scientific reports

OPEN

High throughput screening identifies potential inhibitors targeting trimethoprim resistant DfrA1 protein in *Klebsiella pneumoniae* and *Escherichia coli*

Soharth Hasnat[1,5], Soaibur Rahman[1,5], Meherun Binta Alam[1], Farha Mohi Suin[1], Farzana Yeasmin[3], Tanjila Suha[1], Nahuna Tanjin Supty[1], Sal Sabila[1], Animesh Chowdhury[1], A. D. A. Shahinuzzaman[4], M. Murshida Mahbub[1✉], Tofazzal Islam[3✉] & M. Nazmul Hoque[2✉]

The DfrA1 protein provides trimethoprim resistance in bacteria, especially *Klebsiella pneumoniae* and *Escherichia coli*, by modifying dihydrofolate reductase, which reduces the binding efficacy of the antibiotic. This study identified inhibitors of the trimethoprim-resistant DfrA1 protein through high-throughput computational screening and optimization of 3,601 newly synthesized chemical compounds from the ChemDiv database, aiming to discover potential drug candidates targeting DfrA1 in *K. pneumoniae* and *E. coli*. Through this approach, we identified six promising DCs, labeled DC1 to DC6, as potential inhibitors of DfrA1. Each DC showed a strong ability to bind effectively to the DfrA1 protein and formed favorable chemical interactions at the binding sites. These interactions were comparable to those of Iclaprim, a well-known antibiotic effective against DfrA1. To confirm our findings, we explored how the promising DCs work at the molecular level, focusing on their thermodynamic properties. Additionally, molecular dynamics simulations confirmed the ability of these six DCs to effectively inhibit the DfrA1 protein. Our results showed that DC4 (an organofluorinated compound) and DC6 (a benzimidazole compound) exhibited potential efficacy against the DfrA1 protein than the control drug, particularly regarding stability, solvent-accessible surface area, solvent exposure, polarity, and binding site interactions, which influence their residence time and efficacy. Overall, findings of this study suggest that DC4 and DC6 have the potential to act as inhibitors against the DfrA1, offering promising prospects for the treatment and management of infections caused by trimethoprim-resistant *K. pneumoniae* and *E. coli* in both humans and animals. However, further in vitro validations are necessary.

Keywords Trimethoprim, Antimicrobial resistance, DfrA1, Inhibitors, High-throughput screening, Computational drug discovery

Inventions of new chemical compounds against antibiotic-resistant pathogens are the critical demand for world health security. Computational chemistry plays a central role in many drug discovery applications as computational methods help understand the in-sight of physicochemical properties of small-molecule drugs[1]. Optimizing a vast array of organic compounds, particularly hit-to-lead candidates, has significantly enhanced drug-target binding affinity. Large-scale screening, powered by supercomputing, offers new hope in addressing the growing challenge of antibiotic resistance[2]. The creation of a computational library containing diverse organic compounds, enriched with elements like oxygen, nitrogen, fluorine, sulfur, and chlorine, makes drug discovery

[1]Department of Genetic Engineering and Biotechnology, East West University, Dhaka 1212, Bangladesh. [2]Molecular Biology and Bioinformatics Laboratory, Department of Gynecology, Obstetrics and Reproductive Health, Gazipur Agricultural University (GAU), Gazipur 1706, Bangladesh. [3]Institute of Biotechnology and Genetic Engineering, Gazipur Agricultural University (GAU), Gazipur 1706, Bangladesh. [4]Pharmaceutical Sciences Research Division, Bangladesh Council of Scientific and Industrial Research (BCSIR), Dhaka 1205, Bangladesh. [5]Soharth Hasnat and Soaibur Rahman contributed equally to this work. ✉email: murshida@ewubd.edu; tofazzalislam@bsmrau.edu.bd; nazmul90@bsmrau.edu.bd

more comprehensive and sophisticated[3,4]. The presence of charged electrostatic regions with electronegative atoms facilitates nucleophile attraction, a critical factor in drug-protein interactions[5]. In-silico techniques, such as drug-receptor interaction analysis and chemical attraction identification, now enable precise predictions before in-vivo experiments. Antimicrobial resistance (AMR) has emerged as a global public health crisis, threatening the efficacy of antibiotics against bacterial infections[6]. The antibiotic trimethoprim (TMP) is commonly used to treat various bacterial infections; however, its effectiveness is often compromised by the rapid emergence of TMP-resistant bacterial strains[7,8]. Recent studies have identified a concerning rise in TMP resistance among infections caused by *E. coli* and *Klebsiella* species. This widespread resistance presents substantial challenges in achieving effective treatment outcomes, highlighting the need for alternative therapeutic approaches to address these increasingly difficult-to-treat infections[7,9,10]. TMP mediates its antibacterial effect by targeting bacterial dihydrofolate reductase (DfrA1), a ubiquitous enzyme found in a large group of bacteria including the members of Gram-negative *Enterobacteriaceae*[7,11]. Pathogenic bacteria such as *E. coli*, *K. pneumoniae*, *Pseudomonas aeruginosa* etc. are rapidly evolving resistance mechanisms against TMP, rendering standard treatments increasingly ineffective[7,9,12]. The surge in TMP resistance leads to prolonged illnesses, higher healthcare costs, increased mortality, and a greater risk of spreading infections, underscoring the urgent need for innovative solutions[13]. According to the Centers for Disease Control and Prevention, TMP-resistant bacteria account for more than 9,000 healthcare-associated infections each year in the United States[14]. A significant contributor to this resistance is the widespread occurrence of the *dfrA1* gene, which can confer a 190 to 1,000-fold increase in TMP resistance[15]. In light of the challenges posed by TMP-resistant bacteria, particularly *E. coli* and *K. pneumoniae*, our research aims to identify promising drug candidates (DCs) by targeting the highly virulent DfrA1 protein, a major factor driving TMP resistance in these pathogens.

Computational tools like Gaussian offer a cost-effective and efficient approach to drug design and development, providing significant insights within a limited timeframe[16,17]. These methods enable the prediction of drug geometries, structures, and physicochemical and biological properties. Density functional theory (DFT) technique, particularly when combined with cluster-based supercomputing, allows for the extraction of highly accurate information on biological systems in a short period[18]. The effectiveness of drug-target interactions can also be anticipated by evaluating binding sites. Furthermore, DFT studies on the Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) and electronic spectra provide valuable insights into molecular stability and reactivity[19]. The use of computational platforms like Maestro (Schrödinger LLC) further enhances the drug discovery process, making it more cost-effective by revealing the stability of potential drug candidates in their interactions with receptor proteins[20]. However, despite its value, this approach can sometimes be limited in accurately capturing the inherent reactivity of a molecule within a protein complex, such as the drug-binding pocket of a protein[21]. Chemical compounds interact with the amino acids in these binding pockets, primarily through electrophilic or nucleophilic attraction to the side chain atoms of the amino acids[20]. The computer-based software used in these analyses is often developed from lab-based experiments and incorporates various algorithmic calculations to minimize errors[22]. Despite these advancements, additional methods are needed in drug discovery and development. For instance, Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) calculations are frequently used to estimate the free energy of binding between a ligand and a receptor or protein during molecular dynamics simulations[23]. This approach provides a more accurate computational prediction of receptor-ligand interactions. Numerous computational studies have been employed to identify effective compounds aimed at combating antibiotic resistance[24–27]. However, given the critical importance of this issue, a comprehensive computational study involving a large library of chemical agents is essential for targeting the DfrA1 protein. In this study, we screened 3,601 newly synthesized chemical compounds containing various electronegative chemical species against the DfrA1 protein. While TMP demonstrates complete ineffectiveness against DfrA1, Iclaprim has been recognized as the most potent drug targeting this virulent protein. However, due to concerns regarding its potential risk of liver toxicity, the FDA rejected Iclaprim in 2019[28]. Given the emerging issue of AMR against DfrA1, particularly in the absence of a fully effective drug, there is an urgent need to identify new therapeutic agents to combat resistant pathogens. Thus, the identification of two drug candidates, DC4 and DC6, through the large-scale screening of a library of 3,601 chemical compounds, presents a promising opportunity for developing therapeutics targeting trimethoprim-resistant DfrA1 protein in *K. pneumoniae* and *E. coli*. This discovery opens new avenues for further wet lab experiments for validation of our computational findings.

Results

Protein-DCs binding simulations identify effective inhibitors of DfrA1

A comprehensive screening of 3,601 optimized compounds (File S1) against TMP-resistant DfrA1 protein identified six compounds (DC1 – DC6) as effective binders. These six DCs were highlighted as the promising inhibitors of DfrA1 due to their strong interaction profiles, which included H-bonds and other non-covalent interactions, as well as their high affinity for the drug-binding pockets. Remarkably, the binding free energies of these inhibitors were all lower than −9 kcal/mol (Table 1), indicating a strong binding potential. The analysis of the chemical structures of the DCs revealed the presence of charged atoms, such as oxygen and nitrogen, in the DC inhibitors. The chemical structure, IUPAC name, and SMILES notation of the most promising compounds, as well as the control, are presented in Table 1. The identification of the drug-binding pocket regions in DfrA1 protein showed that the binding pocket is centrally located within the protein, as indicated by the red coloration (Fig. S1). Further analysis of the screened compounds demonstrated that all six potential inhibitors (DCs), along with the control drug, bind within the identified pocket region (Fig. 1). During the initial screening, DC1 formed three H-bonds with the amino acids PRO18, ILE20, and SER49 within the binding pocket, with bond lengths ranging from 2.4 to 3.41 Å. Additionally, three non-covalent interactions were detected, with distances ranging from 2.27 to 5.15 Å (Fig. 1A). The electronegative atoms of DC2, primarily oxygen and nitrogen, formed
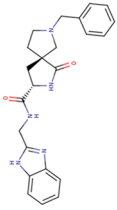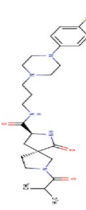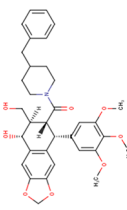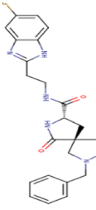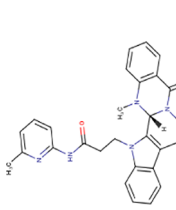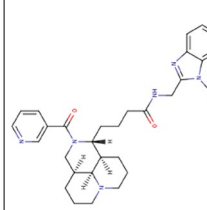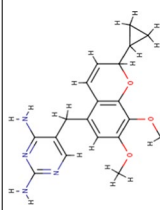
Ligand (DC)	Chemical structure	PubChem ID	IUPAC name	SMILES notation	Binding score (kcal/mol)
DC1		Compound CID: 137,908,608	(3S,5S)-N-(1H-benzimidazol-2-ylmethyl)-7-benzyl-1-oxo-2,7-diazaspiro[4.4]nonane-3-carboxamide	O=C(NCC1=NC2=CC=CC=C2N1)[C@@H]1C[C@]2(CCN(CC3=CC=CC=C3)C2)C(=O)N1	-10.0
DC2		Compound CID: 137,908,787	(3S,5R)-N-[3-[4-(4-fluorophenyl)piperazin-1-yl]propyl]-7-(2-methylpropanoyl)-1-oxo-2,7-diazaspiro[4.4]nonane-3-carboxamide	CC(C)C(=O)N1CC[C@@]2(C[C@H](NC2=O)C(=O)NCCCN2CCN(CC2)C2=CC=C(F)C=C2)C1	-9.4
DC3		(4-benzylpiperidin-1-yl)-[(5R,6R,7R,8R)-8-hydroxy-7-(hydroxymethyl)-5-(3,4,5-trimethoxyphenyl)-5,6,7,8-tetrahydrobenzo[f][1,3]benzodioxol-6-yl]methanone	Compound CID: 110,195,985	[H][C@@]1(CO)[C@@H](O)C2=C(C=C3OCOC3=C2)[C@@H](C2=CC(OC)=C(OC)C(OC)=C2)[C@@]1([H])C(=O)N1CCC(CC2=CC=CC=C2)CC1	-9.3
DC4		Compound CID: 137,908,602	(3S,5S)-7-benzyl-N-[2-(6-fluoro-1H-benzimidazol-2-yl)ethyl]-1-oxo-2,7-diazaspiro[4.4]nonane-3-carboxamide	FC1=CC=C2NC(CCNC(=O)[C@@H]3C[C@]4(CCN(CC5=CC=CC=C5)C4)C(=O)N3)=NC2=C1	-10.4
DC5		Compound CID: 137,915,644	3-[(1R)-21-methyl-14-oxo-3,13,21-triazapentacyclo[11.8.0.0²,¹⁰.0⁴,⁹.0¹⁵,²⁰]henicosa-2(10),4,6,8,15,17,19-heptaen-3-yl]-N-(6-methylpyridin-2-yl)propenamide	[H][C@@]12N(C)C3=C(C=CC=C3)C(=O)N1CCC1=C2N(CCC(=O)NC2=NC(C)=CC=C2)C2=C1C=CC=C2	-9.9
DC6		Compound CID: 110,195,961	N-[(1-methylbenzimidazol-2-yl)methyl]-4-[(5R,6R,9S,13S)-7-(pyridine-3-carbonyl)-1,7-diazatricyclo[7.3.1.0⁵,¹³]tridecan-6-yl]butanamide	[H][C@]12CCN3CCC[C@]([H])([C@@]([H])(CCCC(=O)NCC4=NC5=CC=CC=C5N4C)N(C1)C(=O)C1=CN=CC=C1)[C@]23[H]	-9.9
Control		Compound CID: 213,043	Iclaprim	COC1=C(C2=C(C=CC(O2)C3CC3)C(=C1)CC4=CN=C(N=C4N)N)OC	-8.6

