Targeted redesign of suramin analogs for novel antimicrobial lead development

Debayan Dey[1#], Suryanarayanarao Ramakumar[2], Graeme L. Conn[1,3#]

[1]Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA

[2]Department of Physics, Indian Institute of Science, Bangalore-560012, India

[3]Emory Antibiotic Resistance Center (ARC).

[#]**Corresponding author:** Debayan Dey, email: debayan.dey@emory.edu or Graeme L. Conn, email: gconn@emory.edu

Running title: Suramin analogs as antimicrobial leads

Keywords: Antiviral, antimicrobial, cheminformatics, virtual screening, precision docking, drug design strategies.

Author Contributions: DD, SR and GLC conceptualized the project. DD and GLC wrote the draft. All authors have given approval to the final version of the manuscript.

ABSTRACT

The emergence of new viral infections and drug resistant bacteria urgently necessitates expedient therapeutic development. Repurposing and redesign of existing drugs against different targets is one potential way in which to accelerate this process. Suramin was initially developed as a successful anti-parasitic drug, but has also shown promising antiviral and antibacterial activities. However, due to its high conformational flexibility and negative charge, suramin is considered quite promiscuous towards positively charged sites within nucleic acid binding proteins. Although some suramin analogs have been developed against specific targets, only limited structure activity relationship (SAR) studies were performed, and virtual screening has yet to be used to identify more specific inhibitor(s) based on its scaffold. Using available structures, we investigated suramin's target diversity, confirming that suramin preferentially binds to protein pockets which are both positively charged and enriched in aromatic or leucine residues. Further, suramin's high conformational flexibility allows adaptation to structurally diverse binding surfaces. From this platform, we developed a framework for structure- and docking-guided elaboration of suramin analog scaffolds using virtual screening of suramin and heparin analogs against a panel of diverse therapeutically relevant viral and bacterial protein targets. Use of this new framework to design potentially specific suramin analogs is exemplified using the SARS-CoV-2 RNA-dependent RNA polymerase (RdRp) and nucleocapsid protein, identifying leads that might inhibit a wide range of coronaviruses. The approach presented here establishes a new computational framework for designing suramin analogs against different bacterial and viral targets and repurposing existing drugs for more specific inhibitory activity.

22 **INTRODUCTION**

23 The rise of multi-drug resistant bacteria and emergence of new infectious viral pathogens, as
24 exemplified by SARS-CoV-2 the cause of the COVID-19 pandemic, are dual major threats to
25 human health world-wide. Effective new antimicrobials are thus urgently needed, but processes
26 for their identification and development are often long and expensive. However, repurposing of
27 existing, approved drugs and redesign of their activity against new targets represents one of the
28 most promising opportunities to by-pass many of these hurdles (1).

29 The unique challenges posed by COVID-19 have resulted in testing of a large array of
30 antivirals, anti-inflammatory drugs, and other potential inhibitors of receptors of viral attachment
31 for their effectiveness against SARS-CoV-2. For example, the nucleoside analog remdesivir,
32 originally developed for Ebola virus, was found to be effective against SARS-CoV-2 with high
33 selectivity, leading to its use for treatment of COVID-19 in many countries (2). Other antivirals
34 like favipiravir, ribavirin, darunavir, and lopinavir-ritonavir combination are either already in use
35 or in clinical trials for SARS-CoV-2 treatment (3). In addition, non-antivirals drugs like
36 dexamethasone, which is used commonly to treat inflammatory and autoimmune conditions (4),
37 and the anti-parasitic ivermectin have shown potential as COVID-19 treatments (5, 6).

38 Suramin was among the first anti-infective agents developed in the early 1920s and has
39 also recently been shown to inhibit the progression of SARS-CoV-2 infection in human airway
40 epithelial cell culture model (7). Suramin appears on the World Health Organization's List of
41 Essential Medicines and has a long history of clinical use, with application over the last century in
42 treating African Trypanosomiasis (sleeping sickness) caused by *Trypanosoma brucei* (8), and
43 other parasitic infections including onchocerciasis (African river blindness), leishmaniasis and
44 malaria (9-11). With its long history of successful use as an anti-parasitic, suramin and its analogs

45 have also found use in wide variety of other treatments, such as anti-cancer agents, antivirals, and
46 venom antidotes (12).

