ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Computational Screening of Potential Inhibitors of *Desulfobacter postgatei* for Pyrite Scale Prevention in Oil and Gas Wells

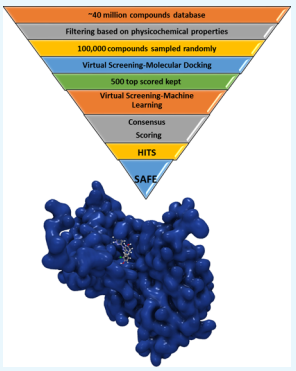
Abdulmujeeb T. Onawole, Ibnelwaleed A. Hussein,* Mohammed A. Saad, Musa E.M. Ahmed, and Hassan Nimir*

Cite This: *ACS Omega* 2021, 6, 10607−10617

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Sulfate-reducing bacteria (SRB), such as *Desulfobacter postgatei* are found in oil wells. However, they lead to the release of hydrogen sulfide. This in turn leads to the iron sulfide scale formation (pyrite). ATP sulfurylase is an enzyme present in SRB, which catalyzes the formation of adenylyl sulfate (APS) and inorganic pyrophosphatase (PPi) from ATP and sulfate. This reaction is the first among many in hydrogen sulfide production by *D. postgatei.* Consensus scoring using molecular docking and machine learning was used to identify three potential inhibitors of ATP sulfurylase from a database of about 40 million compounds. These selected hits ((*S*,*E*)-1-(4-methoxyphenyl)-3-(9-((*m*-tolylimino)methyl)-9,10-dihydroanthracen-9-yl)pyrrolidine-2,5-dione; methyl 2-[[(1*S*)-5-cyano-2-imino-1-(4-phenylthiazol-2-yl)-3-azaspiro[5.5]undec-4-en-4-yl]sulfanyl]-acetate; and (4*S*)-4-(3-chloro-4-hydroxy-phenyl)-1-(6-hydroxypyridazin-3-yl)-3-methyl-4,5-dihydropyrazolo[3,4-b]pyridin-6-ol), known as A, B, and C, respectively) all had good binding affinities with ATP sulfurylase and were further analyzed for their toxicological properties. Compound A had the highest docking score. However, based on the physicochemical and toxicological properties, only compound C was predicted to be both safe and effective as a potential inhibitor of ATP sulfurylase, hence the preferred choice. The molecular interactions of compound C revealed favorable interactions with the following residues: LEU213, ASP308, ARG307, TRP347, LEU224, GLN212, MET211, and HIS309.



1. INTRODUCTION

Sulfate-reducing bacteria (SRB) have been a persistent problem in the oil and gas industry as their presence in oil and gas reservoirs are abundant. SRB are one of the main sources of hydrocarbons as one of the predominant sources of sulfide and sulfur during the maturation of oil reservoirs.[1,2] They help in sulfur formation by the incomplete oxidation of sulfur. This sulfur in turn leads to serious problems, such as iron sulfide (pyrite) scale formation that hinders the injectivity of water injection wells, corrosion of iron, and contamination of produced gas by generating hydrogen sulfide (H_2S).[3−6] These problems sometimes lead to deaths during oil and gas production, such as the fatalities caused by hydrogen sulfide poisoning during offshore operations in the North Sea.[7] SRBs have been tagged has the major perpetrators of microbial-influenced corrosion.[8−10]

SRBs operate mostly in anaerobic conditions and use sulfate as an electron acceptor during their metabolism of energy.[3] They are prevalent in water and land environments as long as they are anoxic and such anoxic environments include oil and natural gas wells.[11] *D. postgatei* is one of the common SRBs that are rod-like or elliptical. Unlike other SRBs, *D. postgatei* converts acetate to carbon dioxide (CO_2) and H_2S[3,12] as described in eq 1.

$$2CH_3COO^- + 2SO_4^{2-} \rightarrow 4CO_2 + 2H_2S + 4H_2O + 6H^+ \quad (1)$$

However, in the reduction of sulfate to H_2S, some enzymes are involved (Table 1), which make the reaction feasible. ATP sulfurylase catalyzed the first reaction in the series.[13,14]

Table 1. Reactions Involved in Respiratory Sulfate Reduction

reaction	enzyme involved
$SO_4^{2-} + ATP^{4-} \rightarrow APS^{2-} + PPi^{4-}$	ATP sulfurylase
$PPi + H_2O \rightarrow 2\ Pi$	inorganic pyrophosphatase
$APS + H^+ + 2e^- \rightarrow HSO_3^- + AMP$	APS reductase
$HSO_3^- + 6H^+ + 6e^- \rightarrow HS^- + 3H_2O$	bisulfite reductase

Existing solutions to alleviate the effects of SRBs include the use of biocides like aldehydes and amines, which are often hazardous and mutagenic.[15] Also, dissolved oxygen and nitrates facilitate an oxidizing environment. However, a few SRBs thrive under aerobic conditions.[7,11,14,16] Also, heavy metals are used, but they are mostly toxic.

Received: December 14, 2020
Accepted: February 2, 2021
Published: April 13, 2021

```
D.      MSKLVAPHGGKGLVCCKL-EGAALEAERKKAAGLKKIEISSQVKGDLIMLGIGGFSPLNG    59
A.      ---MIKPVGSDELRPRFVYDPEQHHRLSSEAESLPSVIVSSQAAGNAVMLGAGYFSPLDG    57
           :: * *.. *    : :    .   .:* .* .: :***. *: :*** * ****:*

D.      FMTKADWKSVCEDFLLANGTFWPVPVMLDASAADAAAIKVGDEITLER----NGEIYATM    115
A.      FMNLADALSSAQSMTLTDGRFFPVPLLCLLES--ADAIAGATRIALRDPNVEGNPVLAVM    115
        **. **  * .:.: *::* *:***::   .:  * **  . .*:*.     .. : *.*

D.      KIEEKFEMTEAEKKWECEKVYKGHGEESDDKVFWEIALKDHPGVQMVMARKEFCLAGPVK    175
A.      DVTAVEQVSDAQMALMTEQVYGTSD------------PKHPGVETFNSQGRTAISGPIQ    162
        .:    ::::*:     *:**   .              .****: . :: . .::**::

D.      VLSEGEFPEKFKGVYLTPAETRAIMDEKGWANVASMQLRNPMHRSHEHLCKIALDV--CD    233
A.      VLNFSYFQTDFPDTFRTAVEIRHEIQERGWQKIVAFQTRNPMHRAHEELCKMAMEAVEAD    222
        **. . *  .* ..: * .* *  ::*:** ::.::* ******:**.***:*::.  .*

D.      GVLIHSLIGNLKPGDIPADVRIKCIDTLIKGYFVPEHVINAGYPLDMRYAGPREALLHAT    293
A.      GVVIHMLLGQLKPGDIPAPVRDAAIRTMAELYFPPNTVMVTGYGFDMLYAGPREAVLHAY    282
        **:** *:*:******** **  .* *: : ** *: *: :** :** *******:***

D.      FRQNYGVNKMIIGRDHAGVGDFYTLFEAQEIFDTIPTPEDPGKRLLCEPLKIDWTFYCHK    353
A.      FRQNMGATHFIIGRDHAGVGDYYGPFDAQTIFDDA----VPTDVLAIEIFRADNTAYSKK    338
        **** *.::***********:*  *:** ***       * . *  * :: * * *.:*

D.      CDGMASMRTCPH-GKDDRVILSGTKLRHALSNNQPVVDHFGREEVLVILREYYASLTEKV    412
A.      LGRVVMMRDAPDHTPDDFIQLSGTRVREMLGQGEAPPPEFSRPEVAQILMDYYRSLPQ--    396
         . :. ** .*.   ** : ****::*. *.:.:    .*.* **  ** :** ** :

D.      EVKLQSHAEGTKM   425
A.      -------------   396
```

Figure 1. Protein sequence of ATP sulfurylase of *D. postgatei* (D.) and *A. vinosum* (A.) (PDB ID: 4DNX) in clustal format. Similar residues in both organisms have the same color code in the sequence.

