



OPEN In silico development of HASDI-G2 as a novel agent for selective recognition of the DNA sequence

Andrii Zaremba✉, Polina Zaremba & Svitlana Zahorodnia

Genetic information, which is mostly encoded in the form of DNA sequence, is the basis of life. Its deviations are often the cause of the most deadly diseases such as cancer. Accordingly, the development of methods to control the transcription of certain DNA parts is an important direction of modern pharmacological and biological research. Within the scope of this work, we are investigating the second generation of a polyintercalating agent that we developed earlier, potentially capable of recognizing 16-bp DNA sequences. In order to confirm its ability for advanced selective DNA recognition a series of simulation experiments was conducted. We differentially investigated the stability of HASDI-G2 complexes with mutated targeting sequences and their native variants. Firstly, we confirmed the ability of HASDI-G2 to clearly discriminate the target sequence (EBNA1) from a random site in the human genome (KCNH2). That repeated the experiment of the polyintercalator's previous version and additionally showed better results of the next-generation structure. Next, we examined HASDI-G2 under conditions where the target sequence differed from the random one increasingly slightly. And we found that even a one-nucleotide mismatch leads to a similar complex destabilization as a mismatch of 3 or 4 nucleotides. Such complexes showed significant conformational rearrangements, accompanied by a sharp decrease in the hydrogen bonds quantity, a drop in the binding free energy, and local melting of the DNA duplex. Moreover, the latter applied not only to sites of direct incompatibility, but also to parts where HASDI-G2 fully corresponded to the sequence of intercalation.

Keywords Targeted therapy, DNA recognition, EBNA1, BCR_ABL1, KRAS, GROMACS

Life, as we imagine it, necessarily presupposes the presence of a certain genetic material that must be realized. Generally speaking, the realization of genetic material is the process we call life, with all its diversity and inexhaustible potential. It is a complex process controlled by complex mechanisms at different levels.

However, sometimes the system fails. And in the vast majority, this leads to negative consequences. If such a violation occurred at the embryo level with its subsequent survival, we can observe the manifestation of a number of birth defects, such as Down syndrome, cystic fibrosis, phenylketonuria, and many others for our species¹. In the case of aberrations that occurred in the post-embryonic period of ontogenesis, affected the mitotic potential of the cell and are sufficiently global, we observe neoplastic transformation. That, in turn, leads to uncontrolled growth and the appearance of cancer as a disease.

Despite the fact that mankind has known about cancer since the third millennium BC, only now is cancer ceasing to be a completely incurable disease². We currently use a whole range of treatments: starting with classics, such as surgical intervention and ending with the latest achievements of target therapy. Despite this, the phenomenon of complete remission in use with II-IV stages of cancer is very rare. Only at the first stage did those methods achieve a good level of efficiency. At the same time, cases of cancer diagnosis at the first stage are still rare³.

If we examine the problem in more detail, we can see that those components of therapy that are directed at the cause of cancer, that is, at genetic information (DNA), are low-selective drugs. In particular, cytostatic agents are very common and classic components of almost every course of chemotherapy⁴. Their mechanism of action mostly consists in complicating the work of the replication machinery due to interaction with the cell DNA. At the same time, this interaction is not limited exclusively to transformed cells. Every cell of the body is a direct target for such compounds because each contains DNA. And the degree of negative impact of a conditional cytostatic agent on a cell is determined solely by its mitotic potential.

Zabolotny Institute of Microbiology and Virology of NASU, 154 Acad. Zabolotny Str., Kyiv 03143, Ukraine. ✉email: vstyp17@gmail.com

The situation is somewhat better in the case of targeted therapy. Directed influence allows us to minimize the negative consequences of side effects. However, in this case, the problem is that these drugs are mostly aimed at the manifestation of the damaged gene expression. So, the issue itself remains, only its consequences are minimized. That often leads to the rapid acquisition of resistance by the tumor to such a well-targeted drug⁵.

With this work, we continue the research started earlier in the direction of the development of a new molecular system capable of selectively recognizing only areas with a certain target sequence within the entire human genome⁶. And, accordingly, potentially devoid of most of the above-described disadvantages of chemotherapeutic drugs and approaches. Here we demonstrate that our improved successor of HASDI — HASDI-G2 is characterized by a reduction in the contribution of non-specific interactions in favor of specific ones, while maintaining a high level of stability of its complexes with the DNA duplex generated on the basis of the targeting sequence. Developing the idea, we show that HASDI-G2 is capable of recognizing the target DNA sequence not only among completely unrelated sites, but also among highly similar ones. HASDI-G2 is able to discriminate the hybrid BCR_ABL1 gene compared to individual BCR or ABL1, and even the KRAS gene (G12S) characteristic to the cancer cell culture A549, which differs from the native version by only one nucleotide substitution⁷.

Methods
Preparation of ligand/DNA complexes

The DNA sequences with a length of 50 nt were retrieved from the GenBank and literature sources and were used to generate classical Watson-Crick duplexes using the molecular editor Avogadro (Table 1)^{8,9}. The resulting double helices were further relaxed during 5 ns of molecular dynamics simulation.

To create the ligand/DNA complexes (at the both stages of iterative development and final research) indazole rings were manually intercalated between the central base pairs of the selected DNA fragments. The process also included an energy minimisation session with the use of the built-in Avogadro capabilities (UFF force field, Steepest algorithm).

Molecular dynamics simulation

The GROMACS 2019.6 software package¹⁰ was used to perform molecular dynamics simulations, with the AMBER99SB as a force field¹¹ and the TIP3P water model¹². Parametrization of the ligands was done with ACPYPE using GAFF force field¹³ and Gasteiger charge method. The system building was completed with the placement of each ligand/DNA complex in a triclinic box filled with an explicit solvent with a NaCl concentration 0.156 M, so the total system charge was equal to zero. For the stability evaluation step the distance between the box walls and the complex was kept at 18 Å and at 30 Å for the development and intercalation steps. All systems after construction underwent a system energy minimization by method of steepest descent to a value of < 1000 kJ/mol/nm (50000 steps maximum). After that, the systems were balanced in two consecutive phases - classical NVT and NPT with 100 ps in length each, with subsequent molecular dynamic simulation for 150 ns.

Most of the parameters were standardly similar for these three stages: 3D periodic boundary conditions, the time step – 2 fs, v-rescale thermostat (modified Berendsen thermostat), DNA_LIG and Water_and_ions as coupling groups with time constant of 0.1 ps for each, reference temperature was equal 300 K, LINCS algorithm for covalent bonds, with constraints for h-bonds, 1 as a number of iterations for rotational lengthening correction and highest order of 4 in the constraint coupling matrix expansion^{14,15}. Neighbor searching was applied within the cutoff scheme Verlet with frequency of the neighbor list update 20 and 1.2 nm as a cut-off distance for the short-range neighbor. At the same time, Van der Waals interactions were calculated with Lennard-Jones potential, 1.2 nm cut-off and smoothly switched the forces to zero between 1.0 nm and 1.2 nm. Electrostatic interactions were described using Fast smooth Particle-Mesh Ewald electrostatics with a 1.2 nm cut-off, interpolation order equal 4 and FFT grid spacing 0.16 nm. The main difference between the stages was in the barostat parameters of the system. In particular, the dimensions of the box were fixed for the NVT stage, the Berendsen barostat was used for the NPT, and for the MD stage - the Parrinello-Rahman barostat. Both barostats were similar by other parameters: time constant equal 2.0 ps, isotropic pressure coupling with time

No.	Target	Source	Sequence for 3D-duplex generation	HASDI-G2 integration sequence
1	EBNA1	MT164472.1	5'-TGGAGGTAGTAAGACCTCCCTTTACAACCTCAGGCGAGGAATTGCCCTTG-3' (204–253)	5'-TCCCTTTACAACCTCA-3' (221–236)
2	KCNH2	NM_000238.4	5'-AGGCGCTGCCCCGAGCCGCGGGCGCTGGAGCGGCTGTCGGCGCGGTGGCAG-3' (28–77)	5'-CGGCGCTGGAGCGGC-3' (45–60)
3	BCR	LC775148.1	5'-ACACCTTTGACCCTGGCCGCTGTGGAGTGTGTTGTGCTGTTGATGCCTTC-3' (109–158)	5'-CTGTGGAGTGTGTTGTG-3' (128–143)
4	ABL1	NG_012034.1	5'-CTTTGTATTTCATATACATTTTAGAGTGGGTTTTATCAGCTTCCATACC-3' (22856–22905)	5'-TTTTAGAGTGGGTTTT-3' (22875–22890)
5	BCR/ABL1	JQ425253.1	5'-ACACCTTTGACCCTGGCCGCTGTGGAGTGGGTTTTATCAGCTTCCATACC-3' (88–137)	5'-CTGTGGAGTGGGTTTT-3' (107–122)
6	KRAS	NG_007524.2	5'-ACTGAATATAAACTTGTGGTAGTTGGAGCTGTTGGCGTAGGCAAGAGTGC-3' (10623–10672)	5'-AGTTGGAGCTGGTGGC-3' (10643–10658)
7	KRAS-G12S	Qanqian Gao et al. [7]	5'-ACTGAATATAAACTTGTGGTAGTTGGAGCTAGTGGCGTAGGCAAGAGTGC-3'	5'-AGTTGGAGCTAGTGGC-3'

Table 1. Sequences used in the study.

constant, 1.0 bar as a reference pressure and compressibility $4.5 \times 10^{-5} \text{ bar}^{-1}$. In addition, NVT differed from the rest by generating velocity according to a Maxwell distribution at a temperature of 300 K with a generator of pseudo-random initial velocities.

All stages were carried out using the GPU as an accelerator.

Binding free energy calculation

Calculation of the binding free energy was done in the gmx_MMPBSA 1.6.1 package using the MM/PBSA (Molecular Mechanics/Poisson-Boltzmann Surface Area) method¹⁶. The trajectory generated by GROMACS 2019.6 during the simulation of target complexes was used as a source of input data. The calculation parameters input file was generated using gmx_MMPBSA 1.6.1. Most variables were left at their default values. However, there were some changes. In particular, the first 150 frames of the original trajectory were not used for analysis, and every 50th frame was selected from the rest (a total of 298 frames). At the same time, the temperature was determined at 300 K, the construction of the topology was carried out on the basis of the topology input file generated by GROMACS 2019.6, and the method of estimating the nonpolar contribution of solvation involved the calculation of its two components - cavity and dispersion terms. The surface-based integration method¹⁷ was used for the dispersion term computation. The cavity term was calculated as linearly proportional to the molecular solvent-accessible-surface area (SASA). Lastly, the ionic strength was changed to 0.156 M.

Obtained data analysis

Standard software provided by GROMACS (trjconv, rms, hbond) and gmx_MMPBSA (gmx_MMPBSA_ana) was used for analysis. Visualization was performed with XMGRACE¹⁸ and Pymol 1.8.

Iterative development of HASDI-G2

The molecular dynamics simulation studies were done for each HASDI/EBNA1-50nt complex. After that, the ligand was modified in the direction of increasing the contribution of its specific component and reducing the non-specific one, as well as improving the relative selectivity. At the same time, the preservation of the complex stability at a level close to one for HASDI was carefully controlled⁶. Each modification was coupled with optimization of the complex geometry by performing a short energy minimization session using the built-in Avogadro functional. The following step was the molecular dynamics simulation of an already modified ligand with the target DNA. The analysis of the simulation results was carried out and the cycle was repeated up until (in our opinion) the affinity level of HASDI to the target sequence was sufficiently improved compared to other sequences of the genome. Ultimately based on HASDI, HASDI-G2 was created as its successor with improved characteristics.

Results

HASDI-G2: general principles

Like HASDI, its successor HASDI-G2 consists of eight segments (units) and each unit is capable of recognising 2 base pairs (Fig. 1)⁶. Accordingly, the total recognition capability of the intercalator is 16 pairs of nucleic bases.

HASDI-G2 DNA sequence recognition is also implemented based on the same features as its predecessor. Actually, the scaffold section of the linker part of each ligand unit did not undergo any changes at all. δ -valerolactone and 2-methylbutyl connected by a covalent C-C bond can be modified (depending on the sequence) in positions 1 and 2 of the methylbutyl aliphatic chain (R_3 and R_4 , respectively) as well as positions 5

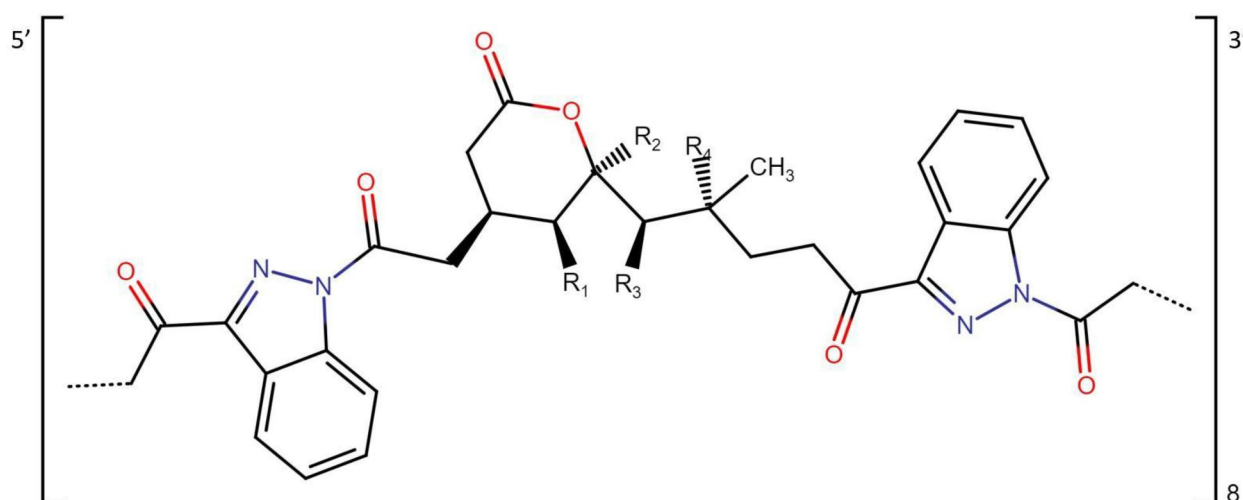


Fig. 1. General view of HASDI-G2 structure presented as an elementary unit that contains a branched linker where R_1 , R_2 , R_3 and R_4 are radicals directly recognizing the targeting sequence, and 2 indazole rings intercalating in between every 2 base pairs. The whole molecule consists of 8 such units for a total recognition of 16 base pairs.

and 6 of the δ -valerolactone ring (R_1 and R_2 , respectively). That provides recognition of two base pairs based on their hydrogen bond donors and acceptors directed towards the major groove of the DNA duplex. However, the list of substituents for R_1 , R_2 , R_3 , and R_4 used in this work is somewhat different and more complete compared to the research that was related to HASDI. Substituents, as well as principles of their combination, are presented in Table 2.