Table 1. Chemical structures, annotation and Docking scores of the best interacting drug candidates (DCs) and the control drug.
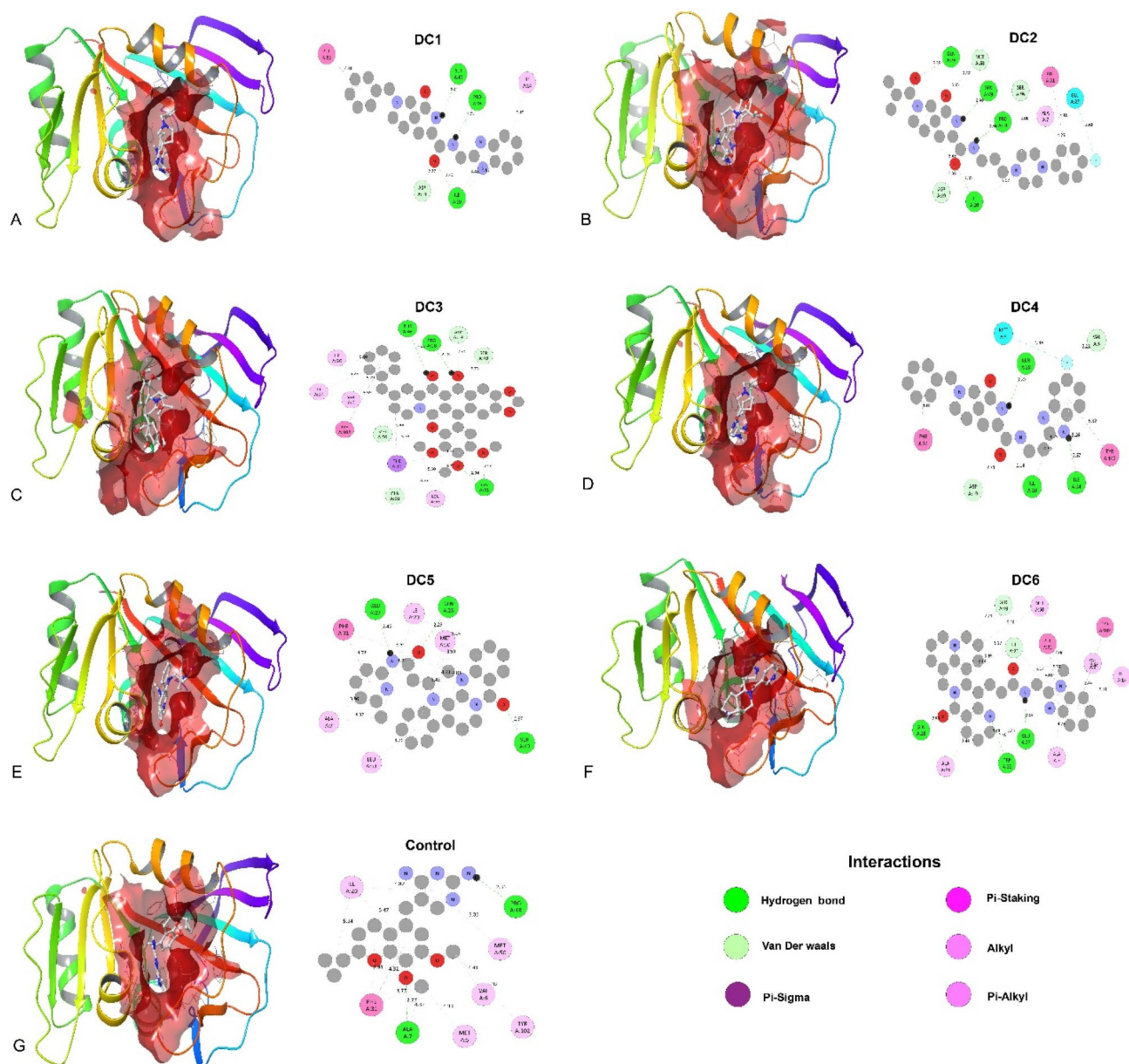
Fig. 1. Interaction of identified potential drug candidates (DCs) with the binding pockets of the DfrA1 protein. The interaction diagrams with highlighted (red) pocket regions and binding details. Interaction of DCs with DfrA1 via pocket residues, showing binding distance (**A**) DC1-DfrA1 complex. (**B**) DC2-DfrA1 complex. (**C**) DC3-DfrA1 complex. (**D**) DC4-DfrA1 complex. (**E**) DC5-DfrA1 complex. (**F**) DC6-DfrA1 complex. (**G**) Control-DfrA1 complex. The initial interaction and binding distance of each contact are presented in detail in each panel.

four H-bonds with the amino acids PRO18, ILE20, GLN28, and SER49, with bond lengths ranging from 2.29 to 2.98 Å. Additionally, seven weak non-covalent interactions were identified between DC2 and the binding pocket amino acids of the DfrA1 (Fig. 1B). DC3 anchored to the binding pockets via three H-bonds and ten non-covalent interactions, including Pi-stacking, Pi-alkyl, Pi-sigma, and Van der Waals interactions, marking the highest number of non-covalent interactions among the six compounds and the control (Fig. 1C). The electronegative oxygen and nitrogen atoms in DC4 formed H-bonds with ILE14, ILE20, GLN28 and SER49, with bond distances ranging from 2.14 to 3.20 Å, along with six non-covalent interactions within the complex (Fig. 1D). The pyridine-rich DC5 exhibited strong binding to the pocket amino acids primarily through non-covalent interactions, with GLU27, GLN28, and SER49 forming hydrogen bonds with DC5 at distances of 2.25 to 2.97 Å (Fig. 1E). DC6, characterized by its nitrogen-rich structure, demonstrated the second-highest number of non-covalent interactions within the pocket regions and formed three H-bonds with TRP22, GLU27, and GLN28, with bond lengths ranging from 2.48 to 2.83 Å (Fig. 1F). The control drug, Iclaprim, anchored within the binding pockets through two hydrogen bonds at distances of 2.3 to 2.77 Å and exhibited six non-covalent

interactions. Comparative analysis showed that the six DCs demonstrated significantly higher affinities to the binding pockets than the control drug (Fig. 1G and Table S1).

Prediction of drug-likeness and pharmacokinetic properties of DCs

The 'Rule of five' stipulates that DCs have good oral bioavailability which satisfied the following requirements: molecular weight (MW) ≤ 500 Da, MLogP (Moriguchi Log P) ≤ 4.15, hydrogen bond donors (HBD) ≤ 5, hydrogen bond acceptors (HBA) ≤ 10. Evaluation of six DCs against Lipinski's 'Rule of five' indicated that compounds DC1, DC2, DC4, and DC5 fulfilled all of the requirements, suggesting that they have acceptable oral bioavailability. DC3 and DC6 satisfied all the criteria except for molecular weight. Reference compound Iclaprim also followed the rule of five conditions. None of the molecules exhibited PAINS (Pan-Assay Interference Compounds) and Brenk alerts except for DC5, which displayed 1 PAINS alert (Table S2). Out of the six compounds analyzed, all were predicted to exhibit significant gastrointestinal (GI) absorption rates, ranging from 51.6 to 100%. Compounds DC1 and DC2 showed the highest water solubility with scores of -3.32 each, while compound DC3 displayed the lowest solubility with a score of -6.10. All six molecules were identified as P-gp substrates, indicating their involvement in drug transport and absorption. Their skin permeation coefficients (Log Kp) ranged from − 8.05 to -6.65 cm/s, suggesting moderate skin permeability. For drugs targeting peripheral sites, BBB (blood-brain barrier) penetration was undesirable due to the potential risk of CNS side effects; however, only DC5 was predicted to be BBB permeable. The steady-state volume of distribution (log L/kg) varied from − 0.32 to 0.89, with VDss classified as low when log VDss < -0.15 and high when log VDss > 0.45. Except for DC3, all molecules exhibited medium to high distribution rates. Metabolic analysis indicated that molecules not inhibiting CYP2D6 were likely to be well metabolized in the liver, with DC1 being the only CYP2D6 inhibitor. All six molecules demonstrated low plasma clearance (CL$_{plasma}$ < 5 ml/min/kg), whereas the reference compound exhibited moderate clearance (5–15 ml/min/kg) (Table S3). The mutagenic potential was closely linked to carcinogenicity, with DC1 and the reference compound showing a high probability of AMES mutagenicity. The remaining molecules exhibited moderate to excellent performance in this test, with probabilities ranging from 0.289 to 0.581, and all six molecules outperformed the reference compound in the carcinogenicity test. HEK293 (human embryonic kidney 293) cells were used for cytotoxicity assessment, where DC3 showed the highest probability (0.958) of cytotoxicity, while the other molecules ranged from moderate to excellent in this test. Regarding RPMI-8226 immunotoxicity, all molecules demonstrated excellent results, showing no immunotoxicity (probability: 0.027–0.176), except for DC3, which had a higher probability (0.447) (Table S4).