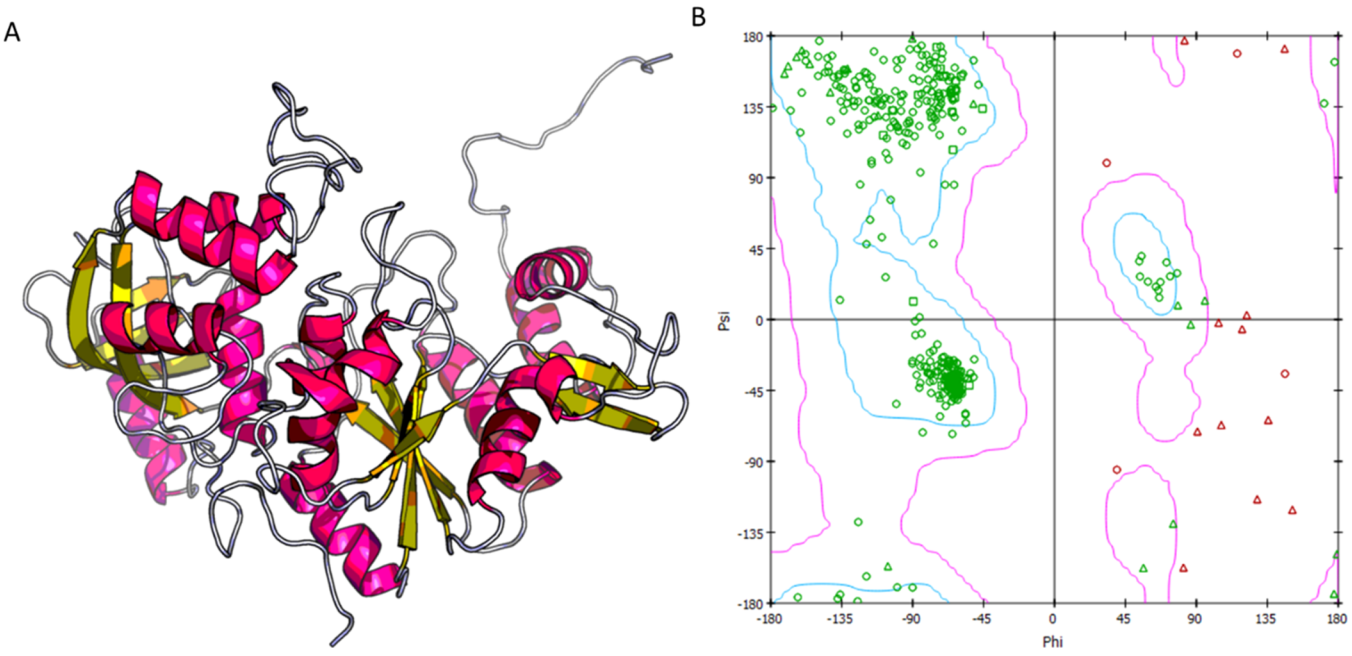


Figure 2. (a) Tertiary structure and (b) the Ramachandran plot of ATP sulfurylase (*D. postgatei*). (Red: helix; yellow: strand; white: coil; squares: proline; triangles: glycine; squares: proline; circle: others).

Recent works[17−19] have proffered environmentally friendly formulations that can be used in scale removal. These formulations include chelating agents such as DTPA (diethylenetriamine pentaacetate) acid, which removes oil field scales by chelating with the iron in the scale. This is much better than using HCl, which exacerbates the situation by producing the toxic hydrogen sulfide gas. However, these formulations do not solve the problem from the origin. Hence, a root cause analysis solution is imperative in preventing iron sulfide scale formation. This could be achieved by eliminating SRBs present in oil and gas reservoirs, which is the central idea of this work. The elimination of SRBs is possible by inhibiting the enzymes responsible for sulfate reduction present in SRB. Computational techniques such as virtual screening using molecular docking and artificial intelligence techniques, such as machine and/or deep learning, have been extensively used in the pharmaceutical industry in discovering and designing novel compounds that inhibit certain proteins to create a

https://dx.doi.org/10.1021/acsomega.0c06078
ACS Omega 2021, 6, 10607−10617

pharmacological effect.[20−24] Moreover, machine-learning tools have been suggested to give a synergistic effect in finding improved hit rates.[25−27] This process can be translated to oil field chemistry in finding a safer and longer lasting solution to sulfur production by SRB. Moreover, these methods can help in screening large databases of compounds and also predict their toxicity properties, which can help as a guide in selecting safe compounds that can be used as inhibitors of ATP sulfurylase.[28−30]

There is a dearth of literature using this approach in reducing sulfur production in oil fields. A relatively recent work[31] screened about 15 compounds to find potential inhibitors for *Archaeoglobus fulgidus,* which though not a bacterium but is also a sulfate-reducing organism. However, the number of compounds screened was quite small and this reduced the possibility of finding an effective compound that will be able inhibit the enzymes responsible for sulfate reduction. Herein, in this work, we use computer-aided drug design (CADD) techniques, such as homology modeling, molecular docking, and machine learning, to select hits via consensus scoring by screening 100,000 compounds from a database of about 40 million compounds. Unlike the earlier work mentioned that screened 15 compounds, this work screens 100,000 compounds from a large database of about 40 million compounds. The selected hits from this work have the potential to inhibit ATP sulfurylase, hence alleviating the threat of SRBs in iron sulfide scale formation particularly those scales formed by pyrite. This will consequently increase production in the oil and gas industry.

2. RESULTS AND DISCUSSION

2.1. Protein Structure and Validation. Homology modeling is a renowned tool in computer-aided drug design that has been used in finding potential inhibitors for diseases.[32,33] The protein sequence of both the template protein (*A. vinosum*) and the target protein (*D. postgatei*) showed similarities (Figure 1)in their sequences using the clustal format.[34] The structure was validated using a Ramachandran plot (Figure 2).

The homology-modeled structure (Figure 2a) depicted about 10 helices, which are connected by strands and coils. With the aid of the Ramachandran plot as a validation tool (Figure 2b). The Ramachandran plot depicts the allowed values of ψ (psi) versus φ (phi) angles corresponding to many confirmations, for a specific amino acid.[35] The structure is made up of 423 residues with 389, 27, and 7 being in the favored, allowed, and outlier regions, respectively. The upper and lower left regions of the Ramachandran plot correspond to the β-pleated and right-handed α-helices, while the middle of the right region indicates the left-handed α-helices. For a superior quality protein, it is expected that the outlier percentage should not exceed 5%, while the favored and allowed regions should have a combined total of more than 90%.[36−38] The seven residues in the outlier make up 1.7%, while the favored and allowed regions both constitute 98.3%; hence, the structure is of excellent quality and hence is reliable for molecular docking.

Besides the Ramachandran plot, the alignment RMSD (root-mean-square deviation) is calculated for only the core residues of both the homology-modeled and template proteins (Table S1). The RMSD value is mostly within the recommended range of 0 to 1.2 Å for most of the core residues.[39] Nevertheless, the quality of the modeled protein can be analyzed by several parameters[40,41] (Table 2). The score is the alignment score that falls between 0 and the domain sequence length. A score of 0 implies that the protein is of the worst quality; hence, the higher the score the better the quality. The score value is 363; hence, it implies it is of good quality. The *P*-value indicates the relative quality of the model such that the smaller the *P*-value, the better the model quality. A *P*-value that is less than 0.0001 is considered a model of good quality. The *P*-value of the modeled protein is 3.80×10^{-13}, which implies a good quality model. The uGDT (Global Distance Test) measures the absolute model quality. For a protein with more than 100 residues, such as the ATP sulfurylase of *D. postgatei* (423 residues), a uGDT value that is greater than 50 indicates a good quality model,[42] which is correct for the homology model (309). The uSeqID is the number of identical residues in the alignment. A higher uSeqID value implies a better model, which is also correct for this model (157).

Table 2. Model Parameters of the Homology-Modeled Structure of ATP Sulfurylase of *D. postgatei*

parameter	value
score	363
P-value	3.80×10^{-13}
uGDT	309
uSeqID	157

2.2. Docking Simulations. Using the vote rank approach, the consensus scoring (Figures 3 and 4) from the two virtual screenings (molecular docking and machine learning) resulted in the selection of three compounds as hits, namely, MCULE-798447760801 (A), MCULE-123221368002 (B), and MCULE-938606270701 (C), which have the following IUPAC names: (*S*,*E*)-1-(4-methoxyphenyl)-3-(9-((*m*-tolylimino)methyl)-9,10-dihydroanthracen-9-yl)pyrrolidine-2,5-dione; methyl 2-[[(1*S*)-5-cyano-2-imino-1-(4-phenylthiazol-2-yl)-3-azaspiro[5.5]undec-4-en-4-yl]sulfanyl]acetate; and (4*S*)-4-(3-chloro-4-hydroxy-phenyl)-1-(6-hydroxypyridazin-3-yl)-3-methyl-4,5-dihydropyrazolo[3,4-b]pyridin-6-ol. These compounds would henceforth be referred to as compounds A, B, and C (Figure 5). Compound A was the 5th and 21st top-scored compound for AutoDock Vina and K_{DEEP}, respectively, while compound B corresponds to the 11th and 3rd top-scored compound for AutoDock Vina and K_{DEEP}, respectively. Compound C was the 17th and 20th top-scored compound for AutoDock Vina and K_{DEEP}, respectively. All the selected compounds had good binding free energies from the molecular docking score that corresponds to −8.7, −8.4, and −8.2 kcal/mol for compounds A, B, and C, respectively. The negative values of the docking score implied that they all bound to the target protein with compound A having the highest binding energy. The pK_d values from K_{DEEP} were 6.56, 7.03, and 6.58 for compounds A, B, and C, respectively, with compound B having the highest pK_d value (Tables S1 and S2, Supporting Information).

