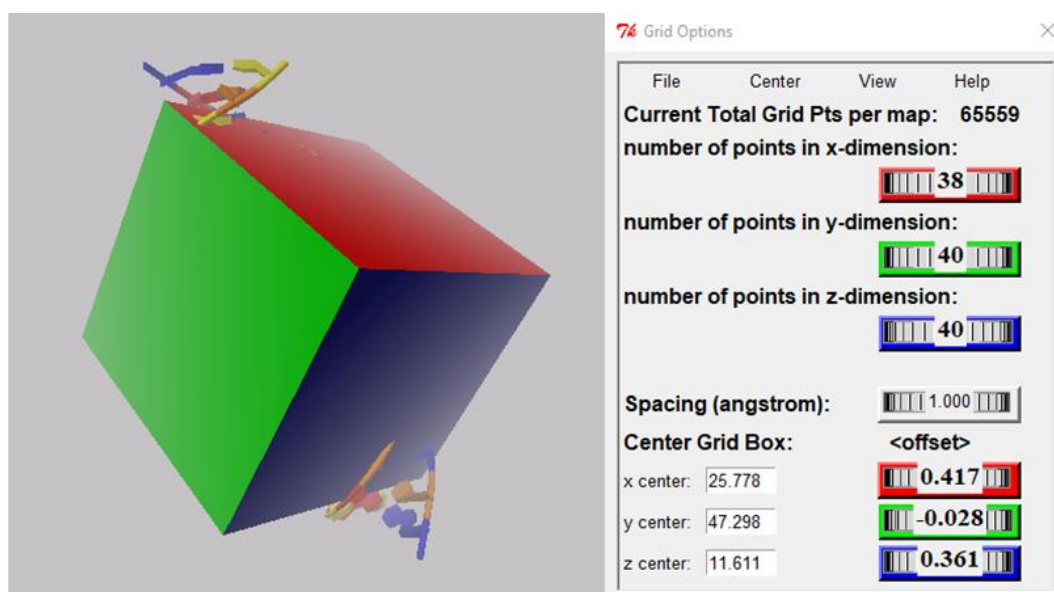


Figure 5: Coordinates of the search area (grid box).

In this step, the parameters of a cubic-like shape in a 3D coordinate system belonging to the search space has to be defined. To do this, we define the rectangles on the x, y and z axes that are the sides of the body of the box. Then, the coordinates of the center of the box has to be specified.

The distances are set in angstroms. The number of the points in the x-dimension is 38. Number of points in y- and z-dimension are both 40. The center of the grid box is set to have the x coordinate at 25.778 (Å), the y is 47.298 (Å), and the z coordinate is at 11.611 (Å).

4. Creation of a configuration file

- All the previously discussed parameters are summarized in a configuration file and in which the receptor's and ligand's pdbqt files are also defined.
- It is structured as the following:

- First, the receptor (DNA) and the ligand (PAHs) files have to be defined.
- Next, the name of the output file has to be assigned.
- and after that, the geometry of the search space should be written.
- It is optional to add the exhaustiveness parameter which defines how much the software has to insist on finding a fitting position for the ligand.

5. Docking

Since the Autodock Vina software doesn't have an executable interface, it has to be executed through the command line.

First, the location of the prepared protein (in this case DNA) and ligand pdbqt files has to be defined.

Next, the Vina software can be executed by locating the executable file in the hard drive. In the same command line, the name of the configuration file as well as the log file have to be mentioned. By pressing ENTER the docking can start.

This docking operation have been repeated for 27 ligands. At the end of the docking calculations, it can be observed that output pdbqt files and log files containing the energetics of the docking were created.

A folder where all the docking files are placed is needed. After a successful docking, the software generates all the possible positions for every PAH. The one with the lowest docking score will be considered and used in the analysis.

3. Results and discussion

The docking results can be visualized using the Autodock tools software that allows us to check the DNA structure and the docked PAHs and display the best fitting locations of the planar and the non-planar PAHs all together.