47 Suramin was found to exhibit broad antiviral properties through distinct activities against
48 components of diverse viruses. For example, through binding to glycoprotein gp120 suramin
49 inhibits attachment of HIV to host T-cells (13). Suramin also limits host cell entry of several other
50 viruses, including herpes, dengue and hepatitis C virus (14-16) and also inhibits enteroviral
51 attachment to human host cells by binding to their nucleocapsid protein (17). Suramin also inhibits
52 some processes in viral replication, for example through binding to Zika virus NS2B/NS3
53 proteinase (18) or to the RNA-dependent RNA polymerase (RdRp) of norovirus, chikungunya
54 virus and SARS-CoV-2 (19-21). Suramin and related compounds thus hold promise as an initial
55 leads for multiple targets, including viral RdRp and nucleocapsid, as well as future structure-
56 activity relationship (SAR) studies or structure-guided design of suramin analogs as new
57 therapeutics.

58 Although suramin and its analogs have been studied for many decades, including some
59 SAR studies (22, 23), many fundamental questions remain regarding the origin of their
60 promiscuity, binding pocket preferences, and mode of on-target interaction. Suramin's
61 effectiveness in targeting viral nucleocapsid and RdRp has not yet been exploited to rationally
62 design more selective and less toxic analogs based on this scaffold. Moreover, to our knowledge,
63 no systematic high-throughput or virtual screening has been performed for suramin analogs for
64 any viral targets (e.g. RdRp or nucleocapsid) to explore potential lead molecules. Here, we
65 examine the structural basis of suramin's ability to bind to diverse proteins and use these insights
66 to develop a framework involving virtual screening, precision docking and docking-guided
67 elaboration of the suramin scaffold against different therapeutically important bacterial and viral

68 targets. Docking studies with a panel of nine diverse proteins reveals high-scoring suramin analogs
69 to be more conformationally selective than suramin and to have greater target specificity. Finally,
70 SARS-CoV-2 RNA-dependent RNA polymerase (RdRp) and nucleocapsid are used to exemplify
71 the process of suramin analog redesign to generate novel analogs with the potential to bind with
72 higher affinity and target specificity. Evolutionary analyses suggest that the optimized leads could
73 potentially inhibit a wide range of coronaviruses (24). This study thus sets out a computational
74 framework using structure- and docking-guided ligand elaboration that can support redesign and
75 repurposing existing antimicrobials for new, specific inhibitory activities.
76

77 **RESULTS**

78 **Structural basis for suramin protein target site diversity**

79 Suramin is a C2 symmetric molecule with each half comprised of a trisulfonated naphthyl ring
80 with negative charge at physiological pH, connected via amide linkages to toluene and phenyl
81 rings (**Fig. 1A**). Suramin is highly flexible with a total of 10 rotatable bonds (**Fig. 1A**; $\alpha_{1\text{-}5}$ and $\alpha_{1'\text{-}}$
82 $_{5'}$) and no substituent groups on the four central rings that might otherwise restrict its
83 conformational freedom. This flexibility, and thus conformational adaptability, is likely a key
84 factor in suramin's ability to interact with diverse target sites in many proteins, but also its high
85 off-target binding. Similarly, another polyanionic class of molecules similar to suramin, the
86 glycosaminoglycans (GAGs) such as heparin (**Fig. 1B**), can bind multiple diverse protein targets
87 but also suffer the limitation of high off-target binding.

88 To begin defining the residue preferences in suramin binding pockets, we examined
89 structures of several suramin complexes with diverse proteins (**Table S1**). These structures
90 included the *Leishmania mexicana* pyruvate kinase (PDB 3PP7**; Fig. 1C**) with suramin bound in

91 its ATP binding site, snake venom protein myotoxin (PDB 1Y4L; **Fig. 1D**), and two human protein
92 targets for treatment of cancer, metabolic and neurological diseases, the NAD^+-dependent protein
93 deacetylase sirtuin 5 (SIRT5; PDB 2NYR; **Fig. 1E**) and the epigenetic reader chromobox homolog
94 7 (PDB 4X3U; **Fig. 1F**) (10, 25-27). A number of general features of the interaction networks
95 within the suramin binding pockets are immediately apparent. The trisulfo-naphthyl ring interacts
96 with positively charged residues (Arg, Lys, His) via its sulfate groups while the naphthyl ring
97 makes π-mediated interactions with aromatic residues, Phe, His, Tyr and Trp (**Fig. 1C-F**).
98 Aromatic residues with a polar ring atom may also make a hydrogen bonding interaction with a
99 sulfonate group, coordinating both features of the trisufonated naphthyl ring. The toluene and
100 central phenyl rings of suramin are additionally stabilized by networks of aromatic and
101 hydrophobic residues (**Fig. 1C-F**).