2.3. Binding Modes and Molecular Interactions of Selected Compounds. The molecular docking provided insight into the mode of binding and the molecular interactions of the selected compounds in the protein target. Compound A showed many favorable molecular interactions with the amino residues in the binding pocket of the target protein (Figure 6). They include hydrogen bonding with GLU287 and 212; alkyl interactions with ALA310, LEU318,

https://dx.doi.org/10.1021/acsomega.0c06078
ACS Omega 2021, 6, 10607−10617

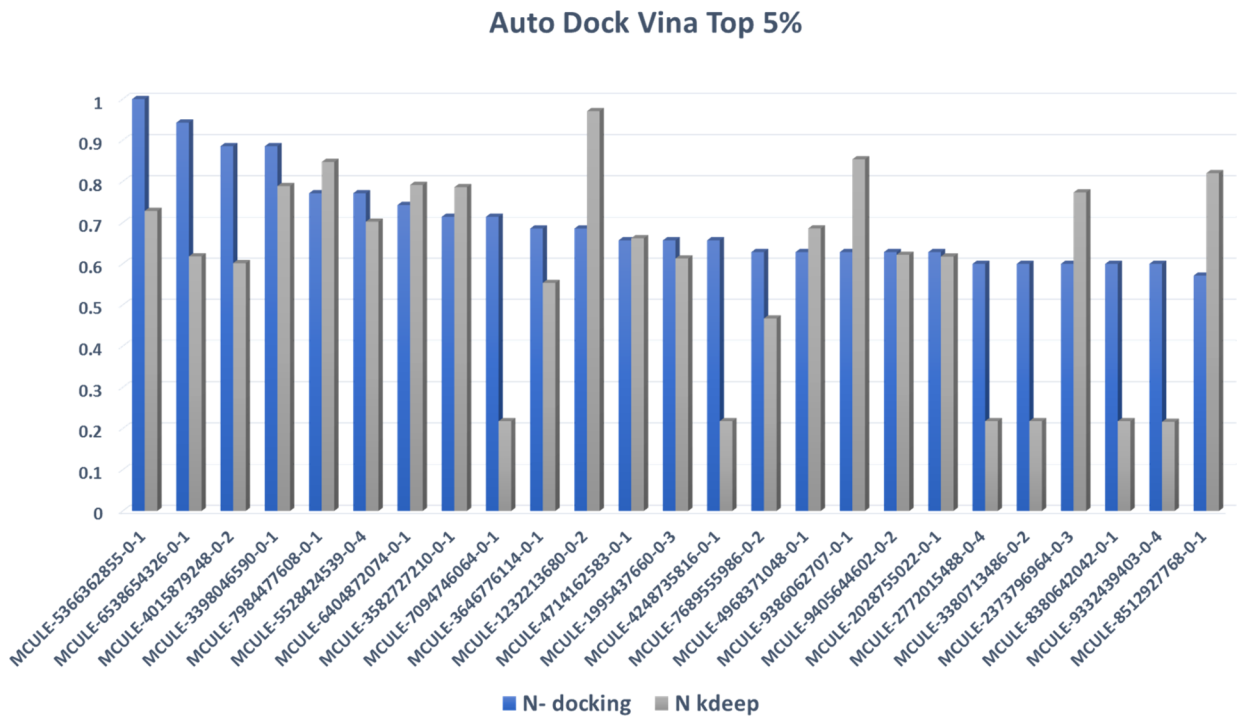


Figure 3. Top 5% normalized scored compounds from AutoDock Vina (molecular docking).

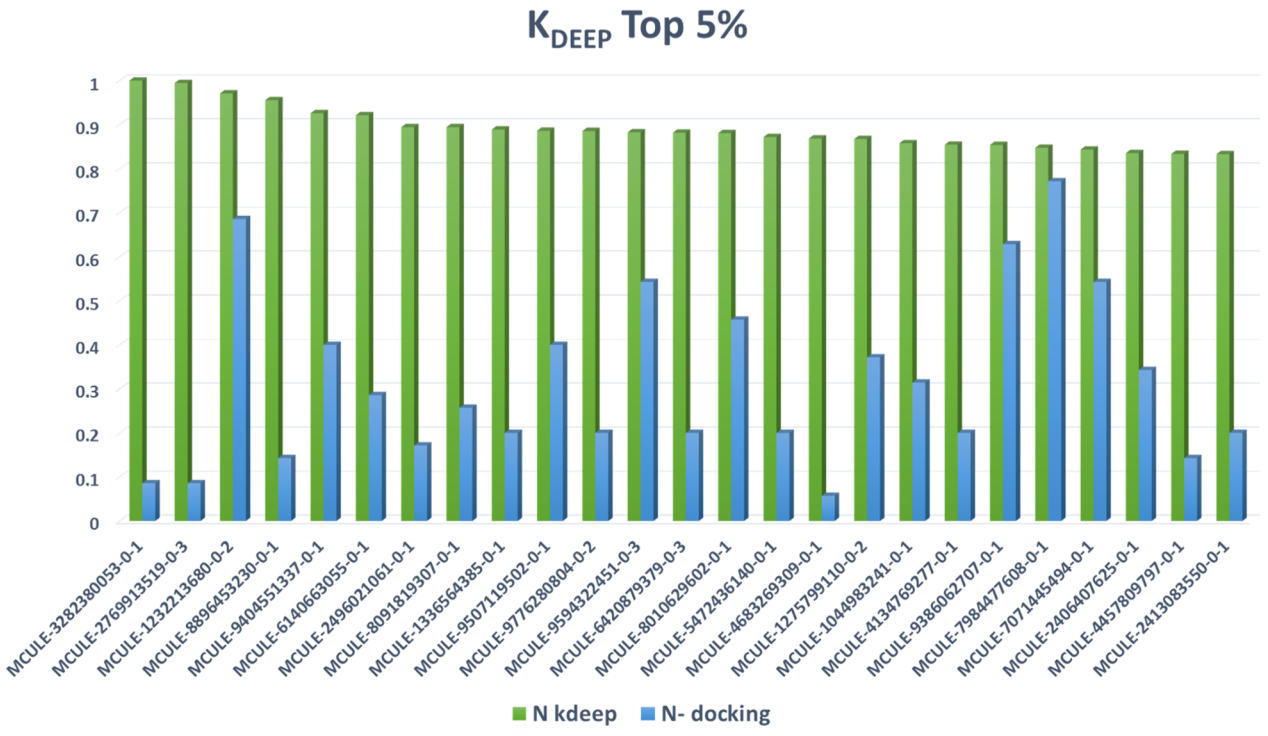


Figure 4. Top 5% normalized scored compounds from K_{DEEP} (machine learning).

and TRP347; sulfur interaction with MET280; pi-anion interaction with ASP308; and pi-pi interactions with HIS221 and HIS309. Compound B also had the favorable interactions. However, unlike compound A, it had one unfavorable interaction with ALA310, which is due to the donor-donor interaction by the two hydrogen atoms (Figure 7). The favorable interactions include hydrogen bonding with ARG307, GLN212, and HIS309; sulfur interaction with HIS291; pi-pi interaction with TRP347; and alkyl interaction with MET280 and TYR282. Like compound B, compound C also has an unfavorable interaction due to the donor−donor interaction between an oxygen atom and nitrogen atom from compound C and ARG214, respectively (Figure 9). However, its favorable interactions include van der Waals interaction with LEU213; hydrogen bonding interactions with ASP308, ARG307, and TRP347; pi-sigma interaction with LEU224; pi-amide interaction with GLN212; and alkyl interactions with MET211 and HIS309. The high number of favorable

https://dx.doi.org/10.1021/acsomega.0c06078
ACS Omega 2021, 6, 10607−10617

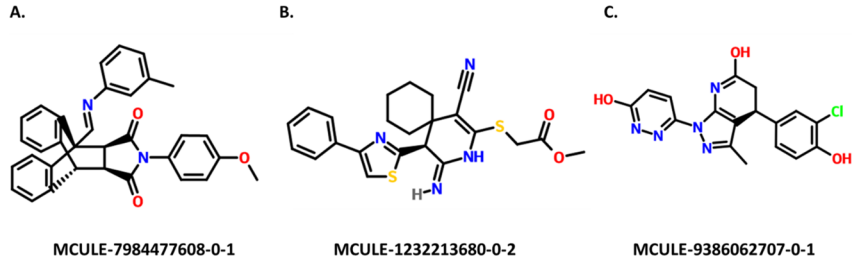


Figure 5. Selected compounds A, B, and C from consensus scoring of two different virtual screening methods.