It was found that there are four locations where the molecules preferred to interact. These locations are going to be referred as 'A', 'B', 'C', and 'D'. (**Figure 6**).

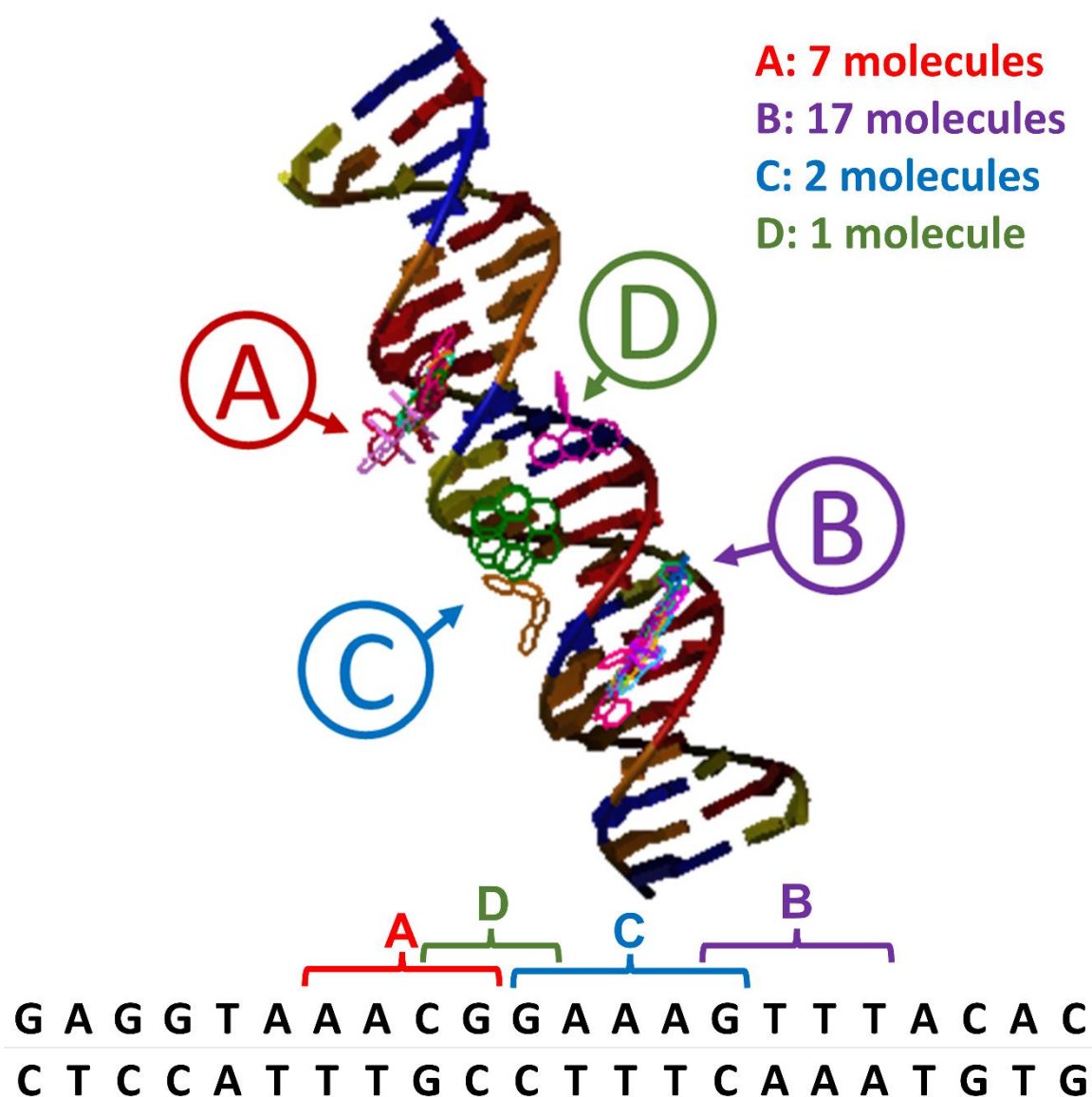


Figure 6: Best fitting position of the studied polycyclic aromatic hydrocarbons (PAHs) around the 22 nucleotide long human DNA sequence (PDB ID: *1HF0*).