102 Suramin complexes with viral proteins have also been determined for norovirus RdRp
103 (PDB 3UR0) (19) and bunyavirus nucleocapsid (PDB 4J4V) (28), but these offer only partial
104 snapshots of favored suramin binding sites as just a portion of suramin could be modeled in each
105 crystal structure (**Fig. S1**). In the RNA binding channel of norovirus RdRp34, the trisulfo-naphthyl
106 ring of suramin is anchored by both positive (Lys171 and Arg392) and aromatic (Trp42) residues,
107 while the central region is stabilized by Gln66, Lys180, Lys181, Arg245 and Lys68. The
108 bunyavirus nucleocapsid forms a pentameric complex with an RNA binding cavity on the inner
109 edge of the ring-like assembly where suramin also interacts (28). Again, each of the modeled
110 suramin rings is stabilized by multiple ionic, hydrogen bonding and aromatic/ hydrophobic
111 stacking interactions, including residues Asn66, Lys67 and Arg95 with the trisulfo-naphthyl ring,
112 and Arg64, Thr63, Gly65, Phe177 and Pro127 the two central rings (**Fig. S1B**). Thus, visual
113 inspection of these viral protein structures reveals overall consistent trends in suramin's binding

114 pocket residue interaction preferences with other proteins of diverse origin and function.

115 More detailed analyses of available suramin-bound protein structures reveal consistent
116 features in the binding pocket physicochemical composition, including strong enrichment of
117 specific residues (**Fig. 1G**). Positively charged Lys and Arg are highly enriched in the pocket
118 compared to the full protein (25% and 15%, respectively, of all interacting residues), while His is
119 more modestly enriched (5%), but, as noted, can make both aromatic stacking and charged
120 interactions. The absence of negatively charged residues (~1.5% combined) also points towards
121 suramin's strong preference for positively charged binding regions, such as nucleic acid binding
122 sites. In total, polar residues are slightly less favored (~12 % total) as compared to aliphatic
123 hydrophobic residues (~14%), with a marked preference for Leu (10.6 %) among the latter group
124 (**Fig. 1G**). This may be due to leucine's aliphatic stacking potential with the aromatic rings of
125 suramin. Aromatic residues (Trp, Tyr and Phe; together ~25% of total) are also strongly enriched
126 in suramin binding sites.

127 Suramin's binding promiscuity for diverse protein targets is likely underpinned by its high
128 inherent conformational flexibility, as revealed by superposition of protein-bound suramin
129 molecules using one trisulfo-naphthyl ring (**Fig. 1H**). Suramin bound to thrombin (PDB 3BF6)
130 adopts the most extended conformation we identified (and is used hereafter as the reference), with
131 the trisulfo-naphthyl rings ~29 Å apart. Suramin bound to SIRT5 (PDB 2NYR) also has a linear
132 conformation but with differences in the α2, α3, α2', and α3' angles (i.e. the dihedrals between the
133 toluene and phenyl rings). In contrast, suramin bound to myotoxin I (PDB 6CE2) has its trisulfo-
134 naphthyl rings only ~14 Å apart and is observed in a highly bent conformation. Similarly, suramin
135 bound to myotoxin II (PDB 1Y4L) adopts a highly bent conformation, but distinguished by
136 additional changes in the α1 and α1' dihedral angles. To further assess the low energy

137 conformations of suramin and two analogs, NF449 and NF023, we used a computational *Monte*
138 *Carlo* conformational scan which revealed 116 potential low energy conformational states for
139 suramin, while NF449 and NF023 each have 143 and 80, respectively.

140 Collectively, these structural and computational conformation analyses confirm suramin to
141 be a highly flexible ligand, with multiple potential hinge points, that can adopt a large variety of
142 conformations to fit a given target protein binding pocket. Further, these favored binding pockets
143 are positively charged and enriched with aromatic and Leu residues, such as most commonly occur
144 in large interfaces or in protein regions which bind to nucleic acids or nucleotides.