Additionally to highly sequence-specific linker regions, HASDI-G2 contains indazole rings located at both ends of each linker that connect a whole molecule like interlinker bridges. Planar indazole rings easily get involved in hydrophobic and Van der Waals interactions on account of a low polarization degree and the presence of a conjugated π -system. That assures their non-specificity for DNA as a double-stranded helix by the classical mechanism for DNA-intercalators¹⁹. At the same time, compared to HASDI, which contains phenazine rings in its composition, the indazole residues are significantly smaller in size, which, accordingly, reduces the overall hydrophobicity of HASDI-G2 and shifts the contribution of each energy component towards their selective share.

In addition, due to the increase in the number of HASDI-G2 variations and the related more urgent need for their clear identification, within the scope of this work we developed a simple scheme for forming the name of a specific HASDI-G2. We suggest adding to “HASDI-G2” in parentheses without a space the generally accepted short name of the gene targeted by the specific HASDI-G2. For example, HASDI-G2(EBNA1) targets a sequence within the EBNA1 gene.

Stability of HASDI-G2 interaction with EBNA1-50nt and KCNH2-50nt

In this part of the study, HASDI-G2 was directed to interact specifically with EBNA1-50nt: the sequence of substituents in positions R_1 , R_2 , R_3 and R_4 corresponded to the duplex generated on the basis of the sequence 5'-TCCCTTTACAACCTCA-3' (HASDI-G2(EBNA1)). To understand the difference between the HASDI-G2(EBNA1)/target complex and a random place in the genome, the developed structure was intercalated into a DNA duplex that corresponds to the sequence 5'-CGGGCGCTGGAGCGGC-3', which is a part of the KCNH2 gene.

Molecular dynamics simulation of the HASDI-G2(EBNA1)/EBNA1-50nt complex

Throughout the entire simulation, the HASDI-G2(EBNA1)/EBNA1-50nt complex demonstrated a firm stability.

Stacking interactions of all indazole rings undergone no changes during simulation: each intercalating planar structure never changed its original location, i.e. at a distance of 2 base pairs from one another (Fig. 2).

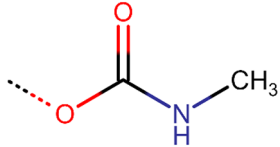
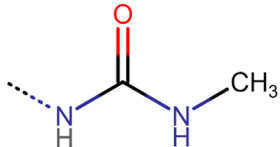
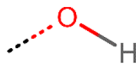
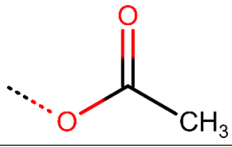
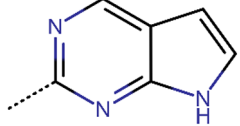
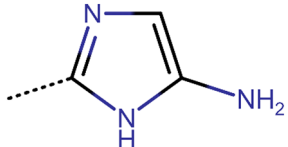
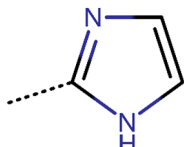
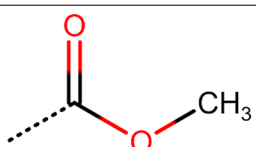
Hydrogen bonds (HBs), as a selective component of the HASDI-G2(EBNA1) interaction with EBNA1-50nt, ranged from 30 to 42 (Fig. 3) with an average value of 36 HB per complex. This is a better number compared to the predecessor in a similar complex (an average of 32 HBs). Accordingly, generally, one linker, which is theoretically capable of recognizing 2 base pairs by forming 6 HBs, interacts with DNA at 4–5 points within this simulation without a tendency to destabilize the interaction.

The high level of stability of the obtained conglomerate is also confirmed by the energy of complexation (ΔG) calculated on the basis of the simulation trajectory by the MM/PBSA method: $\Delta G = -157.72 \pm 6.57$ kcal/mol (Table 3). Nevertheless, the enthalpy (ΔH) of this interaction was equal to -177.31 ± 6.57 kcal/mol, and the entropy penalty ($-T\Delta S$) was 19.59 ± 0.05 kcal/mol. Among the enthalpy components, the stabilizing contribution was made by the Van der Waals forces (Evdw), electrostatic interactions (Eel) and the nonpolar component of solvation energy (Enpol): -395.43 ± 6.72 kcal/mol, -76.33 ± 3.55 kcal/mol and -198.88 ± 2.62 kcal/mol, respectively. Instead, the contributions of the polar energy of solvation (Epb) and the energy of dispersion forces (Edisp) were destabilizing: 140.79 ± 3.59 kcal/mol and 352.54 ± 3.52 kcal/mol, respectively. The significant contribution of the Van der Waals and dispersion forces is obviously related to the relatively large area of the indazole rings in the composition of HASDI-G2(EBNA1). Instead, the input of electrostatic interactions is primarily a consequence of the specific interaction of ligand linker regions with nucleic bases.

Molecular dynamics simulation of the HASDI-G2(EBNA1)/KCNH2-50nt complex

The molecular dynamics simulation of HASDI-G2(EBNA1) in a complex with a random DNA sequence, which was represented by a fragment of the KCNH2 gene, was significantly different from that in the case of the “correct” complex. Already after 6 ns of simulation, the rapid increase in destabilization of the 5'-end of the ligand led to the terminal indazole ring move outside the duplex and, accordingly, to the loss of any orientational interactions with the nucleobases by the corresponding linker. During the next 20 ns, the released part of HASDI-G2(EBNA1) stabilized in an intermediate form. In that form the indazole ring was devoid of stacking interactions with nucleobases, yet was located in a major groove next to the linker of the second segment of the ligand. This configuration lasted until the 80th ns of the simulation. After that, a repeated change in conformation was observed. This time it was also related to the DNA duplex in the area where the first and second HASDI-G2(EBNA1) linkers are located. This was expressed in the local melting of DNA with subsequent exit of C30 of the antisense chain outside the duplex and stabilization of its interaction with the ligand's indazole ring, which was also located outside its original position (Fig. 4). This conformation was further stabilized by two equivalent intramolecular HBs between the R_1 hydroxyl intended for recognition of the sense chain thymine and the oxygen of the keto group on one side and the oxygen of the R_3 substituent intended for cytosine recognition on the other. A similar placement of the ligand and DNA in this area was maintained until the end of the simulation.

In addition to significant destabilization of the DNA duplex in the region of the 5'-end of the ligand, local DNA melting was similarly observed in the area of interaction of the HASDI-G2(EBNA1) 5th linker. After 60 nanoseconds of stable Watson-Crick interaction, G26 of the sense chain almost completely left the duplex. The only remaining HB was represented by the interaction of O6 guanine and 4NH₂ cytosine. This conformation

Radical	The recognized nucleobase	Substituent
R ₁	Adenine (A)	
	Guanine (G)	
	Thymine (T)	
	Cytosine (C)	
R ₂	Adenine (A)	
	Guanine (G)	
	Thymine (T)	
	Cytosine (C)	
Continued		

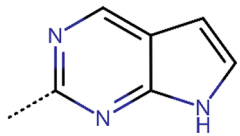
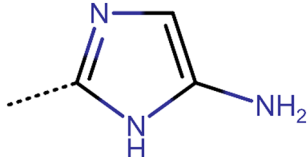
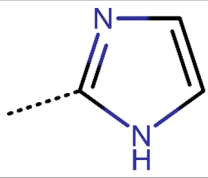
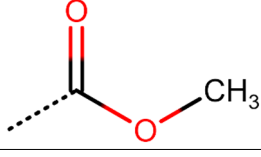
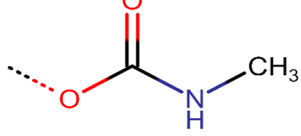
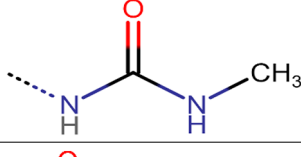

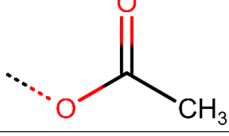
Radical	The recognized nucleobase	Substituent
R ₃	Adenine (A)	
	Guanine (G)	
	Thymine (T)	
	Cytosine (C)	
R ₄	Adenine (A)	
	Guanine (G)	
	Thymine (T)	
	Cytosine (C)	

Table 2. R₁, R₂, R₃ and R₄ of HASDI-G2 linkers and principles of their combination for recognition of target DNA sequence.

of G26 was present until the end of the simulation and was further stabilized by a non-canonical HB between 2NH₂ of guanine and the oxygen of phosphate acid in the negatively charged DNA backbone. At the same time, the indazole ring flanking this linker from the 5'-end was significantly shifted in the direction from G26. This conformation was also stably observed until the end of the simulation.

According to the significant and progressive conformational changes the HB pattern, as the basis of the specific interaction of HASDI-G2 with a certain DNA sequence, was characterized by a significant decrease in their number from the very beginning of the simulation (Fig. 5).

In particular, at the beginning of the simulation, the average HBs number between HASDI-G2(EBNA1) and KCNH2-50nt was 24. Starting from the 8th ns and up to the 140th ns their quantity decreased to 20. At the end of the simulation, an even greater drop was observed – up to 19 intermolecular HBs. Accordingly, an average of 2.5 HBs were present at the level of each individual linker out of 6 possible. This was also confirmed by the analysis of the simulation trajectory. Only in the case of random matches between the EBNA1-50nt and KCNH2-50nt sequences (A28 and C30) did we observe the presence of all HBs. In the remaining cases, only random donor-acceptor interactions were present (1–3 per linker). The first linker after leaving the 5'-indazole had no orienting interactions with the DNA duplex.

The relatively low level of stability of the obtained complex is also confirmed on the basis of the binding free energy calculation, which in this case was equal to -97.25 ± 9.18 kcal/mol (Table 4). At the same time, the

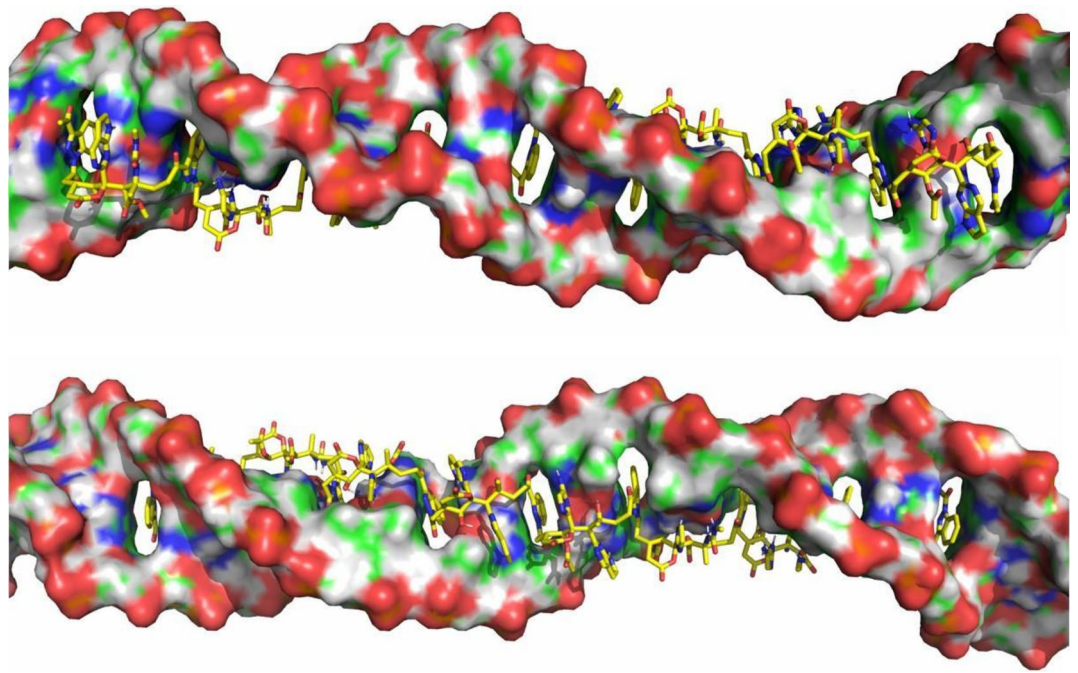


Fig. 2. General view of HASDI-G2(EBNA1) complex with EBNA1-50nt during molecular dynamics simulation (representative trajectory frame).