The most preferred positions for both planar and non-planar molecules was **B**. 17 out of the 27 molecules ended up there (9 planar and 5 non-planar species). The next best fitting position is

A with 7 structures (4 planar and 3 non-planar molecules). 2 non-planar species were placed in **C**, while only one non-planar PAH preferred position **D**.

The planar molecules were located exclusively in positions **A** and **B**. They were generally always connected to the **B** position, with the exception of **P-8**, **P-9**, **P-10**, and **P-13**.

The non-planar species were much more dispersed and fit into all four positions. However, these types of species were also predominantly attached to site **B**, with the exception of **NP-2**, **NP-9**, and **NP-11** which were attached to the site **A**, while **NP-1** and **NP-8** were attached to site **C**, and **NP-4** was preferred **D**.

The results suggest that the most preferred locations are **B** and **A**, but as the non-planar molecules can interact with other regions of the DNA structure (**C** and **D**), these locations could also be important.

More indepth analysis related to what nucleotides sequences constitutes the docking reagions has been taken into considerations.

The docking scores along with the best fitting positions in the DNA structures are listed in **Table 1**.

Table 1: Docking score results along with the preferred fitting positions of the PAHs. Letter codes have been assigned to each molecule (the planar molecules are named P followed by the number of the molecules and the non-planar molecules are named with using NP followed with the number of the structure). The most stable molecules are highlighted in red.

PAH names	Code	Number of rings	Best fitting position	Docking score (kcal/mol)
Naphthalene	P-1	2	B	-5.0
Acenaphthylene	P-2	3	B	-5.6
Acenaphthene	P-3	3	B	-5.5
Fluorene	P-4	3	B	-6.6
Anthracene	P-5	3	B	-6.8
Phenanthrene	P-6	3	B	-6.8
Fluoranthene	P-7	4	B	-7.3
Pyrene	P-8	4	A	-6.8
Chrysene	P-9	4	A	-8.0
Benz[a]anthracene	P-10	4	A	-8.1
Benzo[b]fluoranthene	P-11	5	B	-8.3
Benzo[k]fluoranthene	P-12	5	B	-8.4
Benzo[a]pyrene	P-13	5	A	-8.5
Indeno[1,2,3-cd]pyrene	P-14	6	B	-8.9
Benzo[g,h,i]perylene	P-15	6	B	-8.3
Dibenz[a,h]anthracene	P-16	5	B	-8.9
1,1'-Binaphthalene	NP-1	4	C	-6.5
1,2,3-Triphenylnaphthalene	NP-2	6	A	-7.5
1-Phenylnaphthalene	NP-3	3	B	-6.5
9,9'-Bianthracene	NP-4	6	D	-7.7
9,10-Diphenylanthracene	NP-5	5	B	-7.6
9-Phenylanthracene	NP-6	4	B	-7.7
Benzo(a)perylene	NP-7	6	B	-8.7
Benzo(g)phenanthro(3,4-c6,5-c')diphenanthrene(10)helicene	NP-8	10	C	-8.6
Benzo[c]phenanthrene	NP-9	4	A	-7.8
Hexahelicene	NP-10	6	B	-7.8
Rubrene	NP-11	8	A	-8.3

The best fitting positions of each molecule on the DNA structure and their respective docking scores are collected. It can be noticed that the best fitting molecules in the DNA structure are the two planar species **indeno[1,2,3-cd]pyrene (P-14)** and **dibenz[a,h]anthracene (P-16)**, and the non-planar structure **benzo(a)perylene (NP-7)** with **-8.9 kcal/mol**, **-8.9 kcal/mol**, and **-8.7 kcal/mol** docking scores, respectively.

It is interesting to notice also that these three molecules prefer the position **B**, and that they are constituted of almost the same number of rings (**P-14** and **NP-7** include 6 rings, while **P-16** include 5 rings).