145

146 **Identification of suramin and heparin analogs targeting viral and bacterial proteins**

147 Previous studies have identified inhibitory activity of suramin against bacterial and viral proteins,
148 but no computational or high-throughput screens have been conducted with suramin analogs
149 against such targets. Structure-guided analysis of suramin analogs bound to various viral or
150 bacterial targets using large-scale docking studies could, however, allow for downstream rational
151 design of novel analogs with improved affinity and specificity towards a desired target. We devised
152 a strategy to accomplish this goal using docking studies with suramin and heparin analogs and
153 using docking poses of the latter set of ligands to guide design of new suramin analogs with
154 substituents positioned to make additional protein target-specific interactions (**Fig. 2**).

155 Nucleic acid binding proteins in bacteria and viruses play crucial roles in the processing of
156 genetic information and thus represent important targets for therapeutic development. Due to
157 suramin's affinity for nucleic acid binding proteins, a test panel of diverse protein targets known
158 to interact with DNA or RNA was selected. Viral proteins, known to be bound by suramin,
159 included RdRp of Norovirus and SARS-CoV-2, which is involved in replication and transcription

160 of the viral RNA genome (19, 21); nucleocapsid of bunyavirus and SARS-CoV-2, which is
161 involved in packaging the viral RNA genome and virion assembly, and thus of critical importance
162 in viral infection (17, 28, 29); and, human immunodeficiency virus type 1 (HIV-1) protein gp120,
163 for which binding to galactosylceramide on human colon epithelial cells is blocked by suramin
164 (13, 30). To further diversify the protein target set, we included four bacterial proteins: DNA-
165 binding protein HU (HU), DNA recombination protein RecA (RecA), DNA-binding protein Fis
166 (Fis), and the thiostrepton-resistance methyltransferase (TsnR). HU and Fis are global regulators
167 which play important roles in bacterial gene regulation, biofilm development and maintaining
168 nucleoid architecture (31-34), while RecA is an essential DNA recombination and repair protein
169 known to be targeted by suramin (35). Finally, TsnR is a ribosomal RNA methyltransferase that
170 confers resistance to the antibiotic thiostrepton (36), and represents a wider group of RNA
171 modifying enzymes that are promising targets for novel drugs to break antibiotic resistance. To
172 generate the ligand sets, we next used a small panel of known biosimilars (**Table S2**) in a PubChem
173 search which identified ~2,000 and ~10,000 analogs of suramin and heparin, respectively (**Fig.**
174 **2A**).

175 Using the protein target panel and suramin/ heparin analog sets, high-throughput virtual
176 screening (HTVS; **Fig. 2B**) in the GLIDE module of the Schrödinger software produced a wide
177 range of binding scores, typically ranging from −1.0 to −9.0 kcal/mol, but with large differences
178 in distribution and average score for each protein target (**Fig. 3A**,**B**). Among the bacterial proteins,
179 suramin analogs bind on average more tightly to Fis and TsnR (mean binding score -5.7 and -5.9
180 kcal/mol, respectively) compared to HU and RecA (-5.1 and -4.6 kcal/mol, respectively; **Fig 3A**).
181 HIV protein gp120 was predicted to be bound less tightly with a mean score of -3.6 kcal/mol,

182 whereas both suramin and heparin analogs docked viral RdRp with comparatively higher score
183 (**Fig. 3A,B**).

184 Among the top 100 scoring suramin and heparin analogs identified by HTVS, more than
185 ~50% of these ligands bind to only a single protein among the nine targets (**Fig. 3C**,**D**).
186 Approximately 30% of the remaining ligands in each set bind to only two or three receptors, with
187 just a small fraction showing greater promiscuity (high docking score for 4 or more targets). This
188 result indicates that virtual screening-based using suramin or heparin analogs can identify ligands
189 with greater selectivity for diverse protein targets compared to using either suramin or heparin
190 individually.

191 The top 100 suramin and heparin analogs from HTVS against each target protein were next
192 subject to more accurate docking pose and binding score calculation using Standard Precision (SP)
193 GLIDE (**Fig. 2B** and **Fig. 3E**,**F**). An improvement in the docking score was observed for many
194 analogs due to the ligand conformational sampling used in the SP GLIDE docking method as
195 compared to "rigid" ligand docking of HTVS. However, SP GLIDE docking had a very similar
196 trend in docking score distribution for each set of top 100 analogs (**Fig. 3E**,**F**). Heparin analogs
197 bind modestly better than those of suramin, likely due to their greater overall conformational
198 flexibility and number of hydrogen bond donors/acceptors, which allow them more readily adapt
199 to the wide variety of positively charged surfaces in these target proteins. The top scoring ligands
200 from SP docking (**Tables S3** and **S4**, and Supplementary File "Top 100 docking results") were
201 then selected for subsequent elaboration of the suramin scaffold and are hereafter named beginning
202 with "Sur" and "Hep" for suramin and heparin analogs, respectively.