Molecular orbital analysis highlights the potential of DCs as drug candidates against DfrA1

A key objective in investigating the molecular orbitals of potential DfrA1 protein inhibitors was to elucidate their electronic properties. Our analysis focused on the energy gap (ΔEHOMO-LUMO), which provides insights into the electronic distribution and reactivity of these compounds. For DC1, the HOMO and LUMO energy levels were determined to be -5.72 eV and − 0.50 eV, respectively, yielding an energy gap (ΔEHOMO-LUMO) of 5.22 eV. The density of states (DOS) plot for DC2 revealed an energy gap (ΔEHOMO-LUMO) of 5.03 eV, which is lower than that of DC1. For DC3, the HOMO and LUMO energy levels were − 5.68 eV and − 0.22 eV, respectively, resulting in an energy gap (ΔEHOMO-LUMO) of 5.46 eV. According to the DOS plot for DC4, the energy gap (ΔEHOMO-LUMO) was 4.75 eV, with HOMO and LUMO energy levels at -5.59 eV and − 0.84 eV, respectively. For DC5, the HOMO and LUMO energy levels were − 5.46 eV and − 1.05 eV, respectively, resulting in an energy gap (ΔEHOMO-LUMO) of 4.41 eV. DC6 exhibited the lowest energy gap (ΔEHOMO-LUMO) of 3.18 eV, indicating the smallest energy gap among the six compounds (Table S5). In comparison, the DOS plot for the control compound showed a ΔEHOMO-LUMO of 4.77 eV. These ΔEHOMO-LUMO values suggest that some potential inhibitors are more reactive than the control drug, while others are less reactive (Fig. 2 and Table S5).

Electronic absorption spectra (UV-Vis) provide insights into molecular electronic transitions of DCs

We then focused on the electronic transitions of the molecules using DFT methods. Calculations of electronic absorption spectra confirmed the maximum absorbances and excitation energies for six selected molecules, including the control (Fig. S2). Among the six molecules, DC6 exhibited the highest peak intensity at 401.90 nm in the S0 → S2 excited state, with an excitation energy of 3.0849 eV and configurations of (-0.11274) H-6→L, (0.25326) H-3→L, and (0.62777) H-2→L. DC4 and DC5 showed absorption bands for S0 → S1 at 303.34 nm and 332.43 nm, with energies of 4.0873 eV (configurations: (0.27669) H→L, (0.63590) H→L + 1, (-0.11557) H→L + 2) and 3.7297 eV (configurations: (-0.34646) H-1→L, (0.18515) H-1→L + 1, (0.57349) H→L), respectively. DC1, DC2, and DC3 showed peaks at 278.86 nm, 257.30 nm, and 261.71 nm with excitation energies of 4.4460 eV, 4.8186 eV, and 4.7374 eV, respectively. The control drug, Iclaprim, displayed an absorption peak at 290.07 nm in the S0 → S3 transition with an excitation energy of 4.2743 eV (Table S6).

ESP calculations enhance understanding of the DCs chemical reactivity

As the six compounds preferentially bind to the binding site amino acids of the DfrA1 protein, we proceeded to analyze the surface characteristics of each DC and the control drug. The ESP charge ranges for DC1 through DC6 were − 0.142 to 0.142 a.u., -0.103 to 0.103 a.u., -0.109 to 0.109 a.u., -0.101 to 0.101 a.u., -0.127 to 0.127 a.u., and − 0.141 to 0.141 a.u., respectively. In contrast, the control drug, Iclaprim, exhibited a charge range of -0.0592 to 0.0592 a.u., which was lower than that of all the studied DCs (Fig. 3). The surface analysis of the potential DCs revealed that highly electronegative atoms appeared as red zones, while electropositive regions were depicted in blue. Neutral atoms were represented by green zones, and slightly electronegative and positive atoms formed yellow and turquoise zones, respectively. Notably, the DC compounds did not display extreme positive or
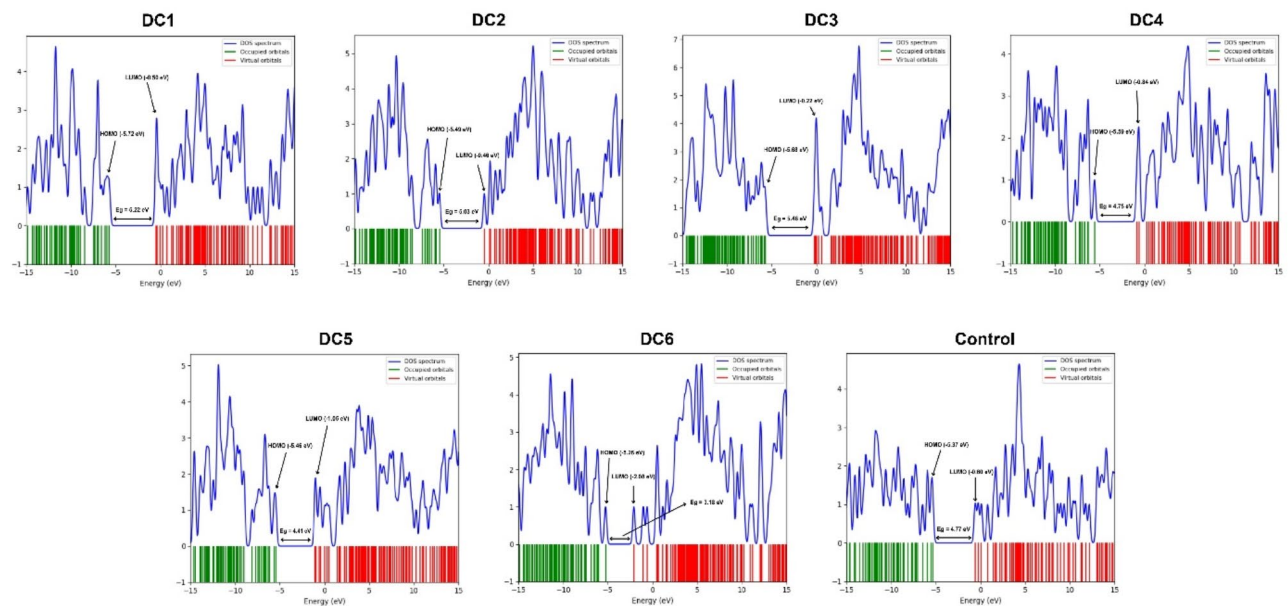
Fig. 2. Density of state (DOS) plot of HOMO–LUMO and energy gap of best-interacting DCs along with the control. The green and red dots in the diagrams show HOMO and LUMO orbitals.
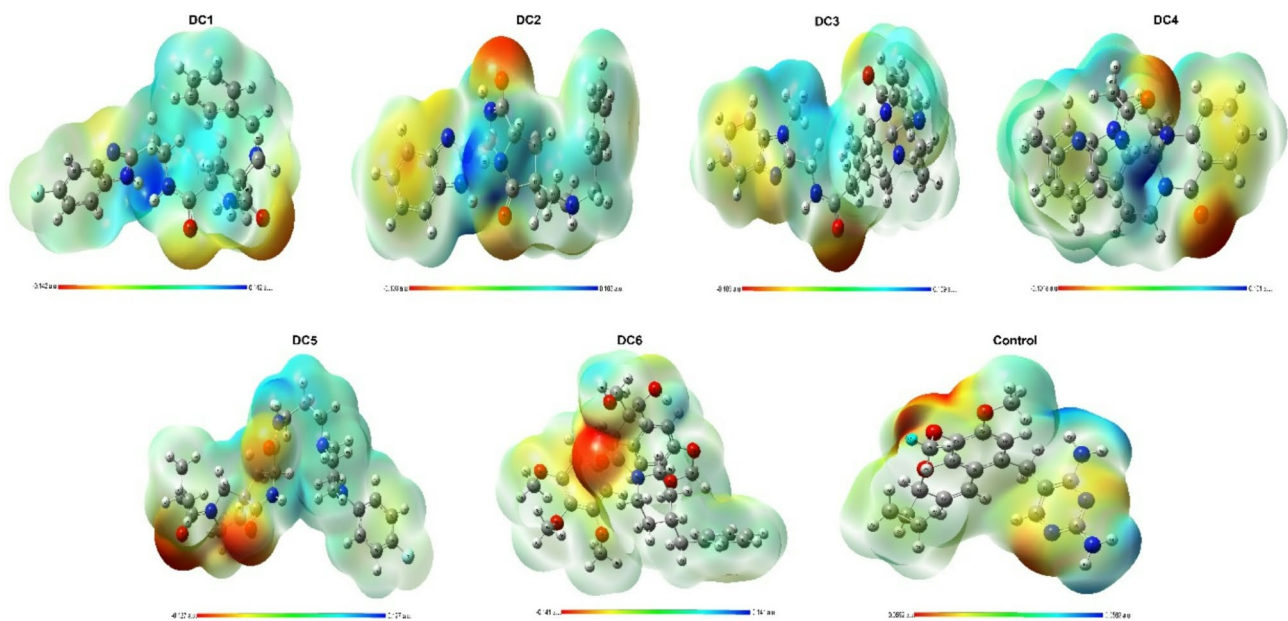

Fig. 3. The electrostatic surface potential (ESP) map of six potential inhibitors of DfrA1 and a control. Different colors are used to specify the charge distribution, such as blue, red, and green colors, which are used to determine the positive charge, negative charge, and neutral charge, respectively. Short electron density and poor interaction are identified in the blue region, but the red region contains high electron density and potential interaction.

negative ESP values, suggesting that they are less likely to experience non-specific interactions or exhibit weak binding with the target. The ESP values for the six compounds were similar and indicated better performance compared to the control drug (Fig. 3). Comparing molecular interactions and ESP surfaces revealed that the charged surfaces of the compounds contributed to covalent interactions with the binding pocket amino acids, while partially charged surfaces were often involved in non-covalent interactions (Figs. 1 and 3).