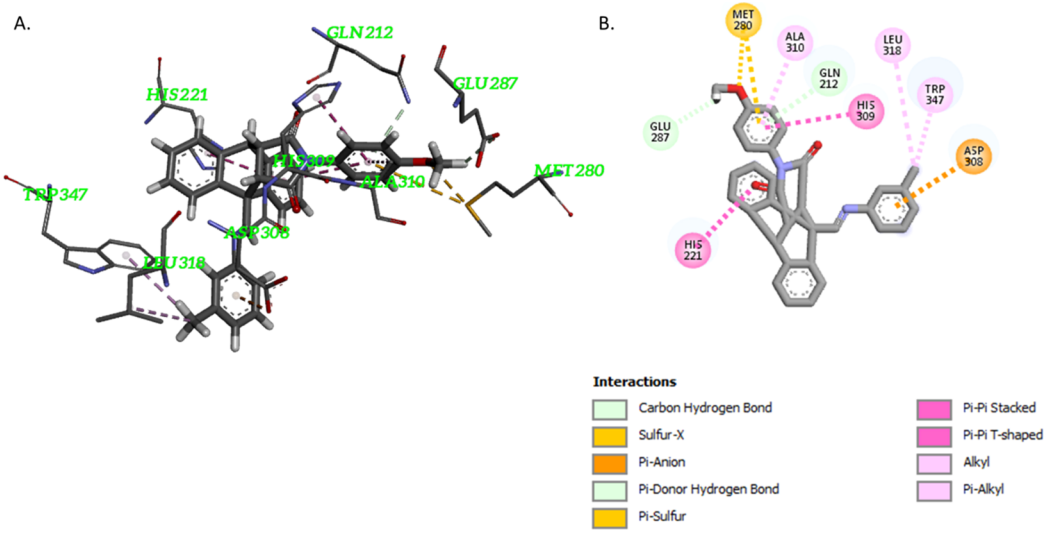


Figure 6. (a) Binding mode and (b) molecular interactions of compound A.

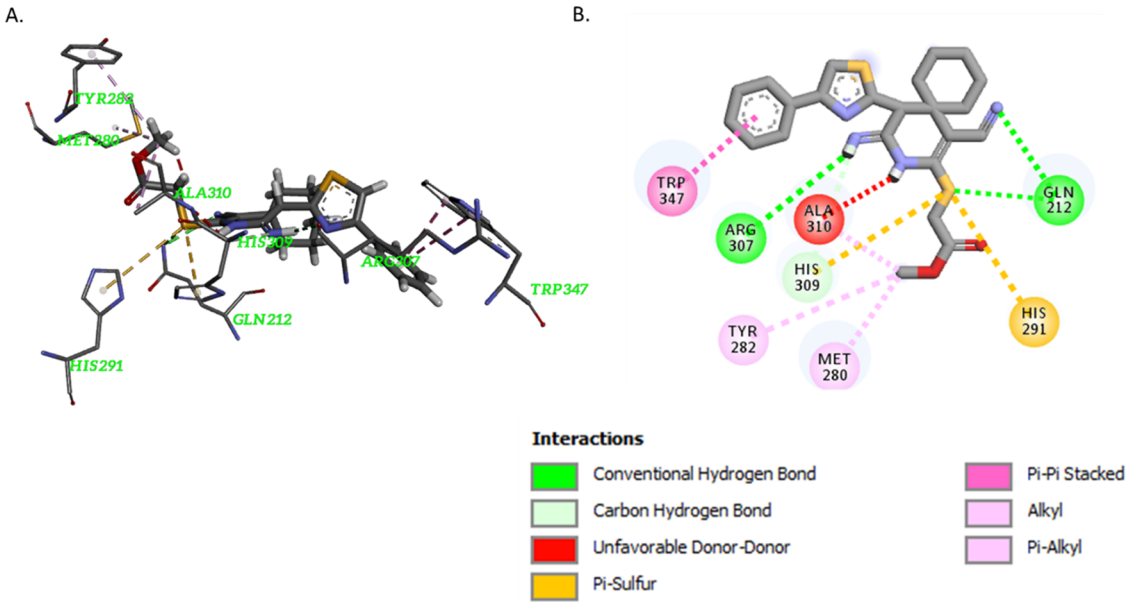


Figure 7. (a) Binding mode and (b) molecular interactions of compound B.

interactions and the absence of unfavorable interaction in compound A is possibly responsible for its highest binding energy (−8.7 kcal/mol) compared to the other selected hits. The presence of GLN212, HIS309, and TRP347 in all the molecular interactions of all three compounds with the protein confirms the same binding pocket. To further validate the homology model, the active sites of both the homology model (*D. postgatei*) and the template protein (*A. vinosum*) were compared with the aid of an active site identification tool, AADS.[43] Though it is important to note that the two proteins are of different species, and they do not have the same number of residues; that is, the homology model contains 423 residues, while the template protein has 396 residues. Nevertheless, there are similarities. All the amino residues present in all the binding pockets of the amino residues were also found in the active site identification tool (GLN212, HIS309, and

https://dx.doi.org/10.1021/acsomega.0c06078
ACS Omega 2021, 6, 10607−10617

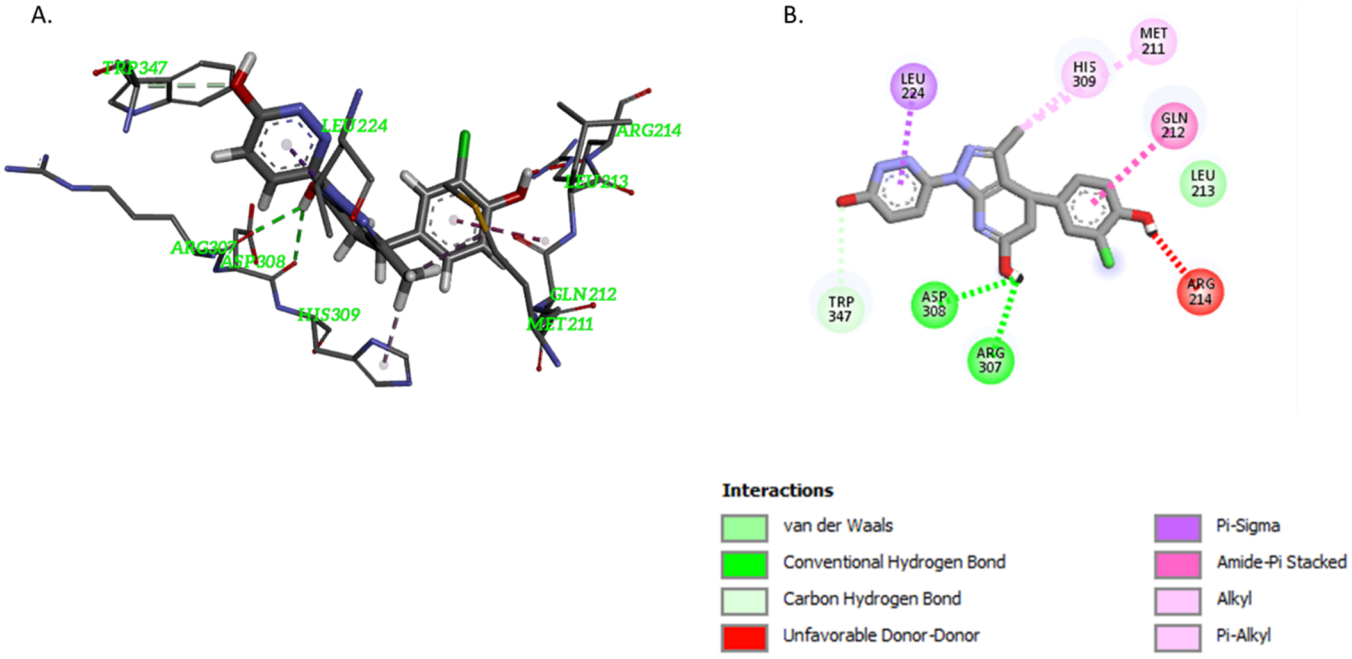


Figure 8. (a) Binding mode and (b) molecular interactions of compound C.