203

204 **Strategy for docking-guided elaboration of the suramin scaffold**

205 To facilitate suramin analog lead redesign for increased target selectivity and affinity, the top
206 scoring suramin analogs for each protein target were visually inspected post-docking and then
207 superposed with the top scoring heparin analog identified in the same binding site. The heparin
208 analog set contains ligands that are typically more conformationally flexible and enriched in
209 potential hydrogen bond donors/ acceptors. While these features would make the heparin analogs
210 themselves poor leads for specific interactors, comparison of the bound analogs can direct
211 identification of sites within the corresponding suramin analog that can be further substituted to
212 make additional interactions with the target protein (**Fig. 2C**). As an example, structural analysis
213 of the top suramin and heparin analogs directed against HU and the resulting suramin analog
214 redesign strategy is described below.

215 HU interacts with DNA as a dimer via a positively charged binding cleft, lined with Lys
216 and Arg residues from both protomers. Sur-1 binds to HU in this cleft with the highest affinity of
217 the suramin analog set (-9.66 kcal/mol), placing two of its naphthyl rings over a surface formed by
218 Pro81 and the adjacent β-strand of both HU protomers at the center of the dimer interface (**Fig.**
219 **4A**). Sur-1 is composed of three sulfo-naphthyl rings (**Fig. 4B**), with both terminal rings substituted
220 with two sulfonyl groups, and a single sulfonate on the central ring. The majority of these groups
221 are coordinated by multiple Arg residues, including Arg53, Arg55 and Arg80 from both protomers,
222 and Arg58 and Lys 86 from a single protomer (**Fig. 4A**). Lys86 of the second protomer is also
223 positioned near (~4-6 Å) the carbonyl and sulfonyl groups of the central and terminal naphthyl
224 rings, respectively, exemplifying how the dimeric interface may be exploited by a pseudo-
225 symmetrical suramin analog. Additionally, Sur-1 and other top HU-binding suramin analogs (Sur-
226 2 and Sur-3; **Table S3**) are all markedly more conformationally rigid than suramin, due to the
227 diazene linker between the naphthyl rings. Sur-1 has only four rotatable bonds with a single low

228 energy conformation (**Fig. 4B**), compared to 10 rotatable bonds and 116 low energy conformations
229 in suramin. The ligand docking process thus identified suramin analogs with both improved
230 binding affinity to HU (based on docking score) and reduced conformational rigidity and net
231 charge. These features make Sur-1 a superior candidate compared to suramin for developing a
232 more specific inhibitor of HU.

233 Our docking procedure with the heparin analog set also identified Hep-41 as having high
234 affinity (-10.02 kcal/mol) for the same pocket as Sur-1 in the DNA binding cleft of HU **(Fig. 4C**).
235 Hep-41 is composed of sulfated iduronic acid and glucosamine with trioxidanylsulfanyl
236 substitution, which make extensive interactions with essentially the same group of residues in the
237 DNA binding cleft (i.e. Arg53, Arg55 Arg58, Arg80 and Lys86; **Fig. 4B**). However, while docking
238 of Sur-1 and Hep-41 are qualitatively similar, due to its larger number of hydrogen bond donors/
239 acceptors, docking of the latter ligand offers detailed structure-guided cues for placement of new
240 substituents in Sur-1. Specifically, superposition of the two analogs reveals sites on Sur-1 for
241 optimal geometric placement of additional polar groups to enhance interaction within the binding
242 site (**Fig. 4A,D**). For example, addition of polar groups on each naphthyl ring could increase target-
243 ligand interactions by further engaging Arg53 and Arg55 (from both protomers) via bifurcated
244 hydrogen bonding. These dual ligand docking comparisons can thus be used as a guide to improve
245 the design of suramin analogs to increase both their target specificity and binding affinity.