MDS of the drug candidates detects stable interactions with DfrA1 protein

One critical aspect of identifying a drug compound targeting a specific protein involved observing stable interactions between the receptor and ligand throughout simulation trials. For DC1, two stable interactions were observed: one between two oxygen atoms of DC1 and amino acids THR46 and TYR58 through the H-bonding, and another H-bond was formed between SER96 and the hydrogens positively charged carbon of DC1 (Fig. 4A). However, MDS for the interaction of the DC2 with binding pocket regions of DfrA1 displayed no stable covalent interactions, besides interaction fractions diagram showed weaker non-covalent interactions (Fig. 4B). For DC3, two negatively charged oxygen atoms of DC1 formed stable bonds throughout the simulation trajectories with amino acids THR46 and SER96. The electro-negative oxygen of the ALA24 anchored to the positively charged hydrogen of DC3, which led to the formation of another H-bond (Fig. 4C). The MDS for the DC4 interaction with binding pocket regions of DfrA1 demonstrated the most robust attachment by forming four H-bonds (2 by electrophilic attack and two by nucleophilic attack) via PRO18, ILE20, THR46 and SER49. The reported interaction fraction was also relatively high compared to other atoms (Fig. 4D). The DC5 showed three H-bonds and one non-covalent interaction, where H-bonds formed via one nucleophilic and two electrophilic
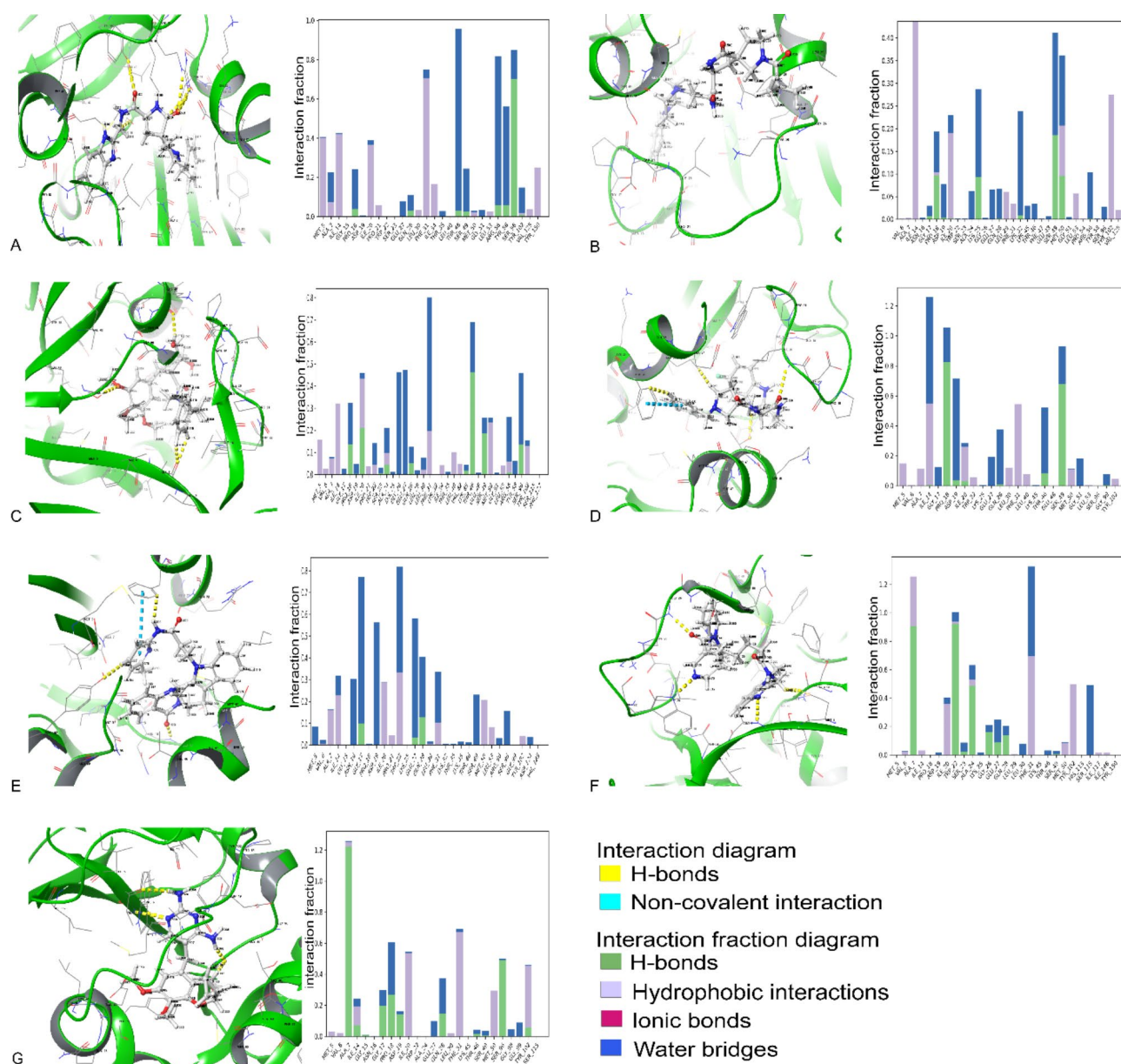

Fig. 4. Stable binding contact of identified potential drug candidates (DCs) with the pocket's residues of the DfrA1 protein. H-bonds and non-covalent interaction of DCs with DfrA1 via pocket residues and diagram of interaction strength (**A**) DC1-DfrA1 complex. (**B**) DC2-DfrA1 complex. (**C**) DC3-DfrA1 complex. (**D**) DC4-DfrA1 complex. (**E**) DC5-DfrA1 complex. (**F**) DC6-DfrA1 complex. (**G**) Control-DfrA1 complex. The interaction contact and strength of each contact are presented in detail in each panel.

reactions, and other non-covalent interaction formed by the pyridine ring (Fig. 4E). Another strong interaction was observed between DC6 and four amino acids: ALA7, TRP22, ALA24, and GLY26. DC6 engaged with ALA7, TRP22, and GLY26 via its electronegative oxygen and nitrogen atoms, while the electronegative oxygen of ALA24 attracted the hydrogen atoms of DC6. These interactions further demonstrate the affinity of DC6 for the binding site, contributing to its potential effectiveness as a drug candidate (Fig. 4F). For the control compound, three covalent interactions were identified, showing strong binding within the DfrA1 binding pocket. ALA7 was attracted to the control compound through highly negative nitrogen atoms and an electronegative side chain, forming a H-bond. Additionally, the electronegative side chain of SER96 attracted the positively charged hydrogen atoms, leading to further bonding (Fig. 4G). The interaction fraction analysis revealed that all the DCs consistently maintained stable interactions with the mutated amino acid GLN28 throughout the study (Fig. 4). However, overall findings indicated that DC1, DC4, and DC6 exhibit robust interactions with the target binding regions, in some cases demonstrating even stronger interactions than the control compound, Iclaprim.

MDS reveals the most potential inhibitors of the DfrA1 protein

Given the stable interactions observed with the DCs, we proceeded to analyze the Root Mean Square Deviation (RMSD) of the protein-ligand complexes during binding, as well as the Root Mean Square Fluctuation (RMSF) of the protein. The RMSD of the DfrA1 protein showed fluctuations of up to 2Å when interacting with all DCs, except DC1 and DC3. In contrast, the control compound exhibited fluctuations around 3.5Å, which is within an acceptable range. However, the lower RMSD values for the protein when bound to the DCs suggest that these compounds offer greater stability, making them more favorable than the control in terms of protein stability during interaction (Fig. 5A). The most significant aspect of the MDS was the ligand RMSD, crucial for assessing the strength and stability of the ligand within the binding pocket of DfrA1. The MDS results revealed that DC4 had the lowest RMSD value, consistently below 1Å, indicating strong stability during interaction. In contrast, DC2 exhibited the highest RMSD displacement, reaching nearly 6Å, suggesting instability when interacting with the binding pocket. Other DCs, as well as the control, exhibited displacements of up to 3Å, remaining within the acceptable stability threshold (Fig. 5B). The RMSF plots proved valuable for assessing local variations along the protein chain. Based on these plots, the amino acid regions at positions 20–35, 65–75, 120–125, and 145–150 displayed the highest fluctuations during the simulation compared to other regions. However, the reported fluctuations for these regions were approximately 3Å. The only anomaly was observed in the control compounds, which exhibited the highest displacement, around 5Å, in the 20–35 amino acid region. Additionally, higher RMSF values were noted for all other regions compared to the DCs (Fig. 5C).

DCs properties disclose their potential as effective drug candidates

The drug-likeness properties, such as Molecular Surface Area (MolSA), were typically calculated based on the van der Waals surface of the molecule. The maximum MolSA value was recorded at 500Å² for DC3, while the control compound showed the lowest value of around 325Å². The MolSA values for the other five DCs ranged between 350 and 450Å² (Fig. 6A). We then calculated the Solvent Accessible Surface Area (SASA) values for the DCs and the control drug. In MDS, SASA refers to the surface area of a compound that is accessible to solvent molecules, typically water. The results indicated that DC2 had the most accessible surface area, with SASA values ranging from 200Å² to 400Å². The second-highest SASA value was observed for DC3, ranging from 100Å² to 250Å². The SASA values for the other four DCs, along with the control, remained below 100Å² throughout the MDS trajectories (Fig. 6B). The Polar Surface Area (PSA) of a compound represents the surface area attributed to polar atoms, such as oxygen and nitrogen, along with their attached hydrogen atoms. The PSA values for DC1 – DC4 were recorded in the 150Å² range, although DC3 initially showed a lower value. Meanwhile, DC6 and DC4 exhibited PSA values of 120Å² and 100Å², respectively. Compared to the control compound, fluctuations were observed in the PSA curve, though the values remained close to those of the DC1 – DC4 compounds (Fig. 6C). The Radius of Gyration (rGyr) measures the compactness or distribution of a DC's atoms around its center of mass. By tracking rGyr over time during the simulation, conformational changes in the DCs were observed. The highest conformational displacement was recorded for DC2, DC1, and DC3. In contrast, the conformational displacement and mass distribution for the other compounds, as well as the control, displayed stable rGyr plots throughout the simulation period (Fig. 6D).

Free energy calculation reveals the strengths of interaction between DCs and DfrA1