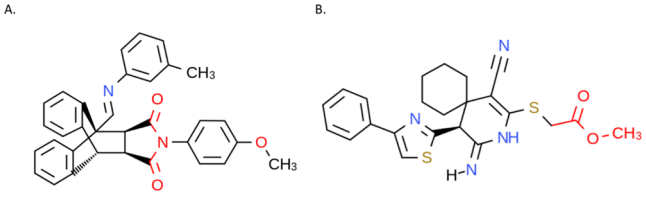


Figure 9. Toxicity alerts are in red font for compounds A and B.

TRP347). Hence, confirming that the binding pocket used was the true binding site. The amino residues ARG201 and ARG214; HIS208 and HIS218; and LEU233 and LEU244 were found at the same position of both the template protein and homology-modeled protein, respectively (Table S4), further validating the structure of the homology model (Figure 8).

2.4. Physicochemical and Toxicological Properties of Selected Compounds. The environmental safety of the use of any of the selected hits is quite important as a compound may be quite effective but due to its toxicological properties may be quite unsafe. In silico methods have been predominantly used in predicting the physicochemical and toxicological properties in the early stages of discovering potential inhibitors as it is much faster and safer and avoids unnecessary deaths of many animals that may be used for tests.[29,44−48] Moreover, their physicochemical properties (Table 3) that were determined using MCULE[49] are often used for gaining insight into their absorption, distribution, and properties. The renowned Lipinski's RO5 (molar mass and Log *P* values should not exceed 500 g/mol and 5, respectively) means that the number of hydrogen bond acceptors (HBA) and hydrogen bond donors (HBD) must not be greater than 10 and 5, respectively; hence, the name RO5 is due to the multiples of five being the maximum limit for the rules.[50] Only compound C did not violate any of the rules; however, both compounds A and B had Log *P* values greater than 5. The refractivity, which contributes to the absorption and distribution properties of the compound, is expected to be

Table 3. Physicochemical Properties of the Selected Hits

property	A	B	C
mass	498.57	452.59	371.78
Log *P*	6.02	5.49	2.59
HBA	5	6	8
HBD	0	2	3
PSA	58.97	152.4	116.65
RO5 violations	1	1	0
refractivity	151.04	128.98	99.43
atoms	64	55	40
rings	7	4	4
heavy atoms	38	31	26
hydrogen atoms	26	24	14
chiral centers	4	1	1
solubility	−4.01	−3.67	−3.18

between 40 and 130[51] of which only compound A exceeded a value of 151.04. Another property, which contributes to the absorption and distribution of a molecule, is the polar surface area (PSA) that is expected to be less than 140 Å^2[52] of which compound B exceeded with a value of 152.4. The solubility of the selected hits was predicted using admetsar.[53] For a compound to be soluble, it is expected to be between −1 and −5[54] of which all the selected hits qualify. Hence, it can be concluded that only compound C has the most suitable physicochemical properties required among the selected hits.

The toxicity properties (Table 4) were predicted using pkCSM webtool,[55] the toxicity checker,[49] and a DL-AOT (deep learning-acute oral toxicity) prediction server.[56] The AMES toxicity refers to the mutagenicity of the molecule, while the maximum recommended tolerated dose (MRTD) signifies an estimation of the toxic dose threshold in humans. It is recommended to be less than or equal to 0.477 log mg/kg/day. Fortunately, all the selected hits are safe with regard to mutagenicity and MRTD. The skin sensitization refers to the hazard involved if the compound is applied dermally. The results show that all the selected hits are safe. The LD_{50} (lethal

https://dx.doi.org/10.1021/acsomega.0c06078
ACS Omega 2021, 6, 10607−10617

Table 4. Toxicological Properties of the Selected Hits

properties	A	B	C	units
AMES toxicity	No	No	No	categorical (Yes/No)
max. tolerated dose (human)	0.302	−0.573	−0.151	numeric (log mg/kg/day)
skin sensitization	No	No	No	categorical (Yes/No)
oral rat acute toxicity (LD_{50})	2.77	2.97	3.16	log(mg/kg)
minnow toxicity	−2.262	−0.987	2.014	numeric (log mM)

dose) value for the oral rat acute toxicity considers the toxic potency of a compound. It is the amount of a compound that if given at once would cause the death of 50% of a group of test animals. The DL-AOT webtool predicts the LD_{50} values in four categories, namely, (1) danger/poison, (2) warning, (3) caution, and (4) none required/safe. The values for compounds A, B, and C correspond to the warning, caution, and safe, respectively. The minnow toxicity refers to the LC_{50} (lethal concentration) value, which is the concentration of a compound required to cause the death of 50% of a group of flathead minnows. A value below 0.5 log mM is regarded as high acute toxicity. Only compound C was predicted to be safe with a value of 2.014. The toxicity checker[49] tool was used to deduce the toxicity alerts in the selected hits. However, it depicted that the N—C═O group in the imidazole ring and the methoxy group are the toxic alerts for compound A (Figure 9), while the ester group is the toxic alert for compound B. Consequently, compound A could be modified using the toxic alerts as a guide. The oxygen in the methoxy group could be replaced with a nitrogen atom making it a tertiary amine. Also, the two carbonyl groups attached to the imidazole ring should be replaced with any halogen atom. Nevertheless, it showed no alerts for compound C confirming it as a safe compound.

3. CONCLUSIONS

The use of virtual screening and applying a consensus scoring method, which combines both molecular docking and machine learning, helped in selecting three compounds as hits from a database of about 40 million compounds. The molecular docking results showed that all the compounds have negative binding energies with compound A having the highest docking score. The molecular interactions revealed that the high binding affinity observed in compound A is most likely due to the high number of favorable interactions and the absence of unfavorable interactions, which occurred in compounds B and C. However, based on the physicochemical and toxicological properties, compound C is the best choice, as it does not violate any of the Lipinski's RO5 and other recommended properties that relate to absorption and distribution properties. Moreover, compound C was predicted to be the only safe compound with respect to LD_{50} and LC_{50} values for both the acute toxicity for both rats and flathead minnows. Nevertheless, compound C can be modified to increase its binding affinity by replacing the hydroxyl group, which is a hydrogen bond donor and is responsible for the unfavorable interaction, with a carbonyl group that is a hydrogen bond acceptor and hence would form a favorable interaction with ARG214. Nevertheless, whatever modification(s) are made to compound C it is important to take into consideration the effect it would have on its physicochemical and toxicological properties. Summarily, when modifying selected hit compounds, there is often a trade-off between increasing binding affinity and acquiring good physicochemical and toxicological properties. It is recommended that compound C should be validated using molecular dynamics and in vitro experiments to confirm its inhibitory activity against *Desulfobacter postgatei* (a prominent sulfate-reducing bacteria). A positive result would reduce the level of sulfur produced in oil and gas wells, which will subsequently prevent formation of iron sulfide scales such as pyrite. These findings will help in directing future experimental research on this subject. However, an important point to note is the stability and compatibility of these potential inhibitors with other existing chemicals, which are already in use for field applications in the industry. This could pose as a challenge if the reactivity of these compounds may lead to production of harmful substances when added with other chemicals such as chelating agents used for field applications. A possible way of solving such a challenge if it occurs is to further modify this compound based on the structural activity relationships such that it is able to carry out both activities of being an inhibitor of SRB and a scale remover. The lesser the number of chemicals are used in the field, the better, provided that such chemicals are multipurpose in providing a safer and more economical solution.

4. METHODS

4.1. Target Protein Preparation. Selecting a protein target is the first and most crucial step in structure-based virtual screening. Unfortunately, as there was no structure of ATP sulfurylase in the protein databank (PDB), hence the structure was designed using homology modeling with the aid of the Raptor program.[40,41] The sequence of the target protein was taken from the National Center for Biotechnology Information (NCBI) GenBank database (https://www.ncbi.nlm.nih.gov/protein) with the accession number: EIM64703.[57,58] As homology models are dependent on the quality of an existing crystal structure,[59] the model was based on the template of the crystal structure of ATP sulfurylase of *Allochromatium vinosum* (PDB ID: 4DNX)[60,61] with a resolution of 1.6 Å, which implies that the resolution is high as electron density maps derived from the X-ray crystallographic data of a protein structure with a resolution of at most 2.0 Å are deemed high-resolution structures.[62]