246 This same strategy was used for the other eight target proteins, with heparin analog
247 superposition identifying multiple sites in each top suramin analog where substitutions could be
248 made to improve affinity and target selectivity (marked * in Fig. S2-S5; also see Supplemental
249 Results for further details). Specifically, we identified sites for targeted redesign of Sur-6 (using
250 Hep-42) for binding to the positively charged cleft of TsnR in an extended conformation (**Fig. S2**

251 **A-D**), and Sur-13 (using Hep-45) and Sur-15 (using Hep-50) for binding with higher affinity and
252 specificity to Fis (**Fig. S2 E-H**) and RecA (**Fig. S3 A-D**), respectively. Among the viral proteins,
253 HIV gp120 (6IEQ) binds to smaller analogs than the other targets with the highest docking score
254 to Sur-22 and sites for targeted redesign identified by docking of Hep-51 (**Fig. S3E-H**). Similarly,
255 for norovirus RdRp **(Fig. S4A-D**) and bunyavirus nucleocapsid **(Fig. S4E, F**) the docking
256 identified Sur-24/ Hep-56 and Sur-27/ Hep-52 binding poses, respectively, for targeted redesign.
257 Corresponding results for SARS-CoV-2 RdRp and nucleocapsid are described in the next section
258 along with subsequent redesign and further docking analyses. In summary, virtual screening and
259 precision docking analyses using suramin and heparin analog chemical space has discovered new
260 lead analogs and a path to rational redesign for specific viral and bacterial protein targets.
261

262 **Suramin analog redesign for specific, high-affinity lead inhibitors of SARS-CoV-2 RdRp and**
263 **nucleocapsid**

264 SARS-CoV-2 RdRp is a large macromolecular complex composed of nsp12 catalytic subunit, with
265 an RNA binding tunnel that presents a potential site for drug targeting (**Fig. S5A,B**), and nsp7-
266 nsp8 cofactors (37). Virtual screening and docking analyses identified Sur-36 as binding with
267 highest docking score to SARS-CoV-2 RdRp, with Hep-58 also binding in the same region (**Fig.**
268 **S5C**). Sur-36 has a central biphenyl ring attached to terminal sulfo-naphthyl groups joined by
269 aconformationally restricted diazene linker (**Fig. 5B**). Two other similarly high-scoring ligands,
270 Sur-37 and Sur-38, were also identified (**Table S3**), but Sur-36 was selected as the lead candidate
271 due to the greater flexibility of these additional hits. Inspection of the RdRp binding pocket reveals
272 that each ring of Sur-36is stabilized by multiple interactions: the distal sulfo-naphthyl ring by
273 Lys551, Ser549 and Arg553, the central phenyl rings by Ile548 and Ala549, and the proximal

274 sulfo-naphthyl ring with Tyr546 and Asn497 (**Fig. 5A, B**).

275 Using the Hep-58 binding pose, four positions for additional substitutions in Sur-36 were
276 identified as described above for the other ligand/ protein target pairs (**Fig. S5D**). A process of
277 chemical enumeration, structure-guided design and iterative docking was then used to
278 systematically alter the functional groups of substituents at the identified attachment positions and
279 calculate the changes in the docking score (**Table S5**). The methoxy groups in the central phenyl
280 rings were first modified to carboxylate, phenyl-methanol, halides and other functional groups,
281 with docking analysis identifying the carboxylate groups (Sur-60; **Fig. S6A**) as conferring the
282 greatest increase in docking score, from -10.14 (Sur-36) to -11.24 kcal/mol (Sur-60). In Sur-60,
283 one phenyl ring carboxyl group engages Lys545, while the second carboxylate makes new contacts
284 with Arg555. Additional docking-guided modifications were then made in Sur-60 by
285 systematically testing additional functional groups in the terminal naphthyl rings and the linker
286 between the central phenyl rings. This optimization process resulted in substitution of one terminal
287 naphthyl group with a phenalene group, addition of a hydroxyl group to the remaining sulfo-
288 naphthyl group, and a carbonyl linker between the central rings to generate Sur-84 (-11.60
289 kcal/mol; **Fig. S6A**). Sur-84 makes additional contacts to His439 and Arg836 via one of the
290 carboxylate groups and a hydrophobic interaction between the phenalene group and Val557.
291 Finally, conversion of a central phenyl ring to naphthyl added hydrophobic contact with Tyr546
292 and generated Sur-85 (-11.96 kcal/mol; **Fig. 5C** and **S6A**).