One of the key aspects of drug discovery is the calculation of the MM/GBSA status, which is widely used to assess the stability and affinity of ligand binding during the drug discovery process. The calculated free energy for the r_psp_MMGBSA_dG_Bind_vdW module was ≤ -60 for all DCs except the control, indicating that DCs are more fit to the receptor than the control drug, Iclaprim. Positive values were reported for r_psp_MMGBSA_dG_Bind_Solv_GB for both the DCs and the control, suggesting that the binding of the DCs would not readily be influenced by the solvation environment. The reported negative values of ≤ -40 for r_psp_MMGBSA_dG_Bind_Lipo indicate the contribution of non-polar solvation energy (lipophilic interactions) to the binding free energy in MM/GBSA calculations. Subsequently, r_psp_MMGBSA_dG_Bind_Coulomb measured the contribution of Coulombic (electrostatic) interactions to the binding free energy in MM/GBSA calculations (Fig. 7). The consistently negative free energy values for this parameter across all DCs highlight the role of Coulombic forces in the binding affinity between the DCs and DfrA1. The final module of MM/GBSA, r_psp_MMGBSA_dG_Bind, was used to estimate the binding strength of DCs to the DfrA1 binding pocket. The ΔG Bind values for all compounds were higher than those for the control, with values ≤ -90, indicating a stronger and more stable binding to the targeted residues compared to the available drug (e.g., Iclaprim) (Fig. 7).
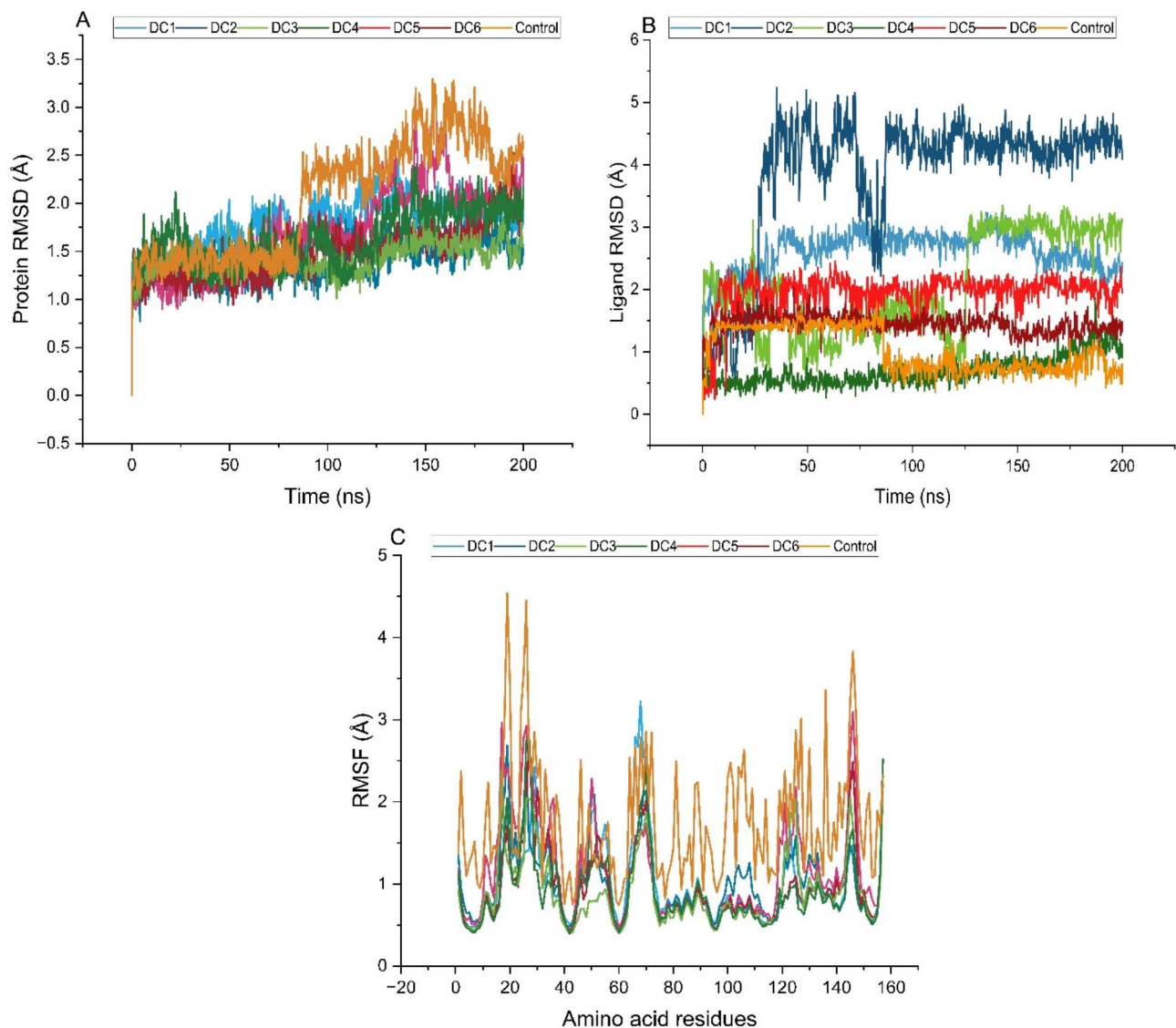
Fig. 5. Thermodynamic stability analysis of the protein, potential drug candidates (DCs), and amino acids over 200 ns simulation trajectories. (**A**) Overall fluctuations of DfrA1 during its interaction with DCs throughout the MDS period. (**B**) Overall fluctuations of all DCs and a control drug during their interaction with DfrA1 in the MDS period. (**C**) The RMSF curve shows the displacement of each amino acid throughout the MDS process.

Discussion

Trimethoprim (TMP) resistance in bacterial pathogens significantly limits treatment options, leading to more severe infections, particularly with *E. coli*, *P. aeruginosa* and *K. pneumoniae*[7,9,12]. The spread of resistance genes exacerbates this issue, increasing healthcare costs, prolonging illness, and posing a growing public health threat due to rising multidrug-resistant organisms. Thus, computational drug discovery approach may overcome TMP-resistance by employing structure-based design, virtual screening, *de-novo* drug design, and MDS profiling to identify novel, effective, and safer antibacterial therapeutics[29]. In this study, we evaluated 3,601 newly synthesized chemical compounds against the TMP-resistant DfrA1 protein using supercomputer-based optimization and screening to identify potential drug candidate(s). This research marks a significant step toward addressing AMR, particularly TMP resistance in bacterial pathogens like *E. coli*, *P. aeruginosa*, and *K. pneumoniae*, offering promising therapeutic solutions for combating infections caused by these superbugs. This innovative approach utilized advanced drug discovery tools for precise compound interaction analysis and semiempirical quantum mechanical methods to refine leading candidates' structures, enhancing their biological activity and paving the way for more targeted and effective therapeutic solutions[30]. The use of a cluster-based supercomputer greatly expedited the optimization process.

A major breakthrough of this study was the identification of six potential inhibitors (DC1 – DC6) targeting the TMP-resistant DfrA1 protein from a library of 3,601 optimized compounds. Large-scale screening combined
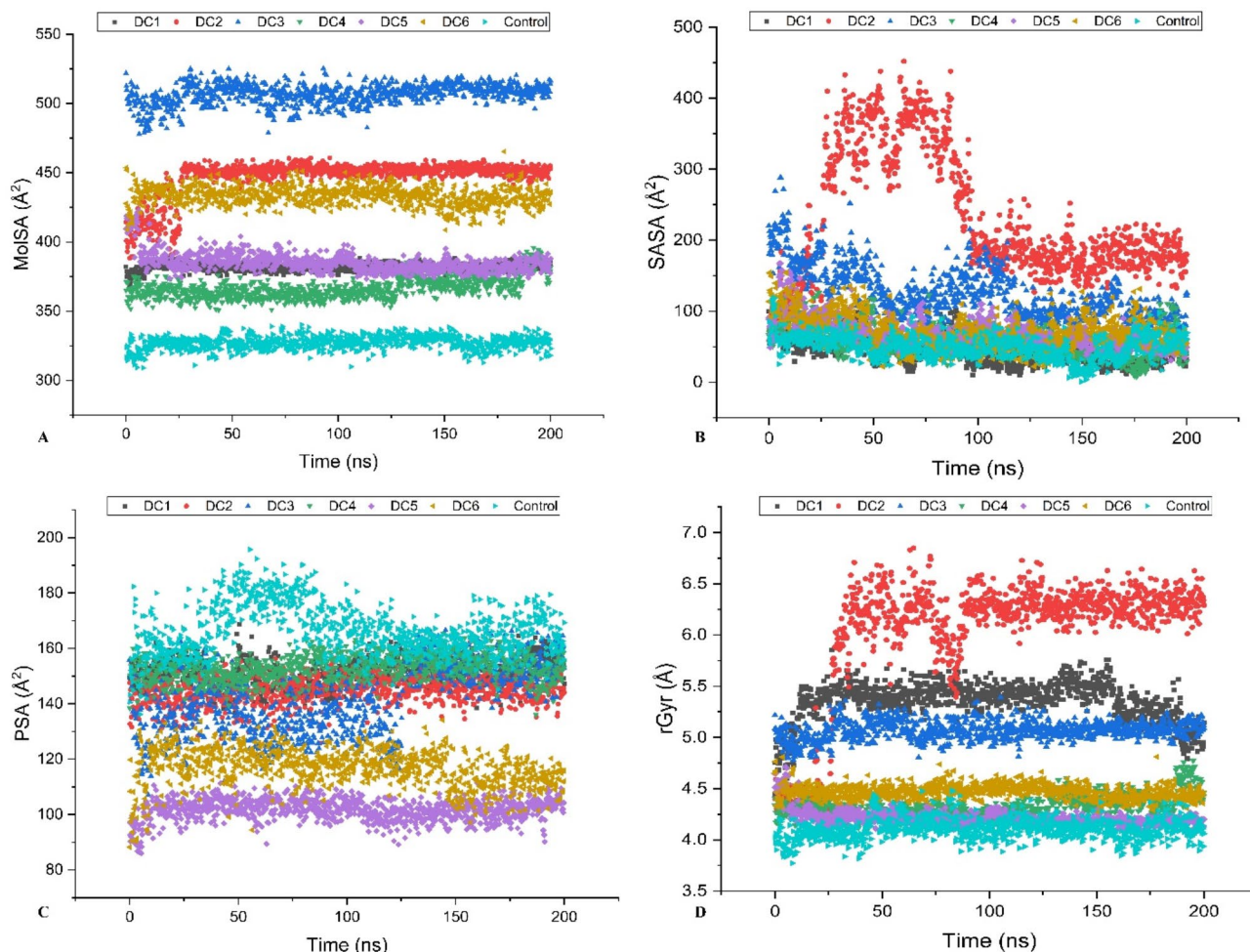
Fig. 6. Graphical representation of four key chemical properties of DCs, focusing on surface area, solvent exposure, polarity, and extendedness. The parameters (**A**) molecular surface area (MolSA), (**B**) solvent accessible surface area (SASA), (**C**) polar surface area (PSA), and (**D**) radius of gyration (rGyr) were calculated over the 200 ns simulation and are depicted in different colors on the plots, illustrating the variations in these properties throughout the simulation.

with computer-aided approaches has emerged as a standard and highly efficient strategy for identifying potential drug candidates, offering both precision and cost-effectiveness in selecting potent compounds[31,32]. In this study, the identification of DCs against TMP-resistant DfrA1 from a library of 3,600 chemical compounds demonstrated a notable improvement in selection accuracy. This aligns with findings from similar studies that highlight the effectiveness of large-scale screening methodologies in drug discovery[33,34]. Besides, these top candidates were selected for their robust H-bonding and non-covalent interactions, which are critical criteria in the drug discovery process[35,36]. These six DCs with strong H-bonding and lower binding free energies demonstrated favorable interactions with the target protein, DfrA1 compared to the control drug, Iclaprim. Moreover, DFT calculations for the six DCs revealed that DC6 had the lowest HOMO-LUMO energy gap (3.18 eV), indicating that DC6 is more reactive than the other DCs. HOMO-LUMO energy gap of Iclaprim was measured at 4.77 eV, closely aligning with the ranges of 4.41 to 5.46 eV observed for the other five DCs. Moreover, UV-Vis calculations evaluated the absorbance characteristics of the DCs, showing that DC6 exhibited the highest epsilon value (absorption coefficient) in comparison to the other candidates and the control drug. Besides, DC1, DC2, and DC3 exhibited absorption peaks at shorter wavelengths than the control, indicating lower excitation energy and enhanced reactivity, thereby suggesting their potential effectiveness as therapeutic agents. The lower excitation energy at these wavelengths suggests increased reactivity for these compounds, indicating their potential effectiveness as therapeutic agents[37,38]. The next phase of investigation focused on the ESP of the DCs. The higher ESP values observed for the DCs, relative to the control, suggest that these compounds possess highly interactive electrostatic surfaces, enhancing their potential for molecular interactions[39]. This characteristic enhances their potential as effective inhibitors of DfrA1, indicating that their electrostatic properties may promote stronger interactions with the target protein.