4.2. Virtual Screening. A blind docking process was employed for the virtual screening as the structure is based on a homology model. This process involves considering the whole of the protein during the molecular docking process to find potential binding sites and is used when a binding site is unknown.[63−65] The PyRx software[66] was used to determine the binding parameters using the maximize option, which enclosed the entire protein molecule, since the binding site is not known and blind docking is being employed. This resulted in a binding site center of 2.0027, −1.9107, and 24.2677 for the *X*, *Y*, and *Z* axes, respectively. The size of the box for the docking was 100 × 100 × 100 Å.[3] These parameters ensured that the whole structure was covered during the virtual screening process. The homology-modeled structure was virtually screened against 100,000 compounds randomly chosen from the MCULE database,[49] which consists of precisely 39,884,964 compounds at the time of this work. To select the 100,000 compounds randomly, the basic properties option was used. This option filters chemical compounds from the database on the basis of the following conditions: The filtered compounds should not violate more than one of the Lipinski's rule of 5 (RO5);[50,67,68] the compounds should not

https://dx.doi.org/10.1021/acsomega.0c06078
ACS Omega 2021, 6, 10607−10617

exceed a maximum of 10 rotatable bonds, a maximum of five halogen atoms, a maximum of five chiral centers; and a minimum of 10 heavy atoms and a minimum of one aromatic ring.[69] The criteria used for filtering 100,000 compounds from the MCULE database are based on common properties found in drug-like compounds. This criteria of filtering ensured that the selected compounds had the desirable physicochemical properties, which are important during absorption and toxicity studies and have been used in our previous works on finding potential inhibitors against the Zika and Ebola virus diseases.[70,71] The filtration is brought to a halt after finding the first 100,000 compounds from the MCULE database that met this criteria. AutoDock Vina was used for the molecular docking of the filtered compounds with the target protein.[72] Only the top-scored 500 compounds were kept from the virtual screening.

4.3. Consensus Scoring. Consensus scoring is a process of combining of two or more scoring algorithms to rank the binding affinities of ligands to a target protein. Consensus scoring has been known to reduce the rate of false positives[73−76] while also serving as a validation tool for the binding affinities. Hence, the use of consensus scoring has led to improved hit rates. Machine learning has proved to be a useful tool in calculating binding affinities in a much shorter time than molecular docking.[77,78] The K_{DEEP} machine-learning tool, which uses convolutional neural networks (CNN), was used for a second virtual screening. The K_{DEEP} gives the binding affinity result in pK_d such that the higher the pK_d value, the stronger the binding affinity. Lately, artificial intelligence techniques, such as machine and/or deep learning, are gaining considerable attention in the drug discovery process. This is partly due to it being much faster than molecular docking. In addition, having a different method rather than another molecular docking program with a different scoring method helps to reduce the number of false positives.[62,79] The datasets used in building K_{DEEP} were from PDBbind (v.2016) database[80] that contained 13,308 protein−ligand complexes, and their corresponding experimentally determined binding affinities were collected from literature and the Protein Data Bank (www.rscb.org).[81] The dataset was split into training and test datasets. The CNN method was applied using the 3D descriptors described in Jiménez et al.'s[82] study and was used to create the model. Keras and Tensorflow were used to implement the model.[83] The model was compared with other scoring functions, such as cyscore[84] and RF-score,[85] and was found to perform better with respect to average correlation. K_{DEEP} also served as an updated benchmark for even unseen data for predicting binding affinity for protein−ligand affinity.[86] The results of both virtual screenings (molecular docking and K_{DEEP}) were normalized such that values close to 1 and 0 corresponded to the top- and low-scored values, respectively. A vote rank method was used in the consensus scoring process to select the hits that mutually appeared in the top 5% scored ligands from the two virtual screenings.[52,69] That is, the top-scored 25 (5% of 500) compounds in molecular docking (AutoDock Vina) were compared to the top-scored 25 compounds in K_{DEEP}. The compounds that appeared on both lists were the selected compounds from the consensus scoring. This method is known as the rank by voting method.[73] The workflow of the methodology employed is depicted in Figure 10.

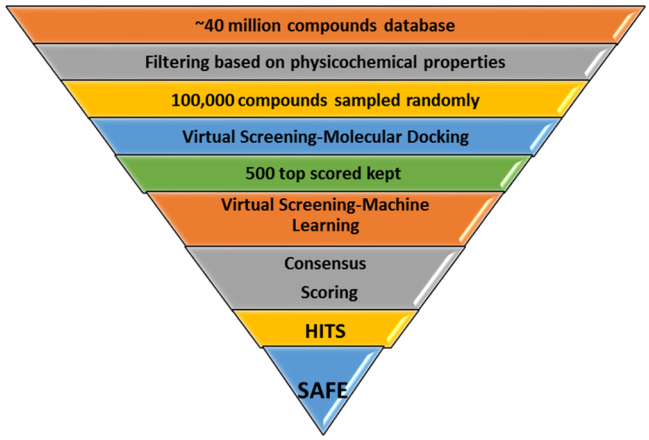


Figure 10. Methodology workflow of hit selection.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.0c06078.

(Table S1) Sequence alignments and RMSD of model and template proteins; (Table S2) AutoDock Vina molecular docking results from first virtual screening; (Table S3) K_{DEEP} machine-learning results from second virtual screening; (Table S4) the amino residues in the active site of *D. postgatei* (homology model) and *A. vinsoum* (template protein) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Ibnelwaleed A. Hussein − *Gas Processing Center, College of Engineering, Qatar University, Doha 2713, Qatar;* orcid.org/0000-0002-6672-8649; Email: ihussein@qu.edu.qa.

Hassan Nimir − *Chemistry Department, College of Arts and Sciences, Qatar University, Doha 2713, Qatar;* Email: hnimir@qu.edu.qa

Authors

Abdulmujeeb T. Onawole − *Gas Processing Center, College of Engineering, Qatar University, Doha 2713, Qatar;* orcid.org/0000-0002-4967-7245

Mohammed A. Saad − *Gas Processing Center, College of Engineering, Qatar University, Doha 2713, Qatar; Chemical Engineering Department, College of Engineering, Qatar University, Doha 2713, Qatar*

Musa E.M. Ahmed − *Gas Processing Center, College of Engineering, Qatar University, Doha 2713, Qatar*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c06078

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by funding from NPRP Grant # 9-084-2-041 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors. Qatar University and the Gas

Processing Center are acknowledged for their support. Open Access funding provided by the Qatar National Library.

LIST OF ABBREVIATIONS USED

SRB- sulfate-reducing bacteria; ATP- adenosine triphosphate; APS- adenosyl phosphosulfate