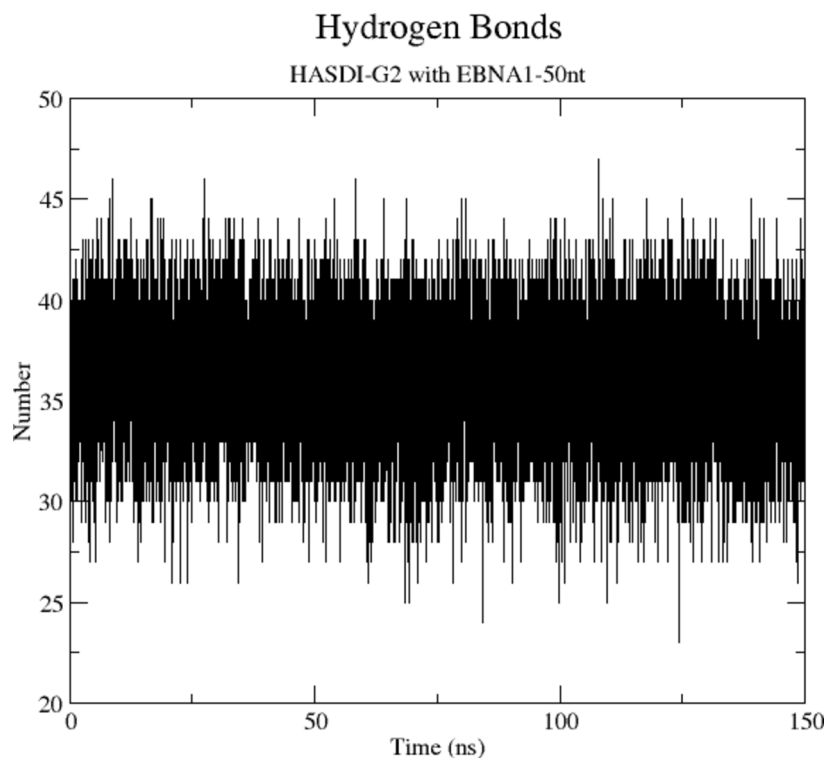


Fig. 3. Dependence of the hydrogen bonds quantity on the molecular dynamics simulation time in the HASDI-G2(EBNA1)/EBNA1-50nt complex.

enthalpy of complexation was -139.65 ± 9.18 kcal/mol, and the entropy penalty was 42.39 ± 0.05 kcal/mol. Among the enthalpy components, as in the case of the HASDI-G2(EBNA1)/EBNA1-50nt complex, the stabilizing contribution was made by Van der Waals forces, electrostatic interactions, and the nonpolar component of solvation energy: -356.74 ± 11.89 kcal/mol, -53.89 ± 5.17 kcal/mol and -180.53 ± 5.76 kcal/mol, respectively.

	Evdw	Eel	Epb	Enpol	Edisp	ΔH	-TAS	ΔG
Average	-395.43	-76.33	140.79	-198.88	352.54	-177.31	19.59	-157.72
SD	6.72	3.55	3.59	2.62	3.52	6.57	0.05	6.57

Table 3. MM/PBSA calculated binding energy of HASDI-G2(EBNA1) and EBNA1-50nt expressed in kcal/mol.

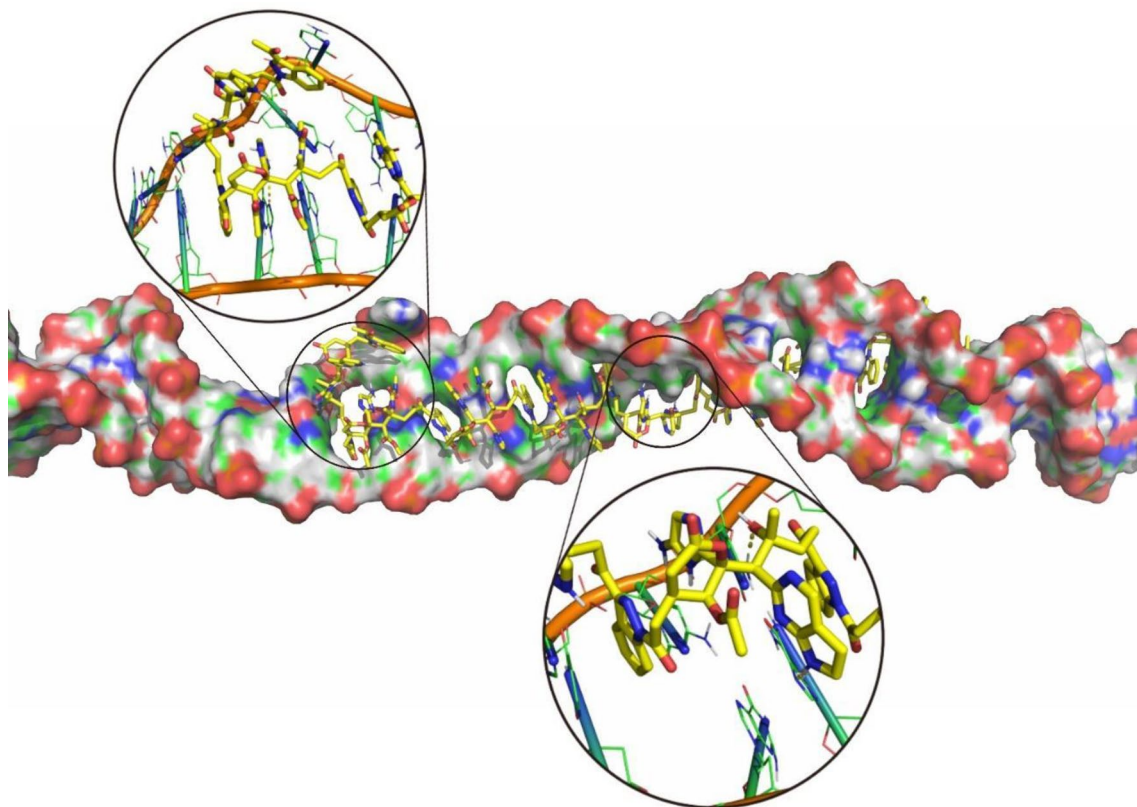


Fig. 4. General view of HASDI-G2(EBNA1) complex with KCNH2-50nt during molecular dynamics simulation (representative trajectory frame).

Instead, the contributions of the polar energy of solvation and the energy of dispersion forces were destabilizing: 123.08 ± 5.88 kcal/mol and 328.43 ± 8.74 kcal/mol, respectively. At the same time, the biggest difference in energy components compared to the original complex was observed in the case of electrostatic interactions (-29.4%). They best reflect the contribution of a specific component of the HASDI-G2(EBNA1)/DNA interaction.

Stability of interaction of HASDI-G2 with BCR/ABL1-50nt, BCR-50nt and ABL1-50nt

In this part of the study, HASDI-G2 substituents in positions R_1 , R_2 , R_3 and R_4 corresponded to the sequence of a part of the BCR_ABL1 fusion gene with a length of 50 nucleotides. Accordingly, HASDI-G2 was targeted to recognize the DNA duplex generated on the basis of the sequence 5'-CTGTGGAGTGGGTTT-3' (HASDI-G2(BCR_ABL1)) as a sense chain. To understand the degree of selectivity of HASDI-G2(BCR_ABL1), in addition to simulating its interaction with the targeting sequence, the stability of HASDI-G2(BCR_ABL1)/BCR-50nt and HASDI-G2(BCR_ABL1)/ABL1-50nt complexes was also investigated. Moreover, the immediate zone of intercalation was always located directly at the point of fusion of BCR with ABL1. That is, HASDI-G2(BCR_ABL1) in any case at least partially corresponded to the sequence of the intercalation site.

Molecular dynamics simulation of the HASDI-G2(BCR_ABL1)/BCR_ABL1-50nt complex

The specific selection of substituents carried out according to Table 2 made it possible to target HASDI-G2 to a different sequence, despite significant changes in it.

During the entire simulation period, the HASDI-G2(BCR_ABL1)/BCR_ABL1-50nt complex maintained the initial configuration of both the ligand and DNA (Fig. 6). All indazole rings were located directly in the intercalation zone, that is, after every two base pairs. Each linker formed numerous HBs with hydrogen bond donors and acceptors located in the major groove of the duplex. And each base pair of DNA-duplex was always within the double helix and was stabilized by classic Watson-Crick interactions.

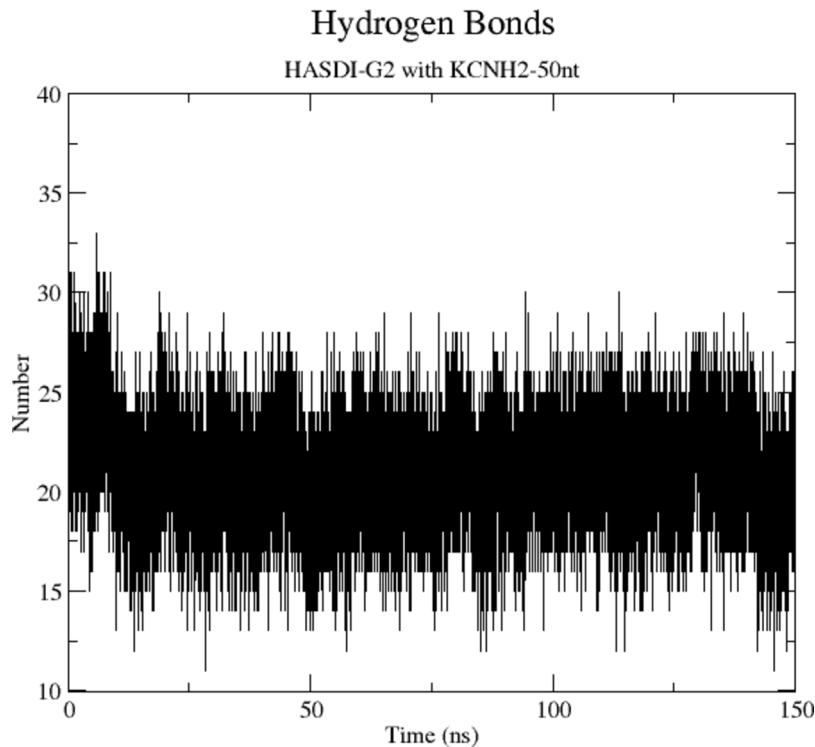


Fig. 5. Dependence of the hydrogen bonds quantity on the molecular dynamics simulation time in the HASDI-G2(EBNA1)/KCNH2-50nt complex.

	Evdw	Eel	Epb	Enpol	Edisp	ΔH	-TΔS	ΔG
Average	-356.74	-53.89	123.08	-180.53	328.43	-139.65	42.39	-97.25
SD	11.89	5.17	5.88	5.76	8.74	9.18	0.05	9.18

Table 4. MM/PBSA calculated binding energy of HASDI-G2(EBNA1) and KCNH2-50nt expressed in kcal/mol.

The quantity of hydrogen bonds when targeting HASDI-G2 to BCR_ABL1-50nt was slightly lower than in the case of its targeting EBNA1-50nt and averaged 34 HBs (Fig. 7). That corresponds to 4.25 selective interactions per linker and is a higher value compared to the previous version of HASDI. It should be noted that from the 1st to the 7th ns of the simulation, the number of HBs was lower and fluctuated around 31 per complex. However, the remaining simulations were characterized by exactly 34 interactions.

The degree of system stability was also confirmed by the calculation of its binding free energy, which in this case was equal to -154.95 ± 6.93 kcal/mol. That is only 2.77 kcal/mol less compared to the HASDI-G2(EBNA1)/EBNA1-50nt reference complex. A similar situation is observed in the case of other energy components (Table 5). At the same time, all stabilizing forces (Van der Waals, electrostatic interactions, nonpolar component of solvation energy) have more negative values compared to those for the reference complex. The opposite is the case of destabilizing forces (the polar energy of solvation and the energy of dispersion interactions). Both of these enthalpy components had a more positive value, moreover Epb differed by 13.52 kcal/mol. All in all, this led to a slightly lower enthalpy of complex formation and, accordingly, a lower binding energy.

Molecular dynamics simulation of the HASDI-G2(BCR_ABL1)/BCR-50nt complex

Molecular dynamics simulation trajectories of HASDI-G2(BCR_ABL1) in complex with native BCR-50nt and “correct” BCR_ABL1-50nt were significantly different (Fig. 8).