While the initial screening suggested strong interactions among six DCs with the DfrA1 protein, MDS revealed that only DC1, DC4, and DC6 maintained stable interactions, whereas DC2 exhibited no significant
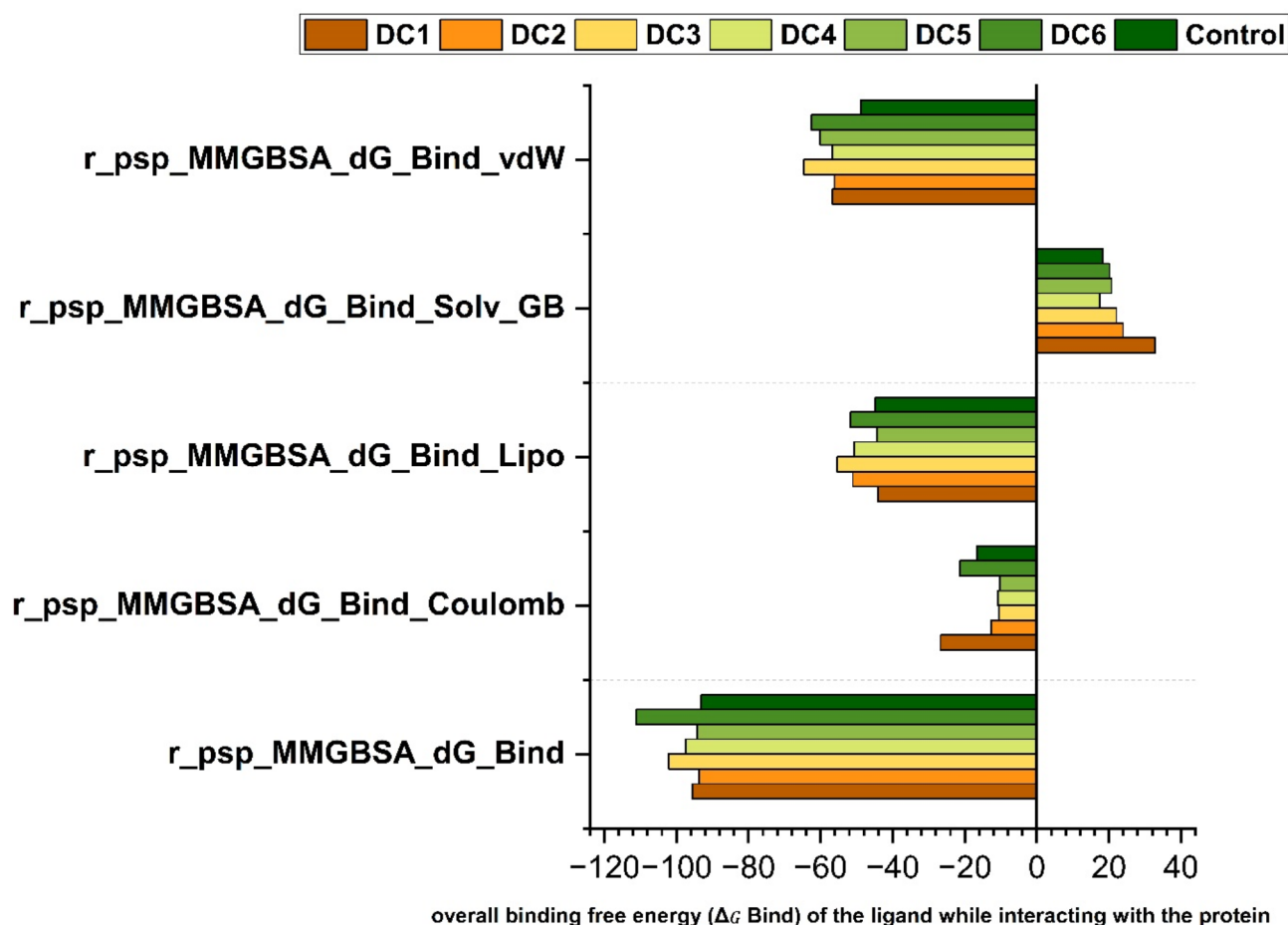
Fig. 7. Validation of thermal MM-GBSA analysis for six drug candidates (DCs) and control while interacting with DfrA1. Total MM-GBSA binding free energy was calculated, considering binding free energy, generalized Born electrostatic solvation energy, lipophilic effects, electrostatic energy (Coulombic effect), and total release free energy during interaction with the pocket residues of DfrA1. A bar graph with distinct colors represents each compound included in the study.

hydrogen bonding with the DfrA1 pocket residues. This stability and higher interaction fractions for DC1, DC4, and DC6 suggest their greater potential as effective inhibitors of DfrA1 (Fig. 4). Interestingly, the RMSD plots of DfrA1 while interacting with DCs showed fluctuations within a range of 1–3 Å, which is perfectly acceptable[40]. Notably, DfrA1 exhibited fewer displacements in the RMSD plots during interactions with the DCs compared to the control drug. Among these DCs, DC2 displayed the highest fluctuations in the RMSD plot. Additionally, the UV-Vis analysis revealed that DC2 exhibited unusual absorption characteristics, likely contributing to its instability. In contrast, only DC4 and DC6 demonstrated excellent RMSD values, remaining below 2 Å, which indicates strong and stable interactions with the DfrA1 binding pocket. This finding aligns with previous studies in computational drug chemistry, suggesting that such low RMSD values are indicative of reliable binding stability[41,42]. The analysis of four key parameters namely MolSA, SASA, PSA, and rGYr revealed that DC4 and DC6 were the most effective inhibitors of DfrA1, with values comparable to or better than the control drug. Furthermore, the MM/GBSA calculations supported these findings by evaluating the overall binding free energy (ΔG Bind) of DC4 and DC6 during their interactions with DfrA1. One of the critical issues for TMP resistance against DfrA1 protein is the amino acid substitutions at D27E and L28Q position[43]. While Iclaprim is effective against mutated DfrA1 because of its capacity to interact with the altered amino acids, the organofluorinated compound DC4 and the benzimidazole compound DC6 showed notable advantages over Iclaprim in terms of thermochemical behavior, molecular dynamics, and binding affinities. Benzimidazole and organofluorinated compounds are potent antibacterial agents commonly employed to combat infections caused by Gram-negative and certain Gram-positive bacteria[44,45]. A major contributor to TMP resistance is a mutation at amino acid position 28 of the DfrA1 protein[15,46,47]. Despite this mutation, Iclaprim remains effective against DfrA1 due to its strong binding affinity at this site. In our study, we observed that the identified compounds, DC4 and DC6, exhibited interaction patterns similar to Iclaprim with the mutated 28GLN residue and had robust interaction with nearby residues (Fig. 4). Notably, Iclaprim remains one of the most effective drugs for addressing infections mediated by DfrA1, particularly in cases involving TMP resistance. The strong binding affinities of DC4 (an organofluorinated compound) and DC6 (a benzimidazole compound) to the mutated residues of DfrA1, coupled

with their favorable pharmacokinetic profiles, suggest their potential as effective alternatives for combating TMP-resistant bacterial infections.

Extensive research has been devoted to the synthesis of organofluorine compounds due to their diverse biological activities, including antimicrobial, anticancer, antitumor, antihypertensive, antidepressant, anti-HIV, and stroke treatment applications[48,49]. In our study, we identified DC4 as a promising antimicrobial agent exhibiting strong activity against the DfrA1 protein, a key target in combating antibiotic resistance. Similarly, substituted benzimidazoles are well-recognized for their broad spectrum of biological activities, such as antiviral, antifungal, antimicrobial, antiprotozoal, anti-inflammatory, anticancer, antioxidant, anticoagulant, antidiabetic, and antihypertensive properties[50–52]. Our findings demonstrated the robust interactions of DC6 with the antibiotic-resistant DfrA1 protein, highlighting its potential as a potential drug candidate to address antibiotic resistance. Although this study provides valuable insights using in-silico computational methods, it is crucial to address its limitations, particularly the absence of experimental validation. Challenges in synthesizing the identified compounds, such as maintaining stability and activity during chemical modification, and difficulties in testing their efficacy in biological systems, could hinder translation to clinical applications. To bridge this gap, specific next steps should include wet-lab experiments, such as compound synthesis, in-vitro binding assays to evaluate their affinity for TMP-resistant DfrA1, and in-vivo studies to assess therapeutic efficacy and toxicity. Additionally, molecular functional characterization could provide deeper insights into the mechanisms of action, enabling the refinement of these drug candidates for enhanced potency and reduced off-target effects. These efforts are critical for advancing the identified compounds toward clinical trials and their potential use as effective treatments.

Conclusions

Antibiotics have long been the cornerstone of combating bacterial infections, yet the growing challenge of AMR underscores the critical need for alternative therapeutic options. This study aimed to identify potential inhibitors targeting the TMP-resistant DfrA1 protein, which contributes to development of AMR in a wide range of bacterial pathogens including *E. coli*, *K. pneumoniae* and *P. aeruginosa* both in humans and animals. Our high-throughput computational analysis reveals the efficacy of DC4 and DC6 as potential inhibitors of DfrA1, highlighting their strong inhibitory affinities. These findings suggest DC4 and DC6 as promising candidates in the battle against DfrA1-carrying resistant pathogens. The better inhibitory activity of the identified DCs compared to the FDA rejected drug Iclaprim highlights their potential as effective alternatives for combating the emerging TMP-resistant DfrA1 protein. While the findings of this study pave the way for discovering novel therapeutic candidates against TMP-resistant bacterial pathogens, further validation through real-world clinical experiments is essential.

Method and materials
Compounds refinement

A total of 3,601 newly synthesized chemical compounds were extracted from the ChemDiv (http://www.chemdiv.com/) database based on their adherence to Lipinski's Rule of five (e.g., molecular weight, log P, drug-likeness scores) and physicochemical features (e.g. chemical diversity, similarity to known inhibitors) indicative of their potentiality as inhibition. This selection ensured chemical diversity and relevance to the target protein[53,54]. The retrieved initial structures of the compounds were converted into molecular data files (.MOL) using the structure transformation tool in MarvinSketch v24.1.3[55]. Subsequently, all compounds underwent preliminary geometry optimization through the Optimize Geometry module in Avogadro[56]. To enhance compatibility with downstream analysis software, the files were converted to the PDB using Avogadro's program extension tool. For further refinement, these molecules were subjected to optimization using Gaussian 09 W (Gaussian, Inc., Wallingford CT, 2016) with the Universal Force Field (UFF) method, simulating a water solvent environment[57]. This optimization was performed on a supermicro-built, server-based supercomputer, significantly reducing processing time and facilitating efficient handling of the large dataset.

Protein system configuration

For our analysis, we targeted the DfrA1 enzyme, which is identical in both *E. coli* and *K. pneumoniae*. Crystallographic structures for this enzyme are available in the PDB database (7REG and 5ECX). We randomly selected PDB ID 5ECX, representing *K. pneumoniae* DfrA1 expressed in *E. coli*. DfrA1 is linked to TMP resistance and has been experimentally validated, was retrieved from the RCSB Protein Data Bank[47]. The three-dimensional (3D) structure was resolved using X-ray diffraction at a resolution of 1.95 Å. The DfrA1 structure was meticulously refined using Maestro v2024-4 (Schrödinger, LLC, New York, NY, 2024), which ensured the removal of undesirable interactions and optimized the engagement between potential binding compounds and the target protein during the molecular screening process[58]. For the refinement, the protein preparation was conducted with the protein preparation workflow of the Maestro with OPLS_2005 force field[59]. During the refinement process, missing side chains were added, hydrogen (H)-bonds were optimized, and excess water molecules were removed from the protein structures.