REFERENCES

(1) Heider, J.; Spormann, A. M.; Beller, H. R.; Widdel, F. Anaerobic Bacterial Metabolism of Hydrocarbons. *FEMS Microbiol. Rev.* **1998**, *22*, 459−473.
(2) Kniemeyer, O.; Musat, F.; Sievert, S. M.; Knittel, K.; Wilkes, H.; Blumenberg, M.; Michaelis, W.; Classen, A.; Bolm, C.; Joye, S. B.; et al. Anaerobic Oxidation of Short-Chain Hydrocarbons by Marine Sulphate-Reducing Bacteria. *Nature* **2007**, *449*, 898−901.
(3) Cord-Ruwisch, R.; Kleinitz, W.; Widdel, F. Sulfate Reducing Bacteria and Their Activities in Oil Production. *J. Pet. Technol.* **1987**, 97−106.
(4) Song, X.; Yang, Y.; Yu, D.; Lan, G.; Wang, Z.; Mou, X. Studies on the Impact of Fluid Flow on the Microbial Corrosion Behavior of Product Oil Pipelines. *J. Pet. Sci. Eng.* **2016**, *146*, 803−812.
(5) Liu, T.; Cheng, Y. F.; Sharma, M.; Voordouw, G. Effect of Fluid Flow on Biofilm Formation and Microbiologically Influenced Corrosion of Pipelines in Oilfield Produced Water. *J. Pet. Sci. Eng.* **2017**, *156*, 451−459.
(6) Sulaiman, K. O.; Onawole, A. T. Quantum Chemical Evaluation of the Corrosion Inhibition of Novel Aromatic Hydrazide Derivatives on Mild Steel in Hydrochloric Acid. *Comput. Theor. Chem.* **2016**, *1093*, 73−80.
(7) Hamilton, W. A. Sulphate-Reducing Bacteria and the Offshore Oil Industry. *Trends Biotechnol.* **1983**, *1*, 36−40.
(8) Enning, D.; Garrelfs, J. Corrosion of Iron by Sulfate-Reducing Bacteria: New Views of an Old Problem. *Appl. Environ. Microbiol.* **2014**, *80*, 1226−1236.
(9) Bertel, D.; Peck, J.; Quick, T. J.; Senko, J. M. Iron Transformations Induced by an Acid-Tolerant Desulfosporosinus Species. *Appl. Environ. Microbiol.* **2012**, *78*, 81−88.
(10) Kuang, F.; Wang, J.; Yan, L.; Zhang, D. Effects of Sulfate-Reducing Bacteria on the Corrosion Behavior of Carbon Steel. *Electrochim. Acta* **2007**, *52*, 6084−6088.
(11) Hao, O. J.; Chen, J. M.; Huang, L.; Buglass, R. L. Sulfate-Reducing Bacteria. *Crit. Rev. Environ. Sci. Technol.* **1996**, *26*, 155−187.
(12) Brandis-Heep, A.; Gebhardt, N. A.; Thauer, R. K.; Widdel, F.; Pfennig, N. Anaerobic Acetate Oxidation to CO2 by Desulfobacter Postgatei. *Arch. Microbiol.* **1983**, *136*, 222−229.
(13) Atkinson, T.; Sherwood, R. F.; Barton, L. L. *Biotechnology Handbooks 8 Sulfate-Reducing Bacteria*; Springer: Boston, MA, 1995, DOI: 10.1007/978-1-4899-1582-5.
(14) Singleton, R. *The Sulfate-Reducing Bacteria: Contemporary Perspectives*; Springer, New York, NY, 1993, DOI: 10.1007/978-1-4613-9263-7.
(15) Lavania, M.; Sarma, P. M.; Mandal, A. K.; Cheema, S.; Lai, B. Efficacy of Natural Biocide on Control of Microbial Induced Corrosion in Oil Pipelines Mediated by Desulfovibrio Vulgaris and Desulfovibrio Gigas. *J. Environ. Sci.* **2011**, *23*, 1394−1402.
(16) Hardy, J. A.; Hamilton, W. A. The Oxygen Tolerance of Sulfate-Reducing Bacteria Isolated from North Sea Waters. *Curr. Microbiol.* **1981**, *6*, 259−262.
(17) Mahmoud, M. A. N.; Badr Salem, B. g.; Hussein, I. A. *Method for Removing Iron Sulfide Scale*, US9783728B2, 2017.
(18) Mahmoud, M.; Hussein, I. A.; Sultan, A.; Saad, M. A.; Buijs, W.; Vlugt, T. J. H. Development of Efficient Formulation for the Removal of Iron Sulphide Scale in Sour Production Wells. *Can. J. Chem. Eng.* **2018**, *96*, 2526−2533.
(19) Buijs, W.; Hussein, I. A.; Mahmoud, M.; Onawole, A. T.; Saad, M. A.; Berdiyorov, G. R. Molecular Modeling Study toward Development of H2S-Free Removal of Iron Sulfide Scale from Oil and Gas Wells. *Ind. Eng. Chem. Res.* **2018**, *57*, 10095−10104.
(20) Sulaiman, K. O.; Onawole, A. T.; Shuaib, D. T.; Saleh, T. A. Quantum Chemical Approach for Chemiluminescence Characteristics of Di-Substituted Luminol Derivatives in Polar Solvents. *J. Mol. Liq.* **2019**, *279*, 146−153.
(21) Evenseth, L. M.; Warszycki, D.; Bojarski, A. J.; Gabrielsen, M.; Sylte, I. In Silico Methods for the Discovery of Orthosteric GABAB Receptor Compounds. *Molecules* **2019**, *24*, 935.
(22) Mohan, R. R.; Wilson, M.; Gorham, R. D., Jr.; Harrison, R. E. S.; Morikis, V. A.; Kieslich, C. A.; Orr, A. A.; Coley, A. V.; Tamamis, P.; Morikis, D. Virtual Screening of Chemical Compounds for Discovery of Complement C3 Ligands. *ACS Omega* **2018**, *3*, 6427−6438.
(23) Maia, E. H. B.; Medaglia, L. R.; Da Silva, A. M.; Taranto, A. G. Molecular Architect: A User-Friendly Workflow for Virtual Screening. *ACS Omega* **2020**, *5*, 6628−6640.
(24) Onawole, A. T.; Sulaiman, K. O.; Kolapo, T. U.; Akinde, F. O.; Adegoke, R. O. COVID-19: CADD to the Rescue. *Virus Res.* **2020**, *285*, 198022.
(25) Fang, X.; Bagui, S.; Bagui, S. Improving Virtual Screening Predictive Accuracy of Human Kallikrein 5 Inhibitors Using Machine Learning Models. *Comput. Biol. Chem.* **2017**, *69*, 110−119.
(26) Pereira, J. C.; Caffarena, E. R.; dos Santos, C. N. Boosting Docking-Based Virtual Screening with Deep Learning. *J. Chem. Inf. Model.* **2016**, *56*, 2495.
(27) Cang, Z.; Mu, L.; Wei, G.-W. Representability of Algebraic Topology for Biomolecules in Machine Learning Based Scoring and Virtual Screening. *PLoS Comput. Biol.* **2018**, *14*, No. e1005929.
(28) Smith, D. A. *Metabolism, Pharmacokinetics, and Toxicity of Functional Groups: Impact of the Building Blocks of Medicinal Chemistry in ADMET*; 2010.
(29) Moroy, G.; Martiny, V. Y.; Vayer, P.; Villoutreix, B. O.; Miteva, M. A. Toward in Silico Structure-Based ADMET Prediction in Drug Discovery. *Drug Discovery Today* **2012**, *17*, 44−55.
(30) Meyer, B.; Kuever, J. Homology Modeling of Dissimilatory APS Reductases (AprBA) of Sulfur-Oxidizing and Sulfate-Reducing Prokaryotes. *PLoS One* **2008**, *3*, e1514−e1517.
(31) Dos Santos, E. S.; De Souza, L. C. V.; De Assis, P. N.; Almeida, P. F.; Ramos-De-Souza, E. Novel Potential Inhibitors for Adenylylsulfate Reductase to Control Souring of Water in Oil Industries. *J. Biomol. Struct. Dyn.* **2014**, *32*, 1780−1792.
(32) Krieger, E.; Nabuurs, S. B.; Vriend, G. Homology Modeling. *Struct. Bioinforma.* **2003**, *857*, 509−523.
(33) Hazai, E.; Bikádi, Z. Homology Modeling of Breast Cancer Resistance Protein (ABCG2). *J. Struct. Biol.* **2008**, *162*, 63−74.
(34) Li, W.; Cowley, A.; Uludag, M.; Gur, T.; McWilliam, H.; Squizzato, S.; Park, Y. M.; Buso, N.; Lopez, R. The EMBL-EBI Bioinformatics Web and Programmatic Tools Framework. *Nucleic Acids Res.* **2015**, *43*, W580−W584.
(35) Ramachandran, G. N.; Sasisekharan, V. Conformation of Polypeptides and Proteins. *Adv. Protein Chem.* **1968**, *23*, 283−437.
(36) Kleywegt, G. J.; Jones, T. A. Phi/Psi-Chology: Ramachandran Revisited. *Structure* **1996**, *4*, 1395−1400.
(37) Hooft, R. W. W.; Sander, C.; Vriend, G. Objectively Judging the Quality of a Protein Structure from a Ramachandran Plot. *Comput. Appl. Biosci.* **1997**, *13*, 425−430.
(38) Lovell, S. C.; Davis, I. W.; Arendall, W. B., III; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure Validation by Cα Geometry: ϕ,ψ and Cβ Deviation. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 437−450.
(39) Kufareva, I.; Abagyan, R. Methods of Protein Structure Comparison. *Methods Mol. Biol.* **2012**, *857*, 231−257.
(40) Peng, J.; Xu, J. Raptorx: Exploiting Structure Information for Protein Alignment by Statistical Inference. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 161−171.
(41) Källberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J. Template-Based Protein Structure Modeling Using the RaptorX Web Server. *Nat. Protoc.* **2012**, *7*, 1511−1522.
(42) Ma, J.; Wang, S.; Zhao, F.; Xu, J. Protein Threading Using Context-Specific Alignment Potential. *Bioinformatics* **2013**, *29*, i257.