It can be assumed that there is a correlation between the number of rings in a molecule and the docking score (**Figure 6**).

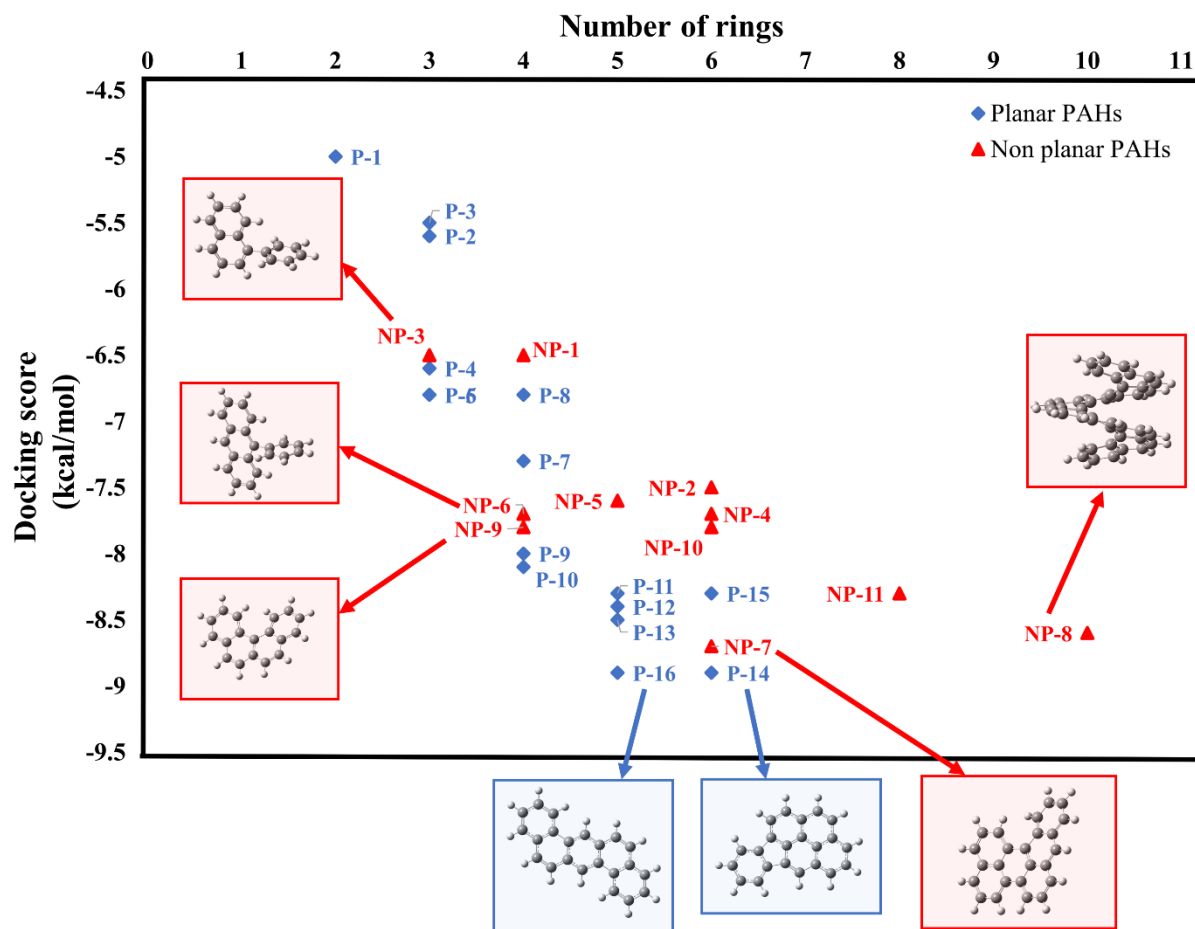


Figure 7: Docking score plotted against the number of rings constituting the corresponding planar and non-planar polycyclic aromatic hydrocarbons (PAHs).