Prediction of drug-likeness and pharmacokinetic properties

In-silico computational pharmacokinetics methods were employed to determine drug absorption, distribution, metabolism, excretion, and toxicity (ADMET) which is a crucial step in drug design to ensure safety and efficacy. Three prediction tools were used in this analysis namely SwissADME, pkCSM, and ADMETlab 3.0. SwissADME is a comprehensive web-based tool for evaluating physicochemical properties such as drug-likeness and pharmacokinetics[60]. pkCSM uses graph-based signatures to predict ADMET features[61]. Another cutting-edge

platform that thoroughly analyzes and evaluates ADMET properties is ADMETlab 3.0[62]. Furthermore, to assess the drug-likeness, the compounds were evaluated against Lipinski's "Rule of five". The SwissADME program was used to assess the compounds for the present study based on their molecular weight, MLOGP, number of hydrogen bond donors and acceptors, TPSA, number of heavy atoms, rotatable bonds, bioavailability score, and warnings for PAINS and Brenk filters. Absorption (water solubility, GI absorption, p- glycoprotein substrate, skin permeation), distribution (BBB permeability, VDss), metabolism (CYP2D6 inhibitor), and excretion (CL_{plasma}) were predicted. In addition to the pharmacokinetic profile, toxicity predictions were conducted using pkCSM and ADMETlab 3.0, with a focus on several significant parameters. The AMES test was employed to estimate the chemicals' mutagenic potential, whereas carcinogenicity predictions examined the likelihood of causing cancer. To provide an extensive toxicity profile, cytotoxicity against HEK293 cells and immunotoxicity against RPMI-8226 (human myeloma) cells both were also assessed. These assessments were crucial for identifying any possible adverse effects that may occur during the development of these natural chemicals as medicinal compounds.

Binding simulation of protein and compounds

Prior to performing binding analysis, the SiteMap module of Maestro (Schrödinger, LLC) was utilized to determine the drug-binding side regions in the targeted DfrA1 protein[58]. The binding simulation between the DfrA1 and compounds was performed using the vina.bat scripts executed via PowerShell for Linux[63]. The configured protein structure was recognized as a biological target by the system, with parameters such as the center and dimensions included to create an optimal environment. The dimensions of the grid were x = 21.22, y = 18.36, and z = 19.96, and the coordinates for the grid center were x = 18.09, y = -20.72, and z = -1.87. The system's input energy was set to 64, and the exhaustiveness was configured to 128 to reduce the likelihood of false interactions between the protein and binding compounds. This approach enabled the simultaneous analysis of the binding efficiency of 3,601 compounds, streamlining the process and reducing complexity. Finally, six DCs such as (DC1 – DC6) that demonstrated strong chemical interactions, including H-bonding and non-covalent forces, with DfrA1 were selected for advanced analysis based on their proven effectiveness and promising results.

Orbital energy analysis

The calculations of HOMO and LUMO were conducted using Gaussian 09 W (Gaussian, Inc.)[64]. Assessing these frontier molecular orbitals (FMOs) was crucial for understanding the chemical reactivity of the DCs. DFT was employed to determine the electronic structure, utilizing the B3LYP method and the 6–31 g (d, p) basis set, with Opt + Freq analysis modules chosen for the calculation process under gas phase conditions to evaluate HOMO-LUMO gaps[65]. All the processors of the cluster-based supercomputer were utilized to perform the calculations. The energy gap calculation plots were generated using GaussSum[66].

Electrostatic surface potential and UV-Vis spectra of drug candidates

The electrostatic surface potential (ESP) and UV-Vis spectra for each of the six DCs, along with the control drug Iclaprim, were determined using Gaussian 09 W (Gaussian, Inc.). For the ESP analysis, the surface-contours module of Gaussian was utilized with the addition of a new cube action for the selected molecules. Initially, a cube was generated to calculate the total electron density of each compound using the SCF density matrix and a coarse grid with a point value of -2. After calculating the electron density, the ESP surface was analyzed based on the resulting molecular surfaces[67]. UV-Vis spectra were obtained using the vibrational spectra module in Gaussian (Gaussian, Inc.). The IR spectrum data, in a text-readable format, were exported from the UV-Vis plot and further analyzed using GraphPad Prism version 10 for Windows (GraphPad Software, San Diego, CA, USA).

Molecular dynamics modeling

To simulate the interactions between the DCs and DfrA1 (the biological target), we selected complexes that exhibited effective binding during the binding simulation analysis. These complexes were then prepared and subjected to energy minimization using the Protein Preparation Wizard, where missing side chains and loops were added using Prime (Schrödinger, LLC). MDS was then carried out using the Desmond MD engine[68]. The systems were solvated in a cubic box with a 10 Å edge length using the SPC water model, and ions were added to ensure a neutral net charge. The final setups included approximately 21,000 atoms. Prior to starting the production simulations, the systems underwent the standard relaxation protocol provided by Desmond. Production simulations were performed under an NPT ensemble, with pressure maintained at 1.01325 bar using the Martyna–Tobias–Klein method and temperature regulated at 300 K with the Nosé–Hoover thermostat[69]. A Coulombic cutoff radius of 9 Å was used, following the default settings. Each system was simulated for a duration of 200 ns (refer to **File S2** for more details).

Quantification and analysis of the drug candidate surface properties during simulation

The solvent accessible surface area (SASA) ($Å^2$), polar surface area (PSA) ($Å^2$), molecular surface area (MolSA) ($Å^2$), and radius of gyration (rGyr) (Å) of the DCs were meticulously quantified utilizing custom simulation.py scripts provided by Maestro (Schrödinger LLC)[58]. These calculations were conducted at 1 ns interval throughout the simulation, ensuring detailed temporal resolution. The surface dynamics of the compounds were then extracted from the trajectory files, resulting in a comprehensive dataset for further analysis. To facilitate clear and insightful visualization, the results were processed and graphically illustrated using OriginPro 2024 (OriginPro software, Northampton, USA) on a Windows platform, allowing for an in-depth exploration of the DCs surface properties over the course of the simulation.

Biding free energy calculation of the drug candidates

The binding free energy for the DCs interacting with the target protein DfrA1 was calculated using the MM/GBSA method, implemented via Maestro (Schrödinger LLC)[58]. Following a comprehensive 200 ns simulation, individual DfrA1 and DC structures were extracted from the simulation trajectory. These extracted structures were then input into the MM/GBSA framework to compute the free energy of interaction[69]. The calculation employed the novel variable dielectric surface generalized born (VSGB) solvation model, optimized for high-resolution protein structures, to ensure accurate estimation of solvation effects[70] (refer to **File S3** for more details).

Data availability

Data availabilityAll data generated or analysed during this study are included in this article [and its supplementary information files].

References

1. Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe, E. W. Jr. Computational methods in drug discovery. *Pharmacol. Rev.* **66**, 334–395. https://doi.org/10.1124/pr.112.007336 (2014).
2. Abdali, N. et al. Reviving antibiotics: efflux pump inhibitors that interact with AcrA, a membrane fusion protein of the AcrAB-TolC multidrug efflux pump. *ACS Infect. Dis.* **3**, 89 (2017).
3. Rosa, L. C. S., Argolo, C. O., Nascimento, C. M. C. & Pimentel, A. S. Identifying substructures that facilitate compounds to penetrate the blood-brain barrier via passive transport using machine learning explainer models. *ACS Chem. Neurosci.* **15**, 2144–2159. https://doi.org/10.1021/acschemneuro.3c00840 (2024).
4. Huo, T. et al. Late-stage modification of bioactive compounds: improving druggability through efficient molecular editing. *Acta Pharm. Sin. B* **14**, 1030–1076. https://doi.org/10.1016/j.apsb.2023.11.021 (2024).
5. Xu, Z. et al. Halogen bond: Its role beyond drug–target binding affinity for drug discovery and development. *J. Chem. Inf. Model.* **54**, 69–78. https://doi.org/10.1021/ci400539q (2014).
6. Miethke, M. et al. Towards the sustainable discovery and development of new antibiotics. *Nat. Reviews Chem.* **5**, 726–749. https://doi.org/10.1038/s41570-021-00313-1 (2021).
7. Kneis, D. et al. Trimethoprim resistance in surface and wastewater is mediated by contrasting variants of the DfrB gene. *ISME J.* **17**, 1455–1466 (2023).
8. Manna, M. S. et al. A Trimethoprim derivative impedes antibiotic resistance evolution. *Nat. Commun.* **12**, 2949 (2021).
9. Somorin, Y. M. et al. Antimicrobial resistance in urinary pathogens and culture-independent detection of trimethoprim resistance in urine from patients with urinary tract infection. *BMC Microbiol.* **22**, 144 (2022).
10. Pham, N. P. et al. Holistic understanding of trimethoprim resistance in *Streptococcus pneumoniae* using an integrative approach of genome-wide association study, resistance reconstruction, and machine learning. *mBio* **15**, e0136024. https://doi.org/10.1128/mbio.01360-24 (2024).
11. Ambrose, S. J. & Hall, R. M. DfrA trimethoprim resistance genes found in gram-negative bacteria: Compilation and unambiguous numbering. *J. Antimicrob. Chemother.* **76**, 2748–2756 (2021).
12. Li, X., Shi, L., Yang, W., Li, L. & Yamasaki, S. New array of aacA4-catB3-dfrA1 gene cassettes and a noncoding cassette from a class-1-integron-positive clinical strain of *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **50**, 2278–2279 (2006).
13. Muteeb, G., Rehman, M. T., Shahwan, M. & Aatif, M. Origin of antibiotics and antibiotic resistance, and their impacts on drug development: A narrative review. *Pharmaceuticals (Basel)* **16**. https://doi.org/10.3390/ph16111615 (2023).
14. Kotb, S. et al. Epidemiology of carbapenem-resistant *Enterobacteriaceae* in Egyptian intensive care units using National Healthcare-associated infections surveillance data, 2011–2017. *Antimicrob. Resist. Infect. Control.* **9** https://doi.org/10.1186/s13756-019-0639-7 (2020).
15. Wróbel, A., Arciszewska, K., Maliszewski, D. & Drozdowska, D. Trimethoprim and other nonclassical antifolates an excellent template for searching modifications of dihydrofolate reductase enzyme inhibitors. *J. Antibiot. (Tokyo)* **73**, 5–27. https://doi.org/10.1038/s41429-019-0240-6 (2020).
16. Hasnat, S. et al. Pantothenate kinase: A promising therapeutic target against pathogenic *Clostridium* species. *Heliyon* **10**, e34544. https://doi.org/10.1016/j.heliyon.2024.e34544 (2024).
17. Baassi, M. et al. Towards designing of a potential new HIV-1 protease inhibitor using QSAR study in combination with molecular docking and molecular dynamics simulations. *PloS One* **18**, e0284539. https://doi.org/10.1371/journal.pone.0284539 (2023).
18. van Mourik, T., Bühl, M. & Gaigeot, M. P. Density functional theory across chemistry, physics and biology. *Philos. Trans. Math. Phys. Eng. Sci.* **372**, 20120488. https://doi.org/10.1098/rsta.2012.0488 (2014).
19. Mumit, M. A. et al. DFT studies on vibrational and electronic spectra, HOMO-LUMO, MEP, HOMA, NBO and molecular docking analysis of benzyl-3-N-(2,4,5-trimethoxyphenylmethylene)hydrazinecarbodithioate. *J. Mol. Struct.* **1220**, 128715. https://doi.org/10.1016/j.molstruc.2020.128715 (2020)
20. Pantsar, T. et al. Decisive role of water and protein dynamics in residence time of p38α MAP kinase inhibitors. *Nat. Commun.* **13**, 569. https://doi.org/10.1038/s41467-022-28164-4 (2022).
21. Hollingsworth, S. A. & Dror, R. O. Molecular dynamics simulation for all. *Neuron* **99**, 1129–1143 (2018).
22. Albahra, S. et al. Artificial intelligence and machine learning overview in pathology & laboratory medicine: A general review of data preprocessing and basic supervised concepts. *Semin. Diagn. Pathol.* **40**, 71–87. https://doi.org/10.1053/j.semdp.2023.02.002 (2023).
23. Forouzesh, N. & Onufriev, A. V. MMGB/SA consensus estimate of the binding free energy between the novel coronavirus spike protein to the human ACE2 receptor. *BioRxiv* https://doi.org/10.1101/2020.08.25.267625 (2020).
24. Baker, S. J., Payne, D. J., Rappuoli, R. & De Gregorio, E. Technologies to address antimicrobial resistance. *Proc. Natl. Acad. Sci. U S A* **115**, 12887–12895. https://doi.org/10.1073/pnas.1717160115 (2018).
25. Hasnat, S. et al. Natural bacteriocins as potential drug candidates targeting core proteins in mastitis pathogens of dairy cattle. *bioRxiv*. https://doi.org/10.1101/2024.11.05.622021 (2024)
26. Fan, C., Basharat, Z., Mah, K. & Wei, C. R. Computational approach for drug discovery against *Gardnerella vaginalis* in quest for safer and effective treatments for bacterial vaginosis. *Sci. Rep.* **14**, 17437. https://doi.org/10.1038/s41598-024-68443-2 (2024).
27. Hossain, M. S. et al. Computational identification of Vernonia cinerea-derived phytochemicals as potential inhibitors of nonstructural protein 1 (NSP1) in dengue virus serotype-2. *Front. Pharmacol.* **15** https://doi.org/10.3389/fphar.2024.1465827 (2024).
28. Bassetti, M., Russo, A., Carnelutti, A. & Wilcox, M. Emerging drugs for treating methicillin-resistant *Staphylococcus aureus*. *Expert Opin. Emerg. Drugs* **24**, 191–204 (2019).