As expected, at the beginning of the simulation the main changes concerned the region of the ligand after the 5th linker, which was in direct contact with BCR-50nt below the point of fusion with ABL1. Only in the regions of random single-nucleotide overlap between BCR and ABL1 (T32 and T34) was observed the preservation of all three HBs characteristic for the recognition of one pair of nucleotides. Accordingly, the part of the ligand that provided recognition of BCR to the point of fusion was characterized by the preservation of all classical HBs for the HASDI-G2(BCR_ABL1)/target complex. However, during the course of the simulation, significant conformational rearrangements of both HASDI-G2(BCR_ABL1) and the DNA duplex were observed. Already after 4 ns of the simulation, the third linker of HASDI-G2(BCR_ABL1) lost almost all HBs with the base pairs that it was actually targeted to recognize. Subsequently, this area was characterized by low stability of the DNA

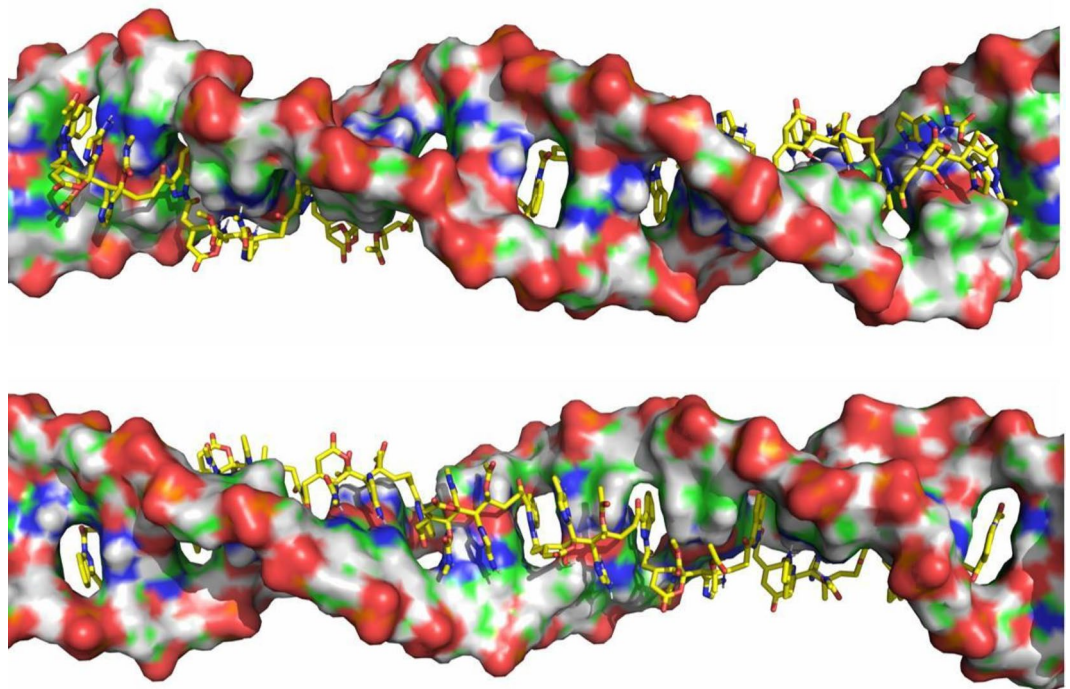


Fig. 6. General view of HASDI-G2(BCR_ABL1) complex with BCR_ABL1-50nt during molecular dynamics simulation (representative trajectory frame).

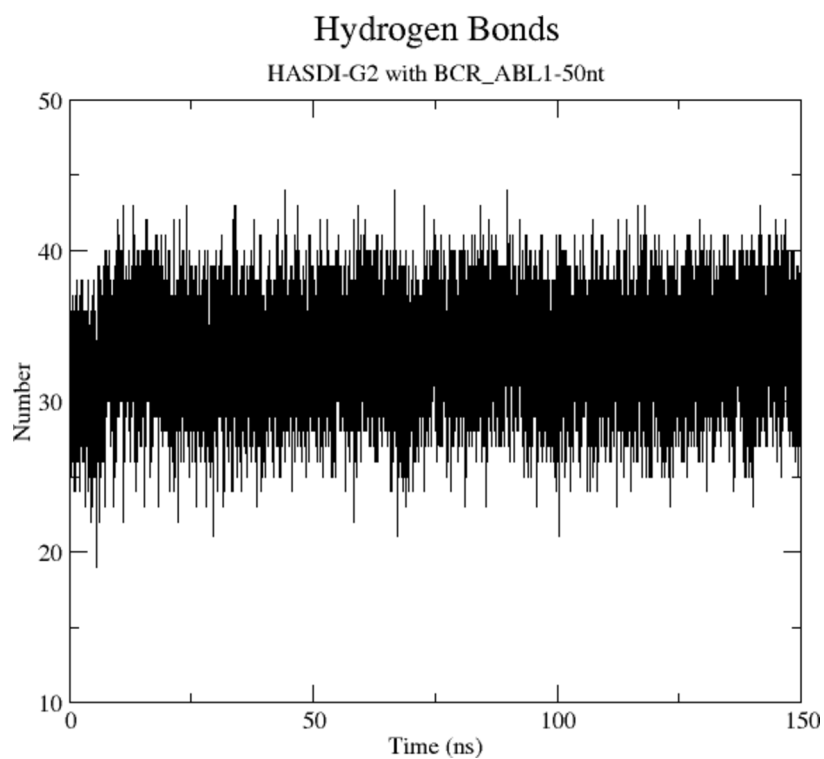


Fig. 7. Dependence of the hydrogen bonds quantity on the molecular dynamics simulation time in the HASDI-G2(BCR_ABL1)/BCR_ABL1-50nt complex.

	Evdw	Eel	Epb	Enpol	Edisp	ΔH	$-T\Delta S$	ΔG
Average	-399.6	-80.36	154.31	-202.17	354.68	-173.15	18.2	-154.95
SD	6.91	3.44	3.68	2.77	3.58	6.93	0.05	6.93

Table 5. MM/PBSA calculated binding energy of HASDI-G2(BCR_ABL1) and BCR_ABL1-50nt expressed in kcal/mol.

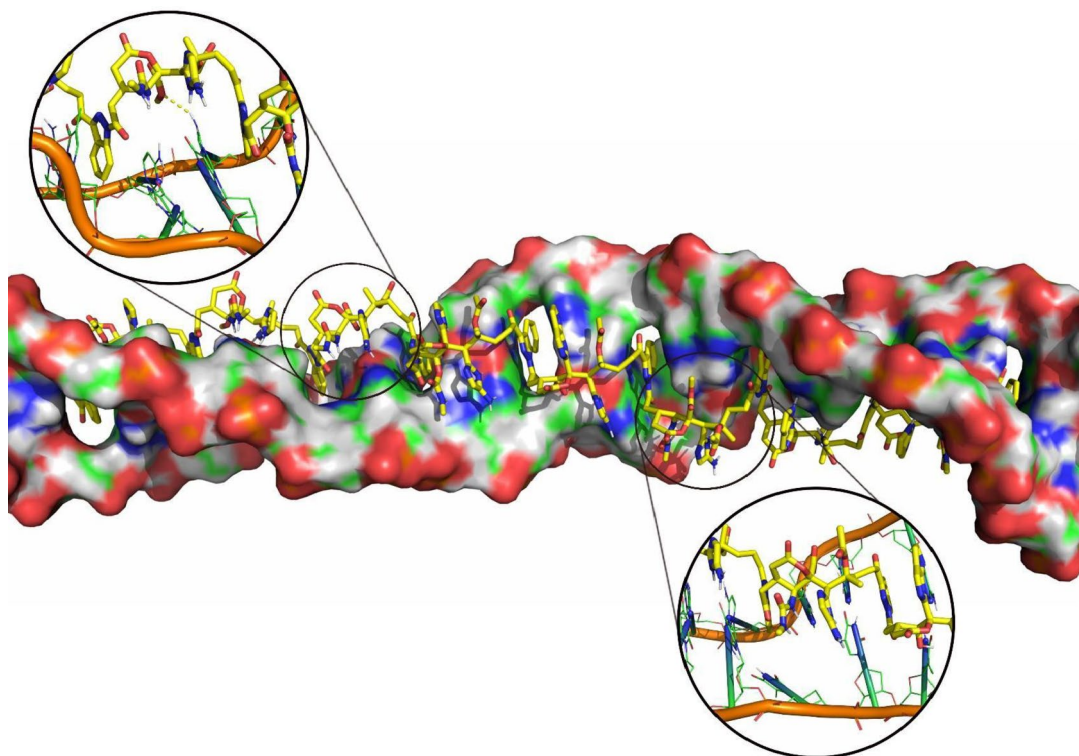


Fig. 8. General view of HASDI-G2(BCR_ABL1) complex with BCR-50nt during molecular dynamics simulation (representative trajectory frame).

duplex interaction with the selective part of the ligand with its gradual complete loss. At 146 ns of the simulation one of the bases G24 that should have been recognized by HASDI-G2(BCR_ABL1) in this region left the double helix.

The latter occurred after T30 (the first nucleobase after the junction) also left the duplex at the 119 ns of the simulation. That suggests a significant contribution of the cooperative effect to the stability of the HASDI-G2(BCR_ABL1)/target interaction. It should also be noted that the 3'-end of the sense chain and the corresponding part of the antisense chain in the duplex formed a stable slightly bent conformation relative to the imaginary straight rest of the double helix until the end of the simulation.

The number of HBs between HASDI-G2(BCR_ABL1) and BCR-50nt changed during the simulation (Fig. 9). In particular, from the 1st to the 3rd ns of the trajectory, their quantity was 29. After that, a sharp decrease to about 26 HBs was observed. At the same time, by the 33rd ns their number had returned to the initial level. However, the rest of the simulation was characterized by their gradual and relentless decrease. Thus, at the 150th ns of simulation, the number was equal to 25 HBs without any tendency to stabilization or growth. Accordingly, at the end of the simulation, one linker on average interacted with the target by forming 3.125 HBs out of 6 possible ones.

The binding free energy calculation generally confirms a significant level of destabilization of the HASDI-G2(BCR_ABL1)/DNA complex in the event of intercalation of the ligand into the incorrect site (Table 6). Even though 12 out of 16 nucleobases that can be recognized by HASDI-G2(BCR_ABL1) correspond to the optimal sequence. In general, the ΔG of the HASDI-G2(BCR_ABL1)/BCR-50nt complex was equal to -146.19 ± 7.34 kcal/mol. This is 8.76 kcal/mol lower compared to the binding free energy of HASDI-G2(BCR_ABL1) to BCR_ABL1-50nt and is a consequence of the overall slightly lower values of each of the calculated interaction enthalpy components. At the same time, the electrostatic component and the polar energy of solvation underwent the largest changes by 5% each. That is obviously related to the decrease in the number of HBs between the components of the complex and the destabilization of the DNA duplex itself with the release of some of the nucleobases outside the double helix.

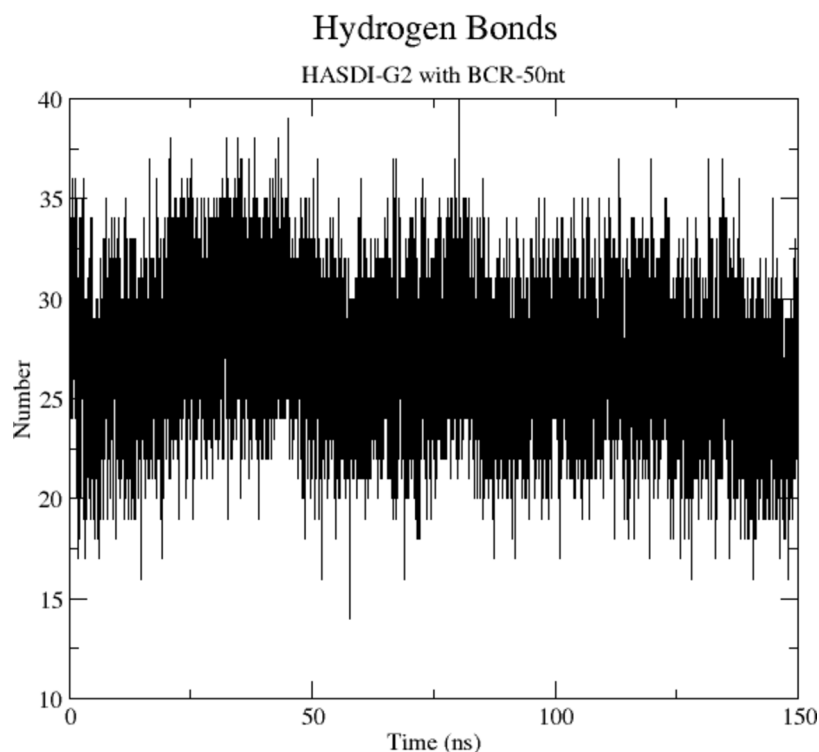


Fig. 9. Dependence of the hydrogen bonds quantity on the molecular dynamics simulation time in the HASDI-G2(BCR_ABL1)/BCR-50nt complex.

	Evdw	Eel	Epb	Enpol	Edisp	ΔH	$-T\Delta S$	ΔG
Average	-387.29	-76.19	146.54	-196.27	348.27	-164.94	18.75	-146.19
SD	8.14	3.78	4.24	3.08	4.41	7.79	0.88	7.34

Table 6. MM/PBSA calculated binding energy of HASDI-G2(BCR_ABL1) and BCR-50nt expressed in kcal/mol.

Molecular dynamics simulation of the HASDI-G2(BCR_ABL1)/ABL1-50nt complex

Similar to molecular dynamics simulations of the HASDI-G2(BCR_ABL1) complex with native BCR-50nt, intercalation to the native ABL1-50nt resulted in the absence of most of the orienting ligand/receptor interactions before the fusion site early in the simulation (Fig. 10). Only in the case of a single nucleotide overlap (T21 and T23) full classical interactions were observed.

However, with the course of the simulation, an increase in conformational changes of both the ligand and DNA in the intercalation zone was observed. That led to the rapid formation of non-canonical HBs and their subsequent rupture. Starting from the 62nd ns, the third linker completely lost all directional interactions with nucleobases in the contact area. This occurred despite G25 being the nucleobase that HASDI-G2(BCR_ABL1) is able to recognize within the “correct” complex. Subsequently, interactions of the linker and G25 were observed. However, the newly formed HBs were extremely unstable and did not relate to the cytidine of this nucleotide pair.

Starting from the 102nd ns of the simulation, a sharp loss of all orientational interactions between the sixth HASDI-G2(BCR_ABL1) linker and the corresponding part of the DNA duplex was observed. That part in turn, is located after the point of fusion of the BCR and ABL1 genes. Accordingly, the cooperative effect of the partial mismatch of the HASDI-G2 sequence is clearly manifested in this case as well. For the sake of fairness, it should be noted that by the end of the simulation, a short period of complete restoration of all interactions between the sixth linker of HASDI-G2(BCR_ABL1) and the corresponding part of DNA was also observed. However, at the end of the simulation, there was no interaction.