It is easily noticeable from the **Figure 7** that the docking score decreases as the number of rings increases in both cases of planar and non-planar molecules resulting in increasingly stable complexes. However, this tendency appears to have some exceptions in the case of non-planar molecules. The most stable non-planar complex (**NP-7**) which includes a PAH with 6 rings has a docking score of **-8.7 kcal/mol**. This molecule is much smaller than the largest molecule of the study **NP-8** with 10 rings (**-8.6 kcal/mol**).

The planarity appears to make a significant difference between the molecules including the same number of rings. In the majority of cases, the docking energy of the planar species is lower than the non-planar ones. The exceptions not obeying this tendency are **NP-3**, **NP-6**, **NP-9**, and

NP-7. The reason for this is that these molecules have either planar parts (**NP-3** and **NP-6** with a 3 ring planar side), or the overall shape of the structure has a very slight torsion (**NP-7** and **NP-9**). The **NP-8** is a confirmation of this theory since it is the largest molecule in the system. Knowing that this molecule has the highest torsion in both groups which make it have a spiral shape, but it got a docking score of only **-8.6** kcal/mol which is less than what is expected from a molecule including 10 rings.

These results suggest that the spatial extent of the molecule has a greater influence on stability, and that the less distorted is the structure, the more it interacts with DNA.

The docking scores of molecules with the same number of rings also shows a difference regardless of the planarity of the molecule. But this is not only due to the effect of the geometry. Molecules with the same number of rings do not necessarily have the same number and nature of atoms, and therefore these species have to be treated as different molecules.

It was also revealed that although some molecules might have exactly the same docking score, they do not have the same number of rings nor atoms. Examples of such phenomenon are the two planar molecules **P-14** and **P-16**.

4. Conclusion

The interactions between polycyclic aromatic hydrocarbons and human DNA has been discussed studied by using the AotoDock Vina molecular docking software to understand the potential carcenogenic effect of PAHs. Calculations on 27 molecules (16 planar + 11 non-planar) have been carried out.

It was found that the two most preferred planar molecules are **indeno[1,2,3-cd]pyrene (P-14)** and **dibenz[a,h]anthracene (P-16)**, with the same docking score of **-8.9** kcal/mol. It can be assumed that these molecules are the most carcenogenic ones among all the considered structures. The third in the ranking is a non-planar structure, **benzo(a)perylene (NP-7)** which got a docking score of **-8.7** kcal/mol.

The conducted work not only confirms that the potential carcenogenic effect increases with the increasing size of the PAHs, but also with the planarity of the molecules. It has been shown that although some involved species were constructed by a high number of molecular rings, their non-planarity prevented them from having a lower docking score from what was expected.

Furthermore, this also verifies the choice of the U.S. Environmental Protection Agency (U.S. EPA) to include into the list of most important PAHs exclusively planar species.

The results showed that most of the molecules (planar and non-planar) prefer two regions (position **A** and **B**) to bind to the selected human DNA structure.

5. Acknowledgements

This research was supported by the European Union and the Hungarian State, co-financed by the European Regional Development Fund in the framework of the GINOP-2.3.4-15-2016-00004 project, which aims to promote cooperation between higher education and industry. Further support was provided by the National Research, Development and Innovation Fund (Hungary) within the TKP2021-NVA-14 project. The GITDA (Governmental Information-Technology Development Agency, Hungary) is gratefully acknowledged for allocating the computing resources used in this work.