29. Islam, M. A. et al. In-silico study unveils potential phytocompounds in *Andrographis paniculata* against E6 protein of the high-risk HPV-16 subtype for cervical cancer therapy. *Sci. Rep.* **14**, 17182 (2024).

30. Christensen, A. S., Kubař, T., Cui, Q. & Elstner, M. Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications. *Chem. Rev.* **116**, 5301–5337. https://doi.org/10.1021/acs.chemrev.5b00584 (2016).

31. Sadybekov, A. V. & Katritch, V. Computational approaches streamlining drug discovery. *Nature* **616**, 673–685. https://doi.org/10.1038/s41586-023-05905-z (2023).

32. Yu, M. et al. Deep learning large-scale drug discovery and repurposing. *Nat. Comput. Sci.* **4**, 600–614. https://doi.org/10.1038/s43588-024-00679-4 (2024).

33. Costa, E. High-Throughput screening reveals new glutaminase inhibitor molecules. *ACS Pharmacol. Transl. Sci.* **4**, 1849–1866. https://doi.org/10.1021/acsptsci.1c00226 (2021).

34. Almihyawi, R. A. H. et al. Integrated computer-aided drug design and biophysical simulation approaches to determine natural anti-bacterial compounds for *Acinetobacter baumannii*. *Sci. Rep.* **12**, 6590. https://doi.org/10.1038/s41598-022-10364-z (2022).

35. Halder, S. K. et al. In Silico identification and analysis of potentially bioactive antiviral phytochemicals against SARS-CoV-2: A molecular Docking and dynamics simulation approach. *Biomed. Res. Int.* **2023**, 5469258. https://doi.org/10.1155/2023/5469258 (2023).

36. Papichettypalle, G. P. & Muthukumaradoss, K. Docking studies and molecular dynamics simulation of triazole benzene sulfonamide derivatives with human carbonic anhydrase IX inhibition activity. *RSC Adv.* **11**, 38079–38093. https://doi.org/10.1039/D1RA07377J (2021).

37. Mishra, J., Sharma, R. K. & Misra, K. in *Management of High Altitude Pathophysiology* (eds Misra, K., Sharma, P., & Bhardwaj, A.) 171–202 (Academic Press, 2018).

38. Negri, G., Kascheres, C. & Kascheres, A. J. Recent development in preparation reactivity and biological activity of enaminoketones and enaminothiones and their utilization to prepare heterocyclic compounds. *J. Heterocycl. Chem.* **41**, 461–491 (2004).

39. Rathi, P. C., Ludlow, R. F. & Verdonk, M. L. Practical high-quality electrostatic potential surfaces for drug discovery using a graph-convolutional deep neural network. *J. Med. Chem.* **63**, 8778–8790. https://doi.org/10.1021/acs.jmedchem.9b01129 (2020).

40. Manandhar, S. et al. Molecular dynamics and structure-based virtual screening and identification of natural compounds as Wnt signaling modulators: Possible therapeutics for Alzheimer's disease. *Mol. Divers.* **26**, 2793–2811. https://doi.org/10.1007/s11030-022-10395-8 (2022).

41. Patel, H. et al. Generating novel T790M/C797S-EGFR tyrosine kinase inhibitors to overcome the problem of mutation and resistance in non small cell lung cancer (NSCLC). *J. Biomol. Struct. Dyn.* **39**, 1–20. https://doi.org/10.1080/07391102.2020.1754918 (2020).

42. Jubb, H., Higueruelo, A. P., Winter, A. & Blundell, T. L. Structural biology and drug discovery for protein–protein interactions. *Trends Pharmacol. Sci.* **33**, 241–248 (2012).

43. Krucinska, J. et al. Structure-guided functional studies of plasmid-encoded dihydrofolate reductases reveal a common mechanism of trimethoprim resistance in gram-negative pathogens. *Commun. Biol.* **5**, 459. https://doi.org/10.1038/s42003-022-03384-y (2022).

44. Lungu, I. A., Moldovan, O. L., Biriş, V. & Rusu, A. Fluoroquinolones hybrid molecules as promising antibacterial agents in the fight against antibacterial resistance. *Pharmaceutics* **14**, 1749 (2022).

45. Song, D. & Ma, S. Recent development of benzimidazole-containing antibacterial agents. *ChemMedChem* **11**, 646–659 (2016).

46. Zhang, Y., Chowdhury, S., Rodrigues, J. V. & Shakhnovich, E. Development of antibacterial compounds that constrain evolutionary pathways to resistance. *Elife* **10** https://doi.org/10.7554/eLife.64518 (2021).

47. Lombardo, M. N., Wright, N. G. D., Anderson, A. C. & D. L. & Crystal structures of trimethoprim-resistant DfrA1 rationalize potent Inhibition by propargyl-linked antifolates. *ACS Infect. Dis.* **2**, 149–156. https://doi.org/10.1021/acsinfecdis.5b00129 (2016).

48. Remete, A., Nonn, M., Escorihuela, J., Fustero, S. & Kiss, L. Asymmetric methods for carbon-fluorine bond formation. *Eur. J. Org. Chem.* **2021** https://doi.org/10.1002/ejoc.202101094 (2021).

49. Inoue, M., Sumii, Y. & Shibata, N. Contribution of organofluorine compounds to pharmaceuticals. *ACS Omega* **5**, 10633–10640. https://doi.org/10.1021/acsomega.0c00830 (2020).

50. Ates-Alagoz, Z. Antimicrobial activities of 1-H-benzimidazole-based molecules. *Curr. Top. Med. Chem.* **16**, 2953–2962. https://doi.org/10.2174/1568026616666160506130226 (2016).

51. Kabi, A. K. et al. In *Nanostructured Biomaterials: Basic Structures and Applications*. 351–378 (eds Swain, B. P.) (Springer Singapore, 2022).

52. Salahuddin, S., Shahar Yar, M. & Mazumder, A. Benzimidazoles: A biologically active compounds. *Arab. J. Chem.* **5** https://doi.org/10.1016/j.arabjc.2012.07.017 (2012).

53. Flores-Padilla, E. A. et al. Chemoinformatic characterization of synthetic screening libraries focused on epigenetic targets. *Mol. Inf.* **41**, 2100285 (2022).

54. Leeson, P. Drug discovery: Chemical beauty contest. *Nature* **481**, 455–456. https://doi.org/10.1038/481455a (2012).

55. Csizmadia, P. MarvinSketch and MarvinView: Molecule applets for the World Wide Web. (1999).

56. Hanwell, M. D. et al. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 1–17 (2012).

57. Tomberg, A. Gaussian 09w tutorial. *An introduction to computational chemistry using G09W and Avogadro software*, 1–36 (2013).

58. Maestro, S. Maestro *Schrödinger, LLC, New York, NY* 682 (2020). (2020).

59. Shivakumar, D. et al. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput.* **6**, 1509–1519 (2010).

60. Daina, A., Michielin, O. & Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **7**, 42717. https://doi.org/10.1038/srep42717 (2017).

61. Pires, D. E., Blundell, T. L. & Ascher, D. B. pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* **58**, 4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104 (2015).

62. Fu, L. et al. ADMETlab 3.0: An updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. *Nucleic Acids Res.* **52**, W422–w431. https://doi.org/10.1093/nar/gkae236 (2024).

63. Ifran, M., Mudalige, H. & Perera, O. Protein-ligand docking study for the identification of binding sites and ligands against the ischemic stroke receptors. (2022).

64. Sevvanthi, S. et al. PES, molecular structure, spectroscopic (FT-IR,), electronic (UV-Vis, HOMO-LUMO), quantum chemical and biological (docking) studies on a potent membrane permeable inhibitor: Dibenzoxepine derivative. *Heliyon* **6** (2020).

65. Jacquemin, D., Planchat, A., Adamo, C. & Mennucci, B. TD-DFT assessment of functionals for optical 0–0 transitions in solvated dyes. *J. Chem. Theory Comput.* **8**, 2359–2372 (2012).

66. O'boyle, N. M., Tenderholt, A. L. & Langner, K. M. Cclib: a library for package-independent computational chemistry algorithms. *J. Comput. Chem.* **29**, 839–845 (2008).

67. Chrostowska, A. et al. UV-Photoelectron spectroscopy of BN Indoles: Experimental and computational electronic structure analysis. *J. Am. Chem. Soc.* **136**, 11813–11820. https://doi.org/10.1021/ja5063899 (2014).

68. Bowers, K. J. et al. In *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*. 84-es.

69. Skrdla, P. J. et al. Complexation mechanisms of aqueous amylose: Molecular dynamics study using 3-Pentadecylphenol. *Mol. Pharm.* **21**, 3540–3552. https://doi.org/10.1021/acs.molpharmaceut.4c00235 (2024).

70. Wang, E. et al. Development and evaluation of MM/GBSA based on a variable dielectric GB model for predicting protein–ligand binding affinities. *J. Chem. Inf. Model.* **60**, 5353–5365 (2020).

Author contributions

S.H., M.M.M., T.I. and M.N.H. conceived and designed the study; S.H., S.R., M.B.A., F.M.S., F.Y., T.S., N.T.S., S.S., A.C. and A.D.A.S. were responsible for data acquisition, analysis, and writing the original draft. M.M.M., T.I. and M.N.H. conducted the critical review and editing. All authors contributed to the article and approved the final submitted version.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-91410-4.

Correspondence and requests for materials should be addressed to M.M.M., T.I. or M.N.H.

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.