As can be seen from Fig. 11, the overall change in the HBs number between ABL1-50nt and HASDI-G2(BCR_ABL1) fluctuates significantly over the course of the simulation. That confirms the features of the trajectory of the ligand-receptor complex described above. From the beginning of the simulation to 112 ns, the average HBs number drops rapidly from 30 to 31 bonds to 25 per complex. After that, a short-term increase of the quantity to 28 is observed. However, starting from 143 ns of the simulation and until its end, the number of HBs rapidly

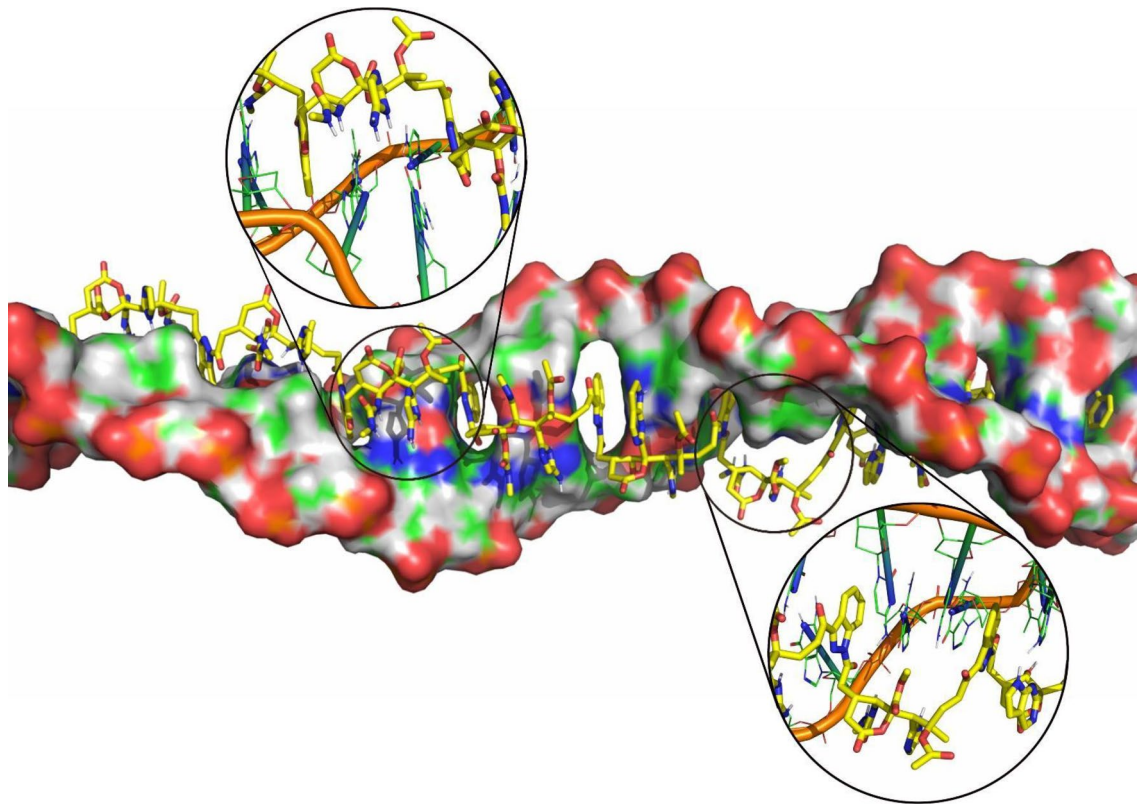


Fig. 10. General view of HASDI-G2(BCR_ABL1) complex with ABL1-50nt during molecular dynamics simulation (representative trajectory frame).

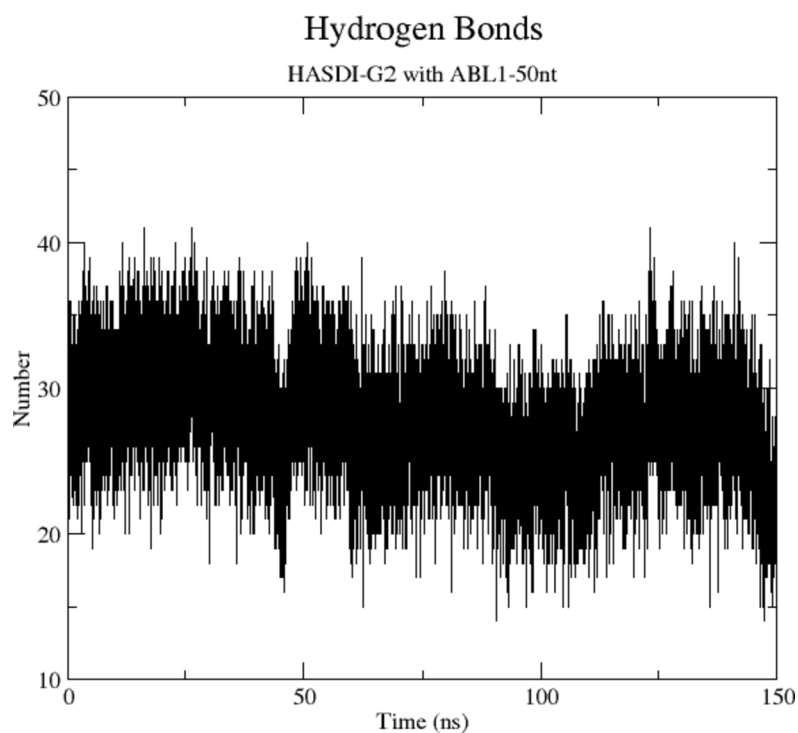


Fig. 11. Dependence of the hydrogen bonds quantity on the molecular dynamics simulation time in the HASDI-G2(BCR_ABL1)/ABL1-50nt complex.

decreases to 22 for the entire complex. That is equal to 2.75 bonds per linker out of 6 possible. And the value is close to that in the case of the almost completely mismatched HASDI-G2(EBNA1)/KCNH2-50nt complex.

Despite the correspondence of 13 out of 16 base pairs to the target sequence for HASDI-G2(BCR_ABL1), the binding energy calculated by the MM/PBSA method is lower than that for the interaction of the ligand with BCR-50nt (Table 7), where 12 out of 16 nucleobases were matched. In this case, $\Delta G = -138.99 \pm 8.36$ kcal/mol, which is 15.96 kcal/mol lower compared to the binding free energy of the reference complex or 7.2 kcal/mol lower compared to the HASDI-G2(BCR_ABL1)/BCR-50nt. The latter is comparable to the difference in free energy of HASDI-G2(BCR_ABL1) binding to BCR_ABL1-50nt and BCR-50nt.

The energy components of the interaction enthalpy were differently correlated with those for the reference complex and for the HASDI-G2(BCR_ABL1)/BCR-50nt. In particular, the energy of Van der Waals interactions had an average value compared to these two complexes as extreme points. Instead, the nonpolar solvation energy and the contribution of dispersion forces were approximated for those in the case of the reference. At the same time, the polar solvation energy was close to that of the HASDI-G2(BCR_ABL1) complex with BCR-50nt, and the electrostatic component — the main reflection of selective interactions — was lower compared to the Coulomb interactions characteristic of this complex. However, the main contribution to such a significant change in the binding free energy was obtained due to the increase of the entropy component by more than 69% compared to the reference complex.

Stability of HASDI-G2 interaction with KRAS_G12S-50nt and KRAS-50nt

In this part of the research, HASDI-G2 due to the selection of substituents in the R₁, R₂, R₃ and R₄ positions of each of the eight linkers is directed to the KRAS_G12S mutant gene, which differs from the native one by only one nucleotide substitution — G31A (numbering from the 5' end of the selected 50nt sequence) (HASDI-G2(KRAS_G12S)). In addition to simulating the complex of the developed intercalator with the target sequence and evaluating its stability, HASDI-G2(KRAS_G12S) in complex with the native KRAS gene was also investigated. This allows to differentially determine the effect of only one nucleotide change on the stability of such a complex.

Molecular dynamics simulation of the HASDI-G2(KRAS_G12S)/KRAS_G12S-50nt complex

The simulation trajectory of the HASDI-G2(KRAS_G12S)/KRAS_G12S-50nt was generally very similar to that of the previous complexes with a targeting sequence (Fig. 12).

Most of the HBs were preserved throughout the simulation. Each pair of nucleobases was recognized by the formation of at most 6 HBs. The indazole rings stably maintained their position within the limits of the initial intercalation — that is, every 2 base pairs. Accordingly, the complex HASDI-G2(KRAS_G12S)/KRAS_G12S-50nt can be considered classically very stable.

The number of HBs was calculated with the help of the GROMACS internal tools, and on average it was equal to 35–36 interactions (Fig. 13). This value was relevant for most of the simulation. Only starting from 138 ns, a slow decrease to ~34 HBs was seen. That, however, did not affect the picture we observed during the analysis of the simulation trajectory of this complex. Accordingly, one linker that recognizes 2 base pairs accounts for 4.25–4.5 HBs. This is a higher value compared to the previous version of HASDI, but in this study it is between HASDI-G2(BCR_ABL1) and HASDI-G2(EBNA1) complexes with targeting sequences.

The binding free energy calculated on the basis of the molecular dynamics simulation trajectory in general indicates in favor of even greater stability of the HASDI-G2(KRAS_G12S) complex with KRAS_G12S-50nt compared to similar complexes and is equal to -159.6 ± 6.61 kcal/mol (Table 8). That is 4.65 kcal/mol more compared to HASDI-G2(BCR_ABL1)/BCR_ABL1-50nt and 1.88 kcal/mol more than for HASDI-G2(EBNA1)/EBNA1-50nt. However, it should be noted that most of the energy components of the enthalpy were comparable to those for HASDI-G2(BCR_ABL1)/BCR_ABL1-50nt. At the same time, the entropy penalty had an intermediate value between the above-mentioned complexes. The only significant difference was expressed in the energy of electrostatic interactions, which in this case was 85.44 ± 3.56 kcal/mol. Accordingly, it is the electrostatic contribution that is key here.

Molecular dynamics simulation of the HASDI-G2(KRAS_G12S)/KRAS-50nt complex

Despite the replacement of only one pair of nucleobases, the trajectory of the HASDI-G2(KRAS_G12S) complex with KRAS-50nt was significantly different from that of the native one (Fig. 14).

From the very beginning of the simulation, in addition to the expected absence of orientational interactions in the area of the 31st bp of the DNA duplex, a rapid destabilization of the interactions of the ligand's 4th linker with its corresponding base pairs was observed. At the same time, A27 changed its own position relative to the plane of the double helix, partially leaving it. In fact, local melting was observed in this area from the very beginning of the simulation. As the time progressed, the number of HBs decreased and ranged from 0 to 2 starting at the 36th ns.

	Evdw	Eel	Epb	Enpol	Edisp	ΔH	-TΔS	ΔG
Average	-392.86	-75.12	146.72	-201.61	353.0	-169.88	30.89	-138.99
SD	9.86	4.56	5.0	3.64	5.5	8.28	1.12	8.36

Table 7. MM/PBSA calculated binding energy of HASDI-G2(BCR_ABL1) and ABL1-50nt expressed in kcal/mol.

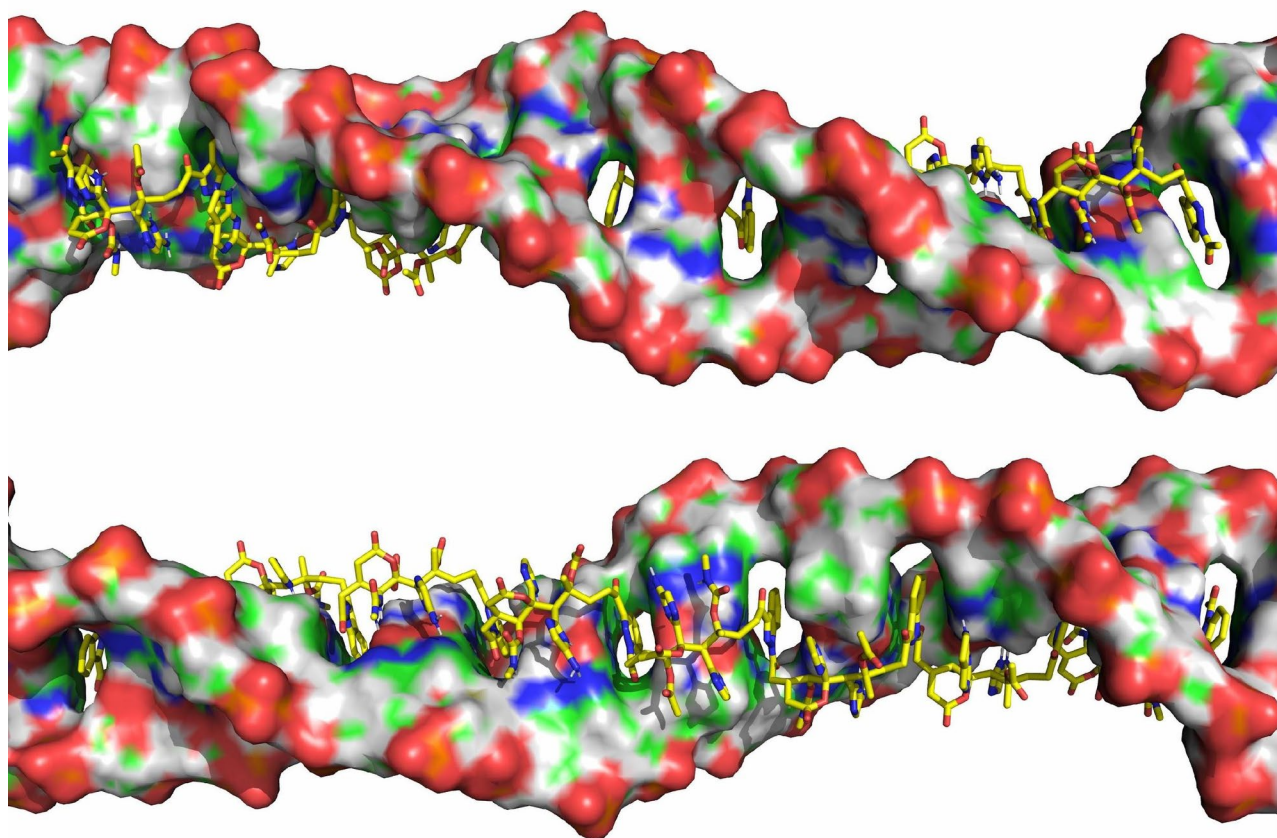


Fig. 12. General view of HASDI-G2(KRAS_G12S) complex with KRAS_G12S-50nt during molecular dynamics simulation (representative trajectory frame).