6. References

- [1] S. Negi, S. Batoye, K. Singh, and J. S. Waraich, “Environmental Pollution, Its Causes and Impact on Ecosystem,” *New Front. Nanomater. Environ. Sci.*, pp. 1–22, 2021.
- [2] W. M. Sweileh, S. W. Al-Jabi, S. H. Zyoud, and A. F. Sawalha, “Outdoor air pollution and respiratory health: A bibliometric analysis of publications in peer-reviewed journals (1900 - 2017),” *Multidiscip. Respir. Med.*, vol. 13, no. 1, pp. 1–12, Jun. 2018.
- [3] S. Manzetti, “Polycyclic Aromatic Hydrocarbons in the Environment: Environmental Fate and Transformation,” <http://dx.doi.org/10.1080/10406638.2013.781042>, vol. 33, no. 4, pp. 311–330, Aug. 2013.
- [4] Y. C. Hsu, S. H. Chang, and M. B. Chang, “Emissions of PAHs, PCDD/Fs, dl-PCBs, chlorophenols and chlorobenzenes from municipal waste incinerator cofiring industrial waste,” *Chemosphere*, vol. 280, p. 130645, Oct. 2021.
- [5] B. G. Loganathan and P. K. S. Lam, “Global Contamination Trends of Persistent Organic Chemicals,” Aug. 2011.
- [6] L. K. Siddens *et al.*, “Polycyclic aromatic hydrocarbons as skin carcinogens: Comparison of benzo[a]pyrene, dibenzo[def,p]chrysene and three environmental mixtures in the FVB/N mouse,” *Toxicol. Appl. Pharmacol.*, vol. 264, no. 3, pp. 377–386, Nov. 2012.
- [7] S. A. Wise, L. C. Sander, and M. M. Schantz, “Analytical Methods for Determination of Polycyclic Aromatic Hydrocarbons (PAHs) — A Historical Perspective on the 16 U.S. EPA Priority Pollutant PAHs,” <https://doi.org/10.1080/10406638.2014.970291>, vol. 35, no. 2–4, pp. 187–247, Mar. 2015.
- [8] B. Ewa and M. Š. Danuta, “Polycyclic aromatic hydrocarbons and PAH-related DNA adducts,” *J. Appl. Genet.*, vol. 58, no. 3, pp. 321–330, Aug. 2017.
- [9] S. Billet *et al.*, “Genotoxic potential of Polycyclic Aromatic Hydrocarbons-coated onto airborne Particulate Matter (PM_{2.5}) in human lung epithelial A549 cells,” *Cancer Lett.*, vol. 270, no. 1, pp. 144–155, Oct. 2008.
- [10] J. Dang, X. Shi, J. Hu, J. Chen, Q. Zhang, and W. Wang, “Mechanistic and kinetic studies on OH-initiated atmospheric oxidation degradation of benzo[a]pyrene in the presence of

- O₂ and NO_x,” *Chemosphere*, vol. 119, pp. 387–393, Jan. 2015.
- [11] T. Rengarajan, P. Rajendran, N. Nandakumar, B. Lokeshkumar, P. Rajendran, and I. Nishigaki, “Exposure to polycyclic aromatic hydrocarbons with special focus on cancer,” *Asian Pac. J. Trop. Biomed.*, vol. 5, no. 3, pp. 182–189, 2015.
 - [12] Z. Liu *et al.*, “Application of molecular docking for the degradation of organic pollutants in the environmental remediation: A review,” *Chemosphere*, vol. 203, pp. 139–150, Jul. 2018.
 - [13] “RCSB PDB: Homepage.” [Online]. Available: <https://www.rcsb.org/>. [Accessed: 14-Nov-2022].
 - [14] “PubChem.” [Online]. Available: <https://pubchem.ncbi.nlm.nih.gov/>. [Accessed: 14-Nov-2022].
 - [15] A. Reményi *et al.*, “Differential Dimer Activities of the Transcription Factor Oct-1 by DNA-Induced Interface Swapping,” *Mol. Cell*, vol. 8, no. 3, pp. 569–580, Sep. 2001.
 - [16] N. Raval, R. Maheshwari, D. Kalyane, S. R. Youngren-Ortiz, M. B. Chougule, and R. K. Tekade, “Importance of Physicochemical Characterization of Nanoparticles in Pharmaceutical Product Development,” *Basic Fundam. Drug Deliv.*, pp. 369–400, Jan. 2019.
 - [17] E. Reizer, B. Viskolcz, and B. Fiser, “Formation and growth mechanisms of polycyclic aromatic hydrocarbons: A mini-review,” *Chemosphere*, vol. 291, p. 132793, Mar. 2022.