https://dx.doi.org/10.1021/acsomega.0c06078
ACS Omega 2021, 6, 10607−10617

(43) Singh, T.; Biswas, D.; Jayaram, B. AADS - An Automated Active Site Identification, Docking, and Scoring Protocol for Protein Targets Based on Physicochemical Descriptors. *J. Chem. Inf. Model.* **2011**, *51*, 2515−2527.
(44) Carmen, A. *Insilico Technologies in Drug Target Identification and Validation*; CRC Press: 2006.
(45) Ekins, S. Progress in Computational Toxicology. *J. Pharmacol. Toxicol. Methods* **2014**, *69*, 115−140.
(46) Hoeng, J.; Peitsch, M. C. *Computational Systems Toxicology*; Springer Science + Business Media: New York, 2015.
(47) Roncaglioni, A.; Toropov, A. A.; Toropova, A. P.; Benfenati, E. In Silico Methods to Predict Drug Toxicity. *Curr. Opin. Pharmacol.* **2013**, *13*, 802−806.
(48) Toropov, A. A.; Toropova, A. P.; Raska, I., Jr.; Leszczynska, D.; Leszczynski, J. Comprehension of Drug Toxicity: Software and Databases. *Comput. Biol. Med.* **2014**, *45*, 20−25.
(49) Kiss, R.; Sandor, M.; Szalai, F. A. A Public Web Service for Drug Discovery. *Aust. J. Chem.* **2012**, *4*, P17.
(50) Lipinski, C. A. Drug-like Properties and the Causes of Poor Solubility and Poor Permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235−249.
(51) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem.* **1999**, *1*, 55−68.
(52) Cerqueira, N. M. F. S. A.; Gesto, D.; Oliveira, E. F.; Santos-Martins, D.; Brás, N.; Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Receptor-Based Virtual Screening Protocol for Drug Discovery. *Arch. Biochem. Biophys.* **2015**, *582*, 56−67.
(53) Cheng, F.; Li, W.; Zhou, Y.; Shen, J.; Wu, Z.; Liu, G.; Lee, P. W.; Tang, Y. AdmetSAR: A Comprehensive Source and Free Tool for Assessment of Chemical ADMET Properties. *J. Chem. Inf. Model.* **2012**, *52*, 3099−3105.
(54) Tsaioun, K.; Kates, S. *ADMET for Medicinal Chemists: A Practical Guide*; Wiley: New Jersey, 2011.
(55) Pires, D. E. V.; Blundell, T. L.; Ascher, D. B. PkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *J. Med. Chem.* **2015**, *58*, 4066−4072.
(56) Xu, Y.; Pei, J.; Lai, L. Deep Learning Based Regression and Multiclass Models for Acute Oral Toxicity Prediction with Automatic Chemical Feature Extraction. *J. Chem. Inf. Model.* **2017**, *57*, 2672−2685.
(57) Nordberg, H.; Cantor, M.; Dusheyko, S.; Hua, S.; Poliakov, A.; Shabalov, I.; Smirnova, T.; Grigoriev, I. V.; Dubchak, I. The Genome Portal of the Department of Energy Joint Genome Institute: 2014 Updates. *Nucleic Acids Res.* **2014**, *42*, D26−D31.
(58) Lucas, S.; Han, J.; Lapidus, A.; Cheng, J.-F.; Goodwin, L.; Pitluck, S.; Peters, L.; Land, M. L.; Hauser, L.; Orellana, R.; et al. *Improved High-Quality Draft Sequence of Desulfobacter Postgatei*; 2012.
(59) Scior, T.; Paiz-Candia, B.; Islas, Á. A.; Sánchez-Solano, A.; Millan-Perez Peña, L.; Mancilla-Simbro, C.; Salinas-Stefanon, E. M. Predicting a Double Mutant in the Twilight Zone of Low Homology Modeling for the Skeletal Muscle Voltage-Gated Sodium Channel Subunit Beta-1 ($Na_v1.4$ $\beta1$). *Comput. Struct. Biotechnol. J.* **2015**, *13*, 229−240.
(60) Parey, K.; Demmer, U.; Warkentin, E.; Dahl, C.; Ermler, U. *The Crystal Structure of the ATP Sulfurylase from Allochromatium Vinosum in the Open State. TO BE Publ,* DOI: 10.2210/PDB4DNX/PDB.
(61) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(62) Stevens, E. *Medicinal Chemistry: The Modern Drug Discovery Process*; Pearson Education, Inc.: Boston, 2014.
(63) Hetényi, C.; van der Spoel, D. Blind Docking of Drug-Sized Compounds to Proteins with up to a Thousand Residues. *FEBS Lett.* **2006**, *580*, 1447−1450.
(64) Hetényi, C.; van der Spoel, D. Efficient Docking of Peptides to Proteins without Prior Knowledge of the Binding Site. *Protein Sci.* **2009**, *11*, 1729−1737.
(65) Grosdidier, A.; Zoete, V.; Michielin, O. Blind Docking of 260 Protein-Ligand Complexes with EADock 2.0. *J. Comput. Chem.* **2009**, *30*, 2021−2030.
(66) Dallakyan, S.; Olson, A. J. *Small-Molecule Library Screening by Docking with PyRx*; Humana Press: New York, NY, 2015, pp. 243−250, DOI: 10.1007/978-1-4939-2269-7_19.
(67) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **2001**, *23*, 3−25.
(68) Lipinski, C. A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug Discov. Today Technol.* **2004**, *1*, 337−341.
(69) Onawole, A. T.; Kolapo, T. U.; Sulaiman, K. O.; Adegoke, R. O. Structure Based Virtual Screening of the Ebola Virus Trimeric Glycoprotein Using Consensus Scoring. *Comput. Biol. Chem.* **2018**, *72*, 170−180.
(70) Sulaiman, K. O.; Kolapo, T. U.; Onawole, A. T.; Islam, A.; Adegoke, R. O.; Badmus, S. O. Molecular Dynamics and Combined Docking Studies for the Identification of Zaire Ebola Virus Inhibitors. *J. Biomol. Struct. Dyn.* **2019**, *37*, 3029−3040.
(71) Onawole, A. T.; Sulaiman, K. O.; Adegoke, R. O.; Kolapo, T. U. Identification of Potential Inhibitors against the Zika Virus Using Consensus Scoring. *J. Mol. Graphics Modell.* **2017**, *73*, 54−61.
(72) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2009**, *31*, 455−461.
(73) Yang, J.-M.; Hsu, D. F. Consensus Scoring Criteria in Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 1134−1146.
(74) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus Scoring for Ligand/Protein Interactions. *J. Mol. Graphics Modell.* **2002**, *20*, 281−295.
(75) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**, *42*, 5100−5109.
(76) Medina-Ruiz, D.; Erreguin-Luna, B.; Luna-Vázquez, F.; Romo-Mancillas, A.; Rojas-Molina, A.; Ibarra-Alvarado, C. Vasodilation Elicited by Isoxsuprine, Identified by High-Throughput Virtual Screening of Compound Libraries, Involves Activation of the NO/CGMP and H2S/KATP Pathways and Blockade of A1-Adrenoceptors and Calcium Channels. *Molecules* **2019**, *24*, 987.
(77) Lavecchia, A. Machine-Learning Approaches in Drug Discovery : Methods and Applications. *Drug Discovery Today* **2015**, *20*, 318−331.
(78) Wójcikowski, M.; Ballester, P. J.; Siedlecki, P. Performance of Machine-Learning Scoring Functions in Structure-Based Virtual Screening. *Sci. Rep.* **2017**, *7*, 46710.
(79) Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The Rise of Deep Learning in Drug Discovery. *Drug Discovery Today* **2018**, *23*, 1241−1250. PM-29366762 M4-Citavi
(80) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **2004**, *47*, 2977−2980.
(81) Burley, S. K.; Berman, H. M.; Christie, C.; Duarte, J. M.; Feng, Z.; Westbrook, J.; Young, J.; Zardecki, C. RCSB Protein Data Bank: Sustaining a Living Digital Data Resource That Enables Breakthroughs in Scientific Research and Biomedical Education. *Protein Sci.* **2018**, *27*, 316−330.
(82) Jiménez, J.; Doerr, S.; Martínez-Rosell, G.; Rose, A. S.; De Fabritiis, G. DeepSite: Protein-Binding Site Predictor Using 3D-Convolutional Neural Networks. *Bioinformatics* **2017**, *33*, 3036−3042.
(83) Team, T. T. D.; Al-Rfou, R.; Alain, G.; Almahairi, A.; Angermueller, C.; Bahdanau, D.; Ballas, N.; Bastien, F.; Bayer, J.;

Belikov, A. Theano: A Python Framework for Fast Computation of Mathematical Expressions. *arXiv Prepr. arXiv1605.02688* **2016**.
(84) Cao, Y.; Li, L. Improved Protein−Ligand Binding Affinity Prediction by Using a Curvature-Dependent Surface-Area Model. *Bioinformatics* **2014**, *30*, 1674−1680.
(85) Ballester, P. J.; Mitchell, J. B. O. A Machine Learning Approach to Predicting Protein-Ligand Binding Affinity with Applications to Molecular Docking. *Bioinformatics* **2010**, *26*, 1169−1175.
(86) Jiménez, J.; Škalič, M.; Martínez-Rosell, G.; de Fabritiis, G. K_{DEEP}: Protein-Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. *J. Chem. Inf. Model.* **2018**, *58*, 287−296.

https://dx.doi.org/10.1021/acsomega.0c06078
ACS Omega 2021, 6, 10607−10617