To a large extent, a similar situation was observed in the case of the seventh linker of HASDI-G2(KRAS_G12S-50nt). Starting from the 7th ns, T33 lost the classical Watson-Crick interactions with its corresponding adenine and also went beyond the duplex plane. Subsequently, short-term periods of a significant increase in the lability of the ligand were observed in this area, with the breaking of most of the orienting interactions. However, all of them ended with stabilization in a conformation where three of the four bases are recognized using the classical pattern of HBs, and T33 forms an exotic hydrogen bond between the 3NH of the heterocycle and the oxygen of the ligand's hydroxyl.

It should also be noted that the 32nd bp, which directly contacts the single nucleotide substitution site, has also lost most of the possible interactions with the corresponding linker and its substituents. Only occasionally did the sixth linker formed weak HBs between R2/R3 and G31. At the same time, the actual conformation of the linker section changed significantly.

In the case of the remaining five HASDI-G2(KRAS_G12S-50nt) linkers, all classical orientational interactions with their corresponding base pairs were observed.

The number of HBs in the HASDI-G2(KRAS_G12S-50nt)/KRAS complex in the period from the 1st to the 14th ns was 30–31 (Fig. 15). However, their rapid decrease to 26–28 was further observed, which remained relatively evenly until the end of the simulation with some tendency towards their decrease at the end. Accordingly, in this case, each individual linker formed an average of 3.375 HBs with the corresponding base pair, which is a relatively high value compared to other mismatched complexes and may indicate the presence of some level of additional stabilization of the orientational interactions of the remaining ligand's linker regions.

The binding free energy of the HASDI-G2(KRAS_G12S-50nt)/KRAS-50nt complex calculated by the MM/PBSA method was -146.13 ± 7.45 kcal/mol (Table 9). That is 13.46 kcal/mol less compared to the native complex. However, it is more compared to the HASDI-G2(BCR_ABL1)/BCR-50nt, where only 12 of 16 nucleobases were matched. All the stabilizing energy components of the enthalpy were expectedly and even classically lower compared to the native complex. At the same time, the contribution of polar solvation was characterized by a smaller energy penalty, and the contribution of dispersion forces was almost identical. At the same time, the entropy penalty was significantly reduced.

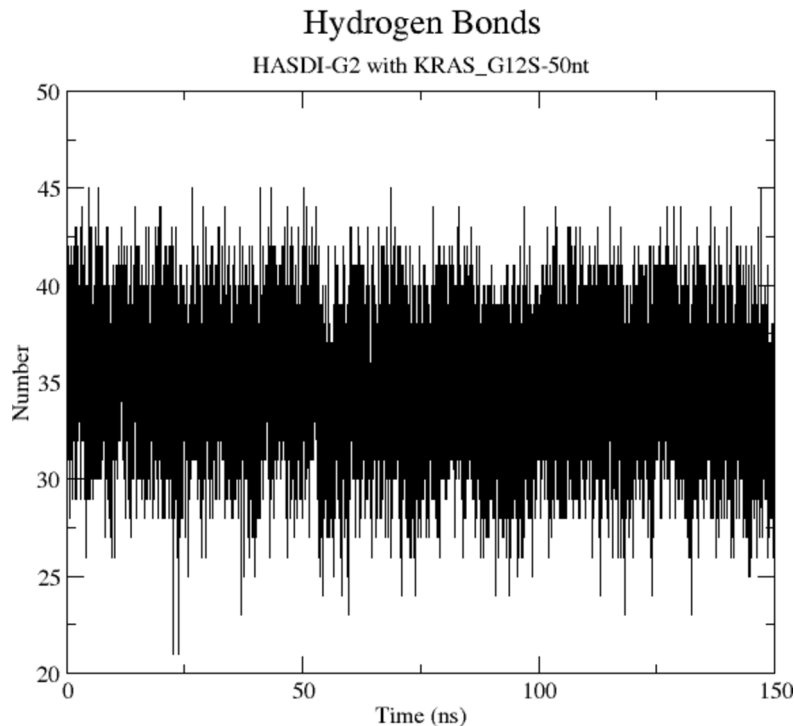


Fig. 13. Dependence of the hydrogen bonds quantity on the molecular dynamics simulation time in the HASDI-G2(KRAS_G12S)/KRAS_G12S-50nt complex.

	Evdw	Eel	Epb	Enpol	Edisp	ΔH	-TAS	ΔG
Average	-398.67	-85.44	154.23	-202.33	353.66	-178.56	18.95	-159.6
SD	7.04	3.56	3.74	2.64	3.94	6.61	0.04	6.61

Table 8. MM/PBSA calculated binding energy of HASDI-G2(KRAS_G12S) ra KRAS_G12S-50nt expressed in kcal/mol.

Discussion

Highly selective agents capable of recognizing a certain irregular DNA sequence are generally attractive for use as anticancer drugs. Currently, a large number of low-molecular compounds with different levels of selectivity for DNA duplex have already been developed^{20–22}. In our opinion, the most promising of the currently existing ones are pyrrole-imidazole polyamides (PIPs) discovered in 1996 by Dervan et al.²³. This is due to their relatively significant ability to target specific DNA sequences. PIPs actually consist of mostly two large fragments, which are built on the basis of N-methylpyrrole and N-methylimidazole connected to each other by classical amide bonds. Instead, these parts are mainly linked together through a linker of mostly γ -aminobutyric acid. The interaction of PIPs with the target DNA sequence involves their placement in the minor groove in such a way that the linker section forms a so-called γ -turn, and the pyrrole-imidazole polyamide fragments pair with each other through Van der Waals interactions into a stable structure where hydrogen bond donors and acceptors lie in one direction towards the nucleobases. Thus forming a DNA-like antiparallel structure where the non-covalent pyrrole-pyrrole pair recognizes A/T or T/A pairs and the imidazole-pyrrole recognizes G/C pairs.

Over the past few decades, a number of studies have been conducted that confirmed the significant potential of pyrrole-imidazole polyamides as agents capable of modulating the transcription of certain parts of the genome. In particular, the group led by Peter Dervan, continuing their own research, developed a polyamide directed at DNA sequences associated with the implementation of the stimulating effect of the NF- κ B factor (considered to be associated with many types of cancer and aging)²⁴. For this, the researchers synthesized the structure [ImImPyPy-(R) α -aminoy-PyPyPyPy] capable of theoretically recognizing the DNA sequence 5'-WGGWWW-3'. Addition of this polyamide to short fragments of the IL6 and IL8 genes (regulated by NF- κ B) resulted in an increase in the melting point of more than 10 °C in both cases. Further studies on A549 cells revealed a 6.5- and 4-fold decrease in IL6 and IL8 mRNA, respectively (incubation of 10 μ M of the construct with cells for 48 h followed by TNF- α stimulation for 12 h). Finally, Peter Dervan and his team tested their developed pyrrole-imidazole polyamide for its targeted activity against the growth of prostate cancer xenografts (LNCaP) in vivo in a male NSG mouse model. They found a pronounced activity of the design they developed applied at a concentration of \sim 1 mg/kg for 9 days: the size of the tumors in the experimental group was 64% smaller

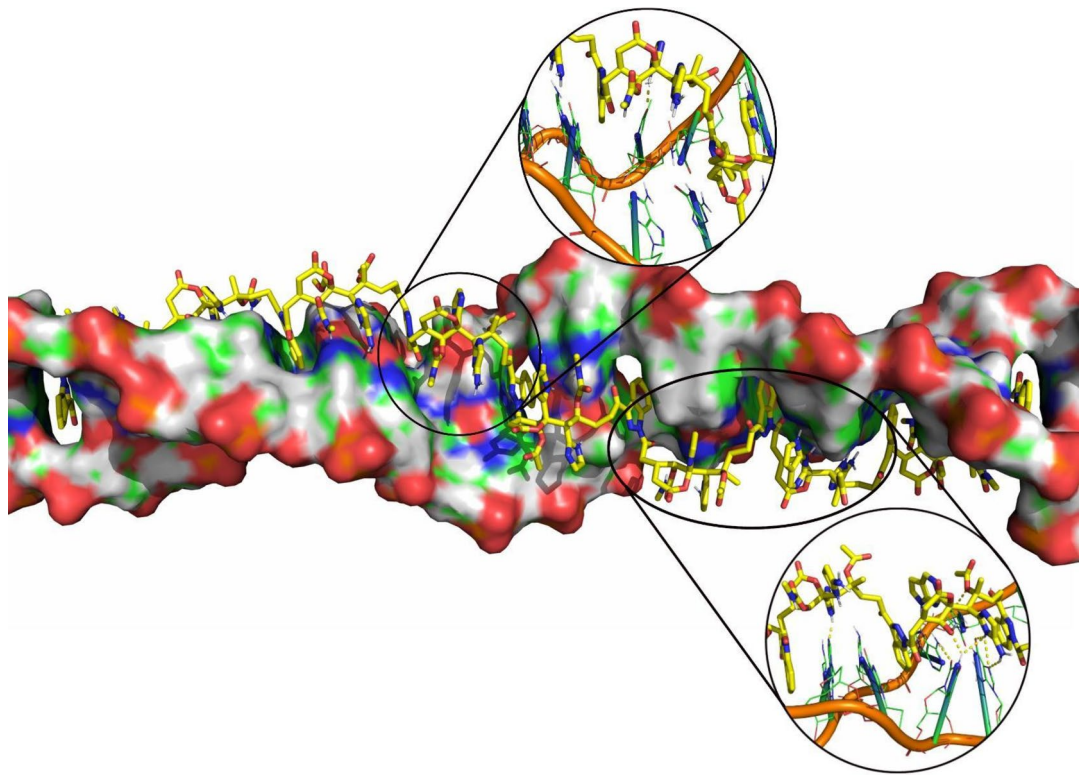


Fig. 14. General view of HASDI-G2(KRAS_G12S) complex with KRAS-50nt during molecular dynamics simulation (representative trajectory frame).

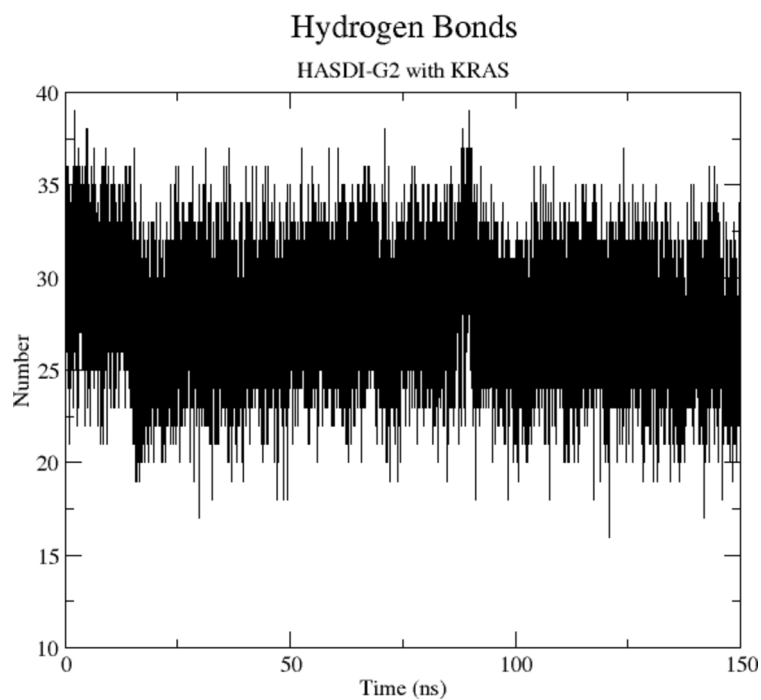


Fig. 15. Dependence of the hydrogen bonds quantity on the molecular dynamics simulation time in the HASDI-G2(KRAS_G12S)/KRAS-50nt complex.

	Evdw	Eel	Epb	Enpol	Edisp	ΔH	-TAS	ΔG
Average	-390.37	-75.82	150.44	-198.28	353.58	-160.46	14.33	-146.13
SD	8.54	5.22	6.17	3.64	5.45	7.45	0.05	7.45

Table 9. MM/PBSA calculated binding energy of HASDI-G2(KRAS_G12S) to KRAS-50nt expressed in kcal/mol.

compared to the control. In addition, the peculiarities of the histological profile and biochemical parameters testified in favor of the rapid necrosis of xenografts.

The studies described above are only a small, albeit representative part of a wide variety of efforts in the direction of the development of anticancer drugs based on pyrrole-imidazole polyamides.

However, despite this, the very approach of using such structures introduces constructive limitations regarding further progress in this direction. In particular, most of the currently available designs have the ability to recognize no more than 6 nucleotides of an irregular DNA sequence²¹. This is related to the larger step of PIP compared to the DNA duplex, which with the growth of the size of the ligand leads to a rapid increase in tension in the conjugate and, accordingly, to a decrease in its stability^{25,26}. At the same time, there were attempts to find a solution to this problem, carried out, in particular, both by Dervan and his team (by incorporating β -alanine into the PIP structure) and by other researchers^{27–29}. This, however, is also valid only up to a certain limit, as it implies a significant increase in the contribution of the entropy penalty to complexation, which is associated with the addition of the ligand by the long aliphatic linker(s)³⁰. Furthermore, the obvious problem of pyrrole-imidazole polyamides remains the degeneracy of their recognition — the non-covalent imidazole-pyrrole pair equally recognizes both G/C and C/G²¹.

In this work, we present our polyintercalating agent HASDI-G2, the linker sites of which are located in the major groove of DNA and are capable of clearly discriminating the target pair of nucleobases among others. In addition to higher selectivity for a specific nucleobase compared to polyamides, HASDI-G2 has the potential for unlimited size increase. That, in perspective, will allow to achieve almost any desired level of selectivity to an irregular DNA sequence. The only obvious limitation may be the pharmacokinetics of the obtained compound, as well as its ability to independently enter the cell nucleus through the nuclear pore complex³¹. Actually, in this and previous studies, the structure developed by us is aimed at recognizing 16 base pairs. According to our preliminary theoretical calculations, this length of the targeting region will allow HASDI-G2 to identify the target sequence among more than 68 billion base pairs⁶.

HASDI-G2 is the second generation of our developed polyintercalating molecular structures. It differs from its predecessor (HASDI) in an intercalating component and some substituents that directly recognize a specific pair of nucleobases (Table 2). These changes were aimed at improving the ratio between the selective component of the ligand/receptor interaction (hydrogen bonds between R₁, R₂, R₃ and R₄ substituents) and the non-selective part of it (hydrophobic and stacking interactions between intercalating planar aromatic structures).

First of all, in order to check the degree of influence of the introduced changes, we repeated the experiment conducted within the framework of previous studies. In particular, HASDI-G2(EBNA1) was created with substituents in the positions of R₁, R₂, R₃ and R₄ of the linker regions to target the sequence 5'-TCCCTTTACA ACCTCA-3' (a 16 bp-long fragment of the EBNA1 gene) according to the principles described in the Table 2. After that, HASDI-G2(EBNA1) was intercalated into a duplex generated on the basis of the EBNA1 gene with a length of 50 base pairs, which contained the targeting sequence, and also into a duplex of the same length generated on the basis of a random sequence of the KCNH2 gene. The obtained two HASDI-G2(EBNA1)/DNA complexes were studied by the method of molecular dynamics simulation.

The trajectories of these simulations were strikingly different from each other. In particular, the HASDI-G2(EBNA1) complex with EBNA1-50nt was characterized by high stability, an average of 36 HBs, and a free energy of complex formation of -157.72 ± 6.57 kcal/mol. Adversely, the complex of HASDI-G2(EBNA1) with KCNH2-50nt was relatively weak: significant conformational rearrangements were observed, which were accompanied by local melting of the DNA duplex in several places and even deintercalation of a few terminal indazole rings. At the same time, the average HBs number fluctuated around 20 with a tendency to decrease, and ΔG was -97.25 ± 9.18 kcal/mol. That is 60.47 kcal/mol or 38.3% less compared to the HASDI-G2(EBNA1)/EBNA1-50nt complex.

A similar difference in the nature of the dynamics of these two complexes was also observed in the case of the first generation structure, but to a much lesser extent. In particular, the difference between complexation energies differed only by 41.83 kcal/mol or by 17.8%. At the same time, such significant conformational changes did not occur, although one local melting of the DNA duplex was also observed. Taken together, the above indicates a much greater potential for HASDI-G2 to specifically sequence-selective interaction with DNA compared to HASDI, as its predecessor.

To better understand the potential of HASDI-G2 a number of additional studies were conducted. We tried to investigate the difference between the affinity of HASDI-G2 to variously altered due to mutation sequences and their native predecessors using the example of real genetic code errors.

As the most vivid example, we considered the fused BCR_ABL1 gene, which is the molecular basis of the so-called Philadelphia chromosome — the main cause of chronic myeloid leukemia³². As in the previous case, HASDI-G2(BCR_ABL1) was designed by selecting the substituents in the positions of R₁, R₂, R₃ and R₄ according to the principles described in Table 2, and was targeted at the sequence 5'-CTGTGGAGTGGGTTT T-3'. It contains the direct fusion region of BCR and ABL1 in the central part, which in turn has a short sequence

5'-GAGTG-3' (25–29) common to both BCR and ABL1. In addition, there are other overlaps between these genes. In particular, four thymines (T21, T23, T32 and T34) are present in the sequences of both genes near the fusion site, as well as in BCR_ABL1. Accordingly, the placement of HASDI-G2(BCR_ABL1) in the fusion site of both BCR and ABL1 implies a relatively small mismatch of the ligand to the DNA sequence. In the case of intercalation of HASDI-G2(BCR_ABL1) to BCR-50nt, 4 nucleotide pairs were mismatched, and in the case of ABL1-50nt — only three.

According to the idea of the previous experiment, we performed three simulations: HASDI-G2(BCR_ABL1) in complex with BCR_ABL1-50nt, with BCR-50nt, and with ABL1-50nt.

In the first case, the molecular dynamics of the complex was similar to that of the HASDI-G2(EBNA1)/EBNA-50nt. In particular, no signs of destabilization of the complex were observed during the entire simulation. All indazole rings were placed in their original positions, and the substituents of the linker sites formed stable hydrogen interactions with the corresponding base pairs. The DNA duplex was also stable. At the same time, the total number of HBs between the ligand and the receptor was equal to ~34 during almost the entire simulation period, and the energy of complex formation was -154.95 ± 6.93 kcal/mol. That also associates the formed complex with the HASDI-G2(EBNA1)/EBNA-50nt.

A much different situation was observed in the case of simulation of HASDI-G2(BCR_ABL1)/BCR-50nt and HASDI-G2(BCR_ABL1)/ABL1-50nt complexes.

In particular, in the first case, there was generally expected absence of orienting interactions between the ligand and four mismatched pairs of nucleotides. That pairs, together with T32 and T34, were located below the fusion zone and interacted with HASDI-G2(BCR_ABL1) in the region 6, 7 and 8 of the linkers. Moreover, significant destructive changes also applied to the zone above the fusion of BCR with ABL1 (the area that fully corresponded to the HASDI-G2 (BCR_ABL1) recognition sequence). In particular, the third linker aimed at recognizing G24, G25 and their corresponding nucleobases of the antisense chain lost all hydrogen interactions with the receptor almost immediately after the start of the simulation. Moreover, shortly after T30 left the DNA duplex at 119 ns of the simulation, local melting was also observed in the area of the third linker — G24 also left the double helix. The described, in our opinion, may be a consequence of the so-called cooperative effect: conformational rearrangements of one part of a macromolecule or conglomerate as a result of changes in relatively distant areas³³. In this case, these changes are manifested in the localization of tension in certain key areas of direct interaction of HASDI-G2 with the DNA duplex.

As expected, the HASDI-G2(BCR_ABL1)/BCR-50nt complex was also inferior to HASDI-G2(BCR_ABL1)/ABL1-50nt at the level of total HBs and binding free energy. Particularly, the total average HBs number was unstable and ranged from 25 to 29 with a tendency to decrease, and the complexation energy was -146.19 ± 7.34 kcal/mol. That, together with the features of the simulation trajectory, supports the apparently lower affinity of HASDI-G2(BCR_ABL1) to BCR-50nt compared to BCR_ABL1-50nt. It should be noted that the difference in affinity is not as significant as in the case of HASDI-G2(EBNA1)/KCNH2-50nt. This, in turn, is associated with a different degree of mismatch between the DNA sequences of these complexes and the targeting sequences.

A similar situation is observed regarding the HASDI-G2(BCR_ABL1)/ABL1-50nt complex. In particular, from the beginning of the simulation random unstable HBs between the duplex and the corresponding linkers of the ligand were observed above the fusion site. Or rather, in the area where three pairs of nucleotides not corresponding to HASDI-G2(BCR_ABL1) are located. Only T21 and T23 formed a full-fledged pattern of orienting interactions during this period. The increase in lability in this area led to the localization of tension within several linker regions of the ligand, which subsequently mostly lost interaction with the DNA duplex. The first of them formed in the region of the third linker at 62 ns of the simulation, and the second one - in the region of the sixth linker at 102 ns. This occurred despite the fact that G25, G30 and G31 fully correspond to HASDI-G2(BCR_ABL1). Accordingly, in this case, there is also a manifestation of the cooperative effect. At the same time, it should be noted that compared to HASDI-G2(BCR_ABL1)/BCR-50nt, the ligand-receptor interactions in these areas were still restored, albeit for a short period, and no signs of local melting of the DNA duplex were observed. According to the graph of the dependence of the quantity of hydrogen bonds on time, their average number in this complex also fluctuated characteristically and amounted to ~25 during most of the simulation with a sharp drop to 22 at the end of the experiment. The energy of complexation was -138.99 ± 8.36 kcal/mol.

As can be seen, the complexation energy and the HBs quantity clearly favor a lower affinity of HASDI-G2(BCR_ABL1) for ABL1-50nt compared to BCR-50nt. However, destructive conformational changes of the DNA duplex, such as local melting, are characteristic of the HASDI-G2(BCR_ABL1)/BCR-50nt complex. This in summary testifies in favor of the process of destabilization of both obtained complexes. However, in the first case, in our opinion, we observe one of the local energy minima on the way to the complete dissociation of the complex. Whereas in the second one, due to a smaller discrepancy in the sequence, this process is more extended in time.

To understand the ability of HASDI-G2 to discriminate a target among truly related sequences, we conducted another simulation experiment. We targeted the developed structure at a mutant variant of the KRAS gene — KRAS_G12S. The sequences of these genetic variants differ by a single nucleotide pair (G→A), which is the cause of the amino acid substitution and, as a result, significant disturbances in the functioning of this factor⁷. Discrimination of one of two sequences that differ by only one nucleotide pair out of sixteen can be considered as the most difficult task that can be put before HASDI-G2 in future.

Molecular dynamics simulation of the complex of HASDI-G2(KRAS_G12S) with KRAS_G12S-50nt confirmed the statement formed on the basis of two previous experiments regarding the absence of a significant difference in the stability of the HASDI-G2/target complex under the condition of aiming at specific sequence. In particular, the trajectory of the obtained complex was characterized by high uniformity and the absence of

any destructive changes or signs of destabilization during the entire simulation period. The number of HBs was ~ 34 , and the binding free energy was -159.6 ± 6.61 kcal/mol.

However, the most interesting thing was about the HASDI-G2(KRAS_G12S) complex with KRAS-50nt. Although HASDI-G2(KRAS_G12S) was able to recognize 15 out of 16 nucleobases, the obtained complex was characterized by a sufficiently low level of stability. In particular, from the very beginning of the simulation, two local single-nucleotide meltings of the DNA duplex were observed: A27 and T33 left the plane of the double helix. Moreover, from this period no more regular orienting connections with the rest of the nucleic bases were observed in the area of A27 and, accordingly, the fourth linker. Instead, in the T33 region, after the change in its position, the interaction of the ligand with the rest of the nucleic bases was stabilized. In which three of the four bases formed classical orientational interactions with the ligand, and T33 formed an exotic HB between the 3NH of the heterocycle and the oxygen of the ligand's hydroxyl. The site of interaction of HASDI-G2(KRAS_G12S) with a mismatched base pair and G32 was characterized by a significant level of lability and the absence of stable HBs, as expected. At the same time, the total average number of HBs between the ligand and the receptor mostly fluctuated within 27 with a tendency to decrease, and the energy of complexation was -146.13 ± 7.45 kcal/mol.

As can be noted, the HASDI-G2(KRAS_G12S)/KRAS-50nt during the molecular dynamics simulation revealed a level of stability comparable to other complexes of a similar type (HASDI-G2(BCR_ABL1)/BCR-50nt and HASDI-G2(BCR_ABL1)/ABL1-50nt). And it was also characterized by similar destructive conformational rearrangements that increased over time. This is an interesting effect because unlike them, KRAS-50nt, as we mentioned earlier, differs from the targeting sequence by only one nucleotide pair.

In summary, HASDI-G2 in three virtual experiments, which involved a total of 7 simulations (each 150 ns in length), demonstrated a significant level of affinity to exactly the sequences it was targeted at. This was manifested in the high stability of HASDI-G2(EBNA1)/EBNA1-50nt, HASDI-G2(BCR_ABL1)/BCR_ABL1-50nt and HASDI-G2(KRAS_G12S)/KRAS_G12S-50nt complexes. In each of them, HASDI-G2 was modified to target EBNA1-50nt, BCR_ABL1-50nt, and KRAS_G12S-50nt, respectively, according to the principles we developed (Table 2). These complexes were characterized by a surprisingly high level of homogeneity of the simulation trajectory, the absence of destructive changes, a high level of orientational interactions, as well as a similar and high calculated complexation energy: -157.72 ± 6.57 kcal/mol, -154.95 ± 6.93 kcal/mol, and -159.6 ± 6.61 kcal/mol, respectively. Instead, HASDI-G2(EBNA1)/KCNH2-50nt, HASDI-G2(BCR_ABL1)/BCR-50nt, HASDI-G2(BCR_ABL1)/ABL1-50nt and HASDI-G2(KRAS_G12S)/KRAS-50nt complexes were characterized by a relatively low level of stability of the ligand-receptor interaction, a significantly smaller number of hydrogen bonds, frequent conformational rearrangements and a relatively low energy of complex formation: -97.25 ± 9.18 kcal/mol, -146.19 ± 7.34 kcal/mol, -138.99 ± 8.36 kcal/mol and -146.13 ± 7.45 kcal/mol, respectively. Each of these associates had a DNA receptor sequence that differed from the HASDI-G2 targeting sequence, however, to varying degrees: KCNH2-50nt overlapped by 2 bp, BCR-50nt overlapped by 12 bp, ABL1-50nt overlapped by 13 bp, and KRAS-50nt overlapped by 15 bp. And although the stability of HASDI-G2(EBNA1)/KCNH2-50nt was obviously the lowest, the rest of the complexes, despite a significant difference in the degree of similarity to the targeting sequence, were generally characterized by similar conformational rearrangements and other parameters. We managed to trace similar patterns in the molecular dynamics of these complexes. In all cases, in our opinion, they are a consequence of the manifestation of a cooperative effect, which begins to show already under the condition of a one-nucleotide difference between the targeting sequence and a specific site of intercalation, and is realized in the form of destructive changes outside the immediate mismatch zone. In more detail, these changes are manifested in the loss of all orientational interactions between a certain HASDI-G2 linker and nucleobases (even if they fully match the targeting sequence), local melting of the DNA duplex, and even deintercalation of some indazole rings. The latter was not observed in the case of HASDI even with the maximum level of targeting sequence mismatch to the polyintercalator interaction site. Accordingly, HASDI-G2 can be characterized as a more selective structure compared to its predecessor.

The high values of the binding free energy obtained in this work should be considered separately. Although we were primarily interested in relative ΔG values, such association energies are sufficiently high for standard non covalent ligand-receptor interactions³⁴, even if the ligand is a high molecular structure. The solution to this issue lies in the peculiarities of the used methods and their combination. In particular, the calculation of the free energy of complex formation is carried out here on the basis of an already ready complex. The formation of the complex involves the intercalation of indazole aromatic rings between pairs of nucleobases, which in this case was carried out manually. However, according to modern data, during the *ad initio* formation of the DNA/intercalator complex, the first stage involves the creation of a cavity between base pairs^{35,36}. This is accompanied by significant energy costs, which were obviously not taken into account in this study. In our next research, we plan to cover this issue in more detail.

It should also be emphasized that during molecular dynamics simulation, the complexes HASDI-G2(EBNA1)/KCNH2-50nt, HASDI-G2(BCR_ABL1)/BCR-50nt, HASDI-G2(BCR_ABL1)/ABL1-50nt and HASDI-G2(KRAS_G12S)/KRAS-50nt demonstrated progressive destabilization of the ligand-receptor interaction. Accordingly, their calculated ΔG can only be compared conditionally with ΔG for the equilibrium complexes HASDI-G2(EBNA1)/EBNA1-50nt, HASDI-G2(BCR_ABL1)/BCR_ABL1-50nt and HASDI-G2(KRAS_G12S)/KRAS_G12S-50nt. In this work, this comparison was used to more clearly demonstrate the rate of progression of HASDI-G2 dissociation from complexes that mimic different levels of nonselectivity of interactions.

Overall, we believe that HASDI-G2 is a decent candidate for further research. We currently plan to investigate this construct both in silico (similar to this work) and in vitro. The latter is particularly important because the research carried out using biological model systems will allow us to clearly assess the degree of success of the molecular structure developed by us.

Conclusion

Our designed molecular polyintercalating construct, named HASDI-G2, was investigated in three virtual experiments. In each of them, HASDI-G2 was targeted to a specific DNA sequence. However, the intercalation was carried out in DNA duplexes generated on the basis of both targeting and foreign sequences. In the case of HASDI-G2 complexes with DNA duplexes generated on the basis of targeting sequences, a consistently high level of stability of the obtained complexes was observed, which was manifested in a high level of homogeneity of the simulation trajectory, the absence of destructive changes, close to the theoretically calculated number of hydrogen bonds, as well as similar and high calculated complexation energy. On the other hand, in the case of intercalation of HASDI-G2 into double helices with sequences of varying degrees of mismatch with the targeting sequences, conformational rearrangements were observed, which often led to local melting of the DNA duplex and, even, deintercalation of part of the indazole rings. These destructive changes were accompanied by a much smaller compared to the theoretically calculated number of orientational interactions, as well as a relatively low binding free energy.

The obtained data clearly indicate the significant potential of HASDI-G2 for highly specific targeting of defined sequences in the DNA genome. This, as well as the absence of obvious structural obstacles in the variation of the targeting sequence, allows us to consider the compound we created as promising in the context of agents capable of transcriptional regulation. And that is particularly important for the development of new anticancer drugs based on HASDI-G2.

Data availability

The datasets generated within the scope of this study are freely available from the corresponding author upon reasoned request. Links to other data used are available in this published article.

Received: 19 August 2024; Accepted: 10 February 2025

Published online: 12 March 2025

References

- Jackson, M., Marks, L., May, G. H. W. & Wilson, J. B. The genetic basis of disease. *Essays Biochem.* **62**, 643–723. <https://doi.org/10.1042/EBC20170053> (2018).
- Faguet, G. B. A brief history of cancer: age-old milestones underlying our current knowledge database. *Int. J. Cancer.* **136**, 2022–2036. <https://doi.org/10.1002/ijc.29134> (2015).
- Koo, M. M., Hamilton, W., Walter, F. M., Rubin, G. P. & Lyrtzapoulos, G. Symptom signatures and diagnostic timeliness in cancer patients: a review of current evidence. *Neoplasia* **20**, 165–174. <https://doi.org/10.1016/j.neo.2017.11.005> (2017).
- Schirmacher, V. From chemotherapy to biological therapy: a review of novel concepts to reduce the side effects of systemic cancer treatment (review). *Int. J. Oncol.* **54**, 407–419. <https://doi.org/10.3892/ijo.2018.4661> (2018).
- Huang, L., Guo, Z., Wang, F. & Fu, L. KRAS mutation: from undruggable to druggable in cancer. *Signal. Transduct. Target. Ther.* **6**, 386. <https://doi.org/10.1038/s41392-021-00780-4> (2021).
- Zaremba, A. A., Zaremba, P. Y. & Zahorodnia, S. D. In silico study of HASDI (high-affinity selective DNA intercalator) as a new agent capable of highly selective recognition of the DNA sequence. *Sci. Rep.* **13**, 5395. <https://doi.org/10.1038/s41598-023-32595-4> (2023).
- Gao, Q. et al. Selective targeting of the oncogenic KRAS G12S mutant allele by CRISPR/Cas9 induces efficient tumor regression. *Theranostics* **10**, 5137–5153. <https://doi.org/10.7150/thno.42325> (2020).
- Bansal, A., Kaushik, S. & Kukreti, S. Non-canonical DNA structures: diversity and disease association. *Front. Genet.* **13**. <https://doi.org/10.3389/fgene.2022.959258> (2022).
- Hanwell, M. D. et al. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 17. <https://doi.org/10.1186/1758-2946-4-17> (2012).
- Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001> (2015).
- Tian, C. et al. ff19SB: amino-acid-specific protein backbone parameters trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **16**, 528–552. <https://doi.org/10.1021/acs.jctc.9b00591> (2020).
- Mark, P. & Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **105**, 9954–9960. <https://doi.org/10.1021/jp003020w> (2001).
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174. <https://doi.org/10.1002/jcc.20035> (2004).
- Ke, Q., Gong, X., Liao, S., Duan, C. & Li, L. Effects of thermostats/barostats on physical properties of liquids by molecular dynamics simulations. *J. Mol. Liq.* **365**, 120116. <https://doi.org/10.1016/j.molliq.2022.120116> (2022).
- Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(1997\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(1997)18:12<1463::AID-JCC4>3.0.CO;2-H) (1997).
- Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A. & Moreno, E. gmx_MMPBSA: a New Tool to perform end-state Free Energy calculations with GROMACS. *J. Chem. Theory Comput.* **17**, 6281–6291. <https://doi.org/10.1021/acs.jctc.1c00645> (2021).
- Tan, C., Tan, Y. H. & Luo, R. Implicit Nonpolar Solvent models. *J. Phys. Chem. B* **111**, 12263–12274. <https://doi.org/10.1021/jp073399n> (2007).
- Turner, P. J. *XMGRACE, Version 5.1.19*. Beaverton (Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, 2005).
- Mukherjee, A. & Sasikala, W. D. Chapter One - Drug-DNA Intercalation: From Discovery to the Molecular Mechanism, in: Karabencheva-Christova, T. (Ed.), *Advances in Protein Chemistry and Structural Biology, Dynamics of Proteins and Nucleic Acids*. Academic Press, 1–62. <https://doi.org/10.1016/B978-0-12-411636-8.00001-8> (2013).
- Wang, M., Yu, Y., Liang, C., Lu, A. & Zhang, G. Recent advances in developing small molecules targeting nucleic acid. *Int. J. Mol. Sci.* **17**, 779. <https://doi.org/10.3390/ijms17060779> (2016).
- Bhaduri, S., Ranjan, N. & Arya, D. P. An overview of recent advances in duplex DNA recognition by small molecules. *Beilstein J. Org. Chem.* **14**, 1051–1086. <https://doi.org/10.3762/bjoc.14.93> (2018).
- Scott, F. J. et al. An evaluation of minor groove binders as anti-lung cancer therapeutics. *Bioorg. Med. Chem. Lett.* **26**, 3478–3486. <https://doi.org/10.1016/j.bmcl.2016.06.040> (2016).
- Trauger, J. W., Baird, E. E. & Dervan, P. B. Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **382**, 559–561. <https://doi.org/10.1038/382559a0> (1996).

24. Raskatov, J. A. et al. Modulation of NF- κ B-dependent gene transcription using programmable DNA minor groove binders. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1023–1028. <https://doi.org/10.1073/pnas.1118506109> (2012).
25. Blackledge, M. S. & Melander, C. Programmable DNA-binding small molecules. *Bioorganic & Medicinal Chemistry, oligonucleotides as targets and Cellular Probes* **21**, 6101–6114. <https://doi.org/10.1016/j.bmc.2013.04.023> (2013).
26. Lin, J. & Nagase, H. The Road not taken with pyrrole-imidazole polyamides: off-target effects and genomic binding. *Biomolecules* **10**, 544. <https://doi.org/10.3390/biom10040544> (2020).
27. Wang, C. C. C., Ellervik, U. & Dervan, P. B. Expanding the recognition of the minor groove of DNA by incorporation of β -alanine in hairpin polyamides. *Bioorg. Med. Chem.* **9**, 653–657. [https://doi.org/10.1016/S0968-0896\(00\)00282-0](https://doi.org/10.1016/S0968-0896(00)00282-0) (2001).
28. Kawamoto, Y. et al. Tandem trimer pyrrole-imidazole polyamide probes targeting 18 base pairs in human telomere sequences. *Chem. Sci.* **6**, 2307–2312. <https://doi.org/10.1039/C4SC03755C> (2015).
29. Kawamoto, Y. et al. Targeting 24 bp within Telomere repeat sequences with Tandem Tetramer pyrrole-imidazole Polyamide Probes. *J. Am. Chem. Soc.* **138**, 14100–14107. <https://doi.org/10.1021/jacs.6b09023> (2016).
30. Siebert, X. & Amzel, L. M. Loss of translational entropy in molecular associations. *Proteins Struct. Funct. Bioinform.* **54**, 104–115. <https://doi.org/10.1002/prot.10472> (2004).
31. Keminer, O. & Peters, R. Permeability of single nuclear pores. *Biophys. J.* **77**, 217–228. [https://doi.org/10.1016/S0006-3495\(99\)76883-9](https://doi.org/10.1016/S0006-3495(99)76883-9) (1999).
32. Osman, A. E. G. & Deininger, M. W. Chronic myeloid leukemia: modern therapies, current challenges and future directions. *Blood Rev.* **49**, 100825. <https://doi.org/10.1016/j.blre.2021.100825> (2021).
33. Hilser, V. J., Dowdy, D., Oas, T. G. & Freire, E. The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. *Proc. Natl. Acad. Sci.* **95**, 9903–9908. <https://doi.org/10.1073/pnas.95.17.9903> (1998).
34. Smith, A. J. T., Zhang, X., Leach, A. G. & Houk, K. N. Beyond Picomolar affinities: quantitative aspects of noncovalent and covalent binding of drugs to proteins. *J. Med. Chem.* **52**, 225–233. <https://doi.org/10.1021/jm800498e> (2009).
35. Mukherjee, A. & Sasikala, W. D. Drug-DNA intercalation: from discovery to the molecular mechanism. *Adv. Protein Chem. Struct. Biol.* **92**, 1–62. <https://doi.org/10.1016/B978-0-12-411636-8.00001-8> (2013).
36. Maganti, L. & Bhattacharyya, D. Sequence specificity in DNA-drug intercalation: MD simulation and density functional theory approaches. *J. Comput. Aided Mol. Des.* **34**, 83–95. <https://doi.org/10.1007/s10822-019-00268-y> (2020).

Author contributions

Andrii Zaremba: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - Original Draft. Polina Zaremba: Writing - Review & Editing. Svitlana Zahorodnia: Supervision.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025