293 The same process was also applied to the SARS-CoV-2 nucleocapsid RNA binding domain
294 (29) which has a shallow positively charged surface (**Fig. S5E-G**). Smaller suramin analogs with
295 phenyl, stilbene and naphthyl moieties typically docked with higher scores than longer suramin
296 analogs, with Sur-31 among the highest docking scores (-7.53 kcal/mol) and selected for further

297 redesign to improve its affinity. Sur-31 is composed of two terminal naphthyl groups with a central
298 unsubstituted phenyl ring linked connected by two conformationally restricted amide linkers (**Fig.**
299 **S5H**). Inspection of the RNA binding pocket reveals Sur-31 is stabilized by multiple interactions:
300 one sulfo-naphthyl ring with Arg93, Arg108 and Arg150, the central phenyl rings with Tyr110
301 and Tyr112, and the second sulfo-naphthyl ring with Tyr112, Thr92 and Arg89 (**Fig. 5D, E**). Our
302 heparin screen also identified Hep-59 as binding to the same region, revealing potential for
303 extending the Sur-31 scaffold at several sites (indicated * in **Fig. S5H**). Again using docking-
304 guided elaboration, one naphthyl ring was replaced by two sulfonyl-substituted phenyl rings
305 connected to the central phenyl ring via a conformationally restricted azide linker resulting in Sur-
306 114 with an improved of docking score of -8.79 kcal/mol (**Fig. S6B**). Sur-114 makes additional
307 interactions with backbone carbonyls of Ala157 and Ile158, along with an increased van der Waals
308 interaction with the protein. Further guided substitutions in the second naphthyl ring, adding
309 another disulfonyl-hydroxy naphthyl ring linked by a flexible linker, generated Sur-121 with a
310 further improvement in docking score to -9.30 kcal/mol (**Fig. S6B**). Final optimization of this
311 naphthyl ring with a sulfate group resulted in Sur-127, with a final docking score of -9.34 kcal/mol.
312 Sur-127 engages the overall same set of residues but with increased H-bonding and electrostatic
313 interactions compared to Sur-31 (**Fig. 5F** and **S6B**).

314 This process thus identified an initial suramin analog for both SARS-CoV-2 protein targets,
315 RdRp (Sur-36) and nucleocapsid (Sur-31), for which predicted affinity could be substantially
316 improved by heparin docking-guided redesign efforts, resulting in final analogs Sur-85 and Sur-
317 127, respectively. Overall, heparin docking- and structure-guided redesign of suramin analogs
318 resulted in a collection of potential leads with markedly improved docking scores and other

319 properties, compared to analogs identified in precision docking (compare green and purple scores,
320 respectively in **Fig. 5G**).

321 Finally, we investigated residue conservation to gain insights into the evolutionary
322 constrains on the potential inhibitor binding sites in SARS-CoV-2 RdRp and nucleocapsid
323 proteins. Across all coronaviruses the identified binding sites appear moderately and very highly
324 conserved in nucleocapsid and RdRp, respectively (**Fig. 6**). As such, the phylogenetic relationship
325 reveals the high similarity of both SARS-CoV-2 proteins to homologs in other β-coronavirus
326 family members, comprising predominantly bat-origin viruses such as MERS-CoV and HCoV-
327 HKU1 (**Fig. 6A, B; Fig S7**). The conservation of the RdRp binding site also extends to α-
328 coronavirus family members of bat origin, which originated the human pathogens HCoV-229E
329 and HCoV-NL63 (**Fig. 6A, B**). More broadly, the generally strong conservation in each binding
330 site implies that the designed suramin analogs, Sur-85 and Sur-127, would potentially be equally
331 effective in targeting these proteins in different coronaviruses. Thus, together the ligand docking/
332 elaboration strategies and evolutionary analyses provide a basis for designing specific inhibitors
333 with broad anti-coronaviral activity.

334

335 **DISCUSSION**

336 Suramin has garnered wide clinical interest for treating parasitic infections, as well as diseases
337 ranging from viral infections (e.g. HIV, Dengue and Ebola) to human cancers (12). Despite not
338 satisfying typical drug likeliness criteria, such as Lipinski's "rule of five", suramin and its analogs
339 have found success in clinical application. Suramin is negatively charged and binds predominantly
340 positively charged surfaces in proteins involved in DNA and RNA processing, e.g., DNA and RNA
341 polymerases, telomerase and chromodomain proteins, histone methyltransferases, and sirtuin

342 histone deacetylases (19, 21, 23, 27). Suramin also inhibits several membrane channels and
343 signaling proteins. Despite suramin having been used for more than 100 years and shown to inhibit
344 diverse protein families, there still exists a significant gap in our understanding of its promiscuous
345 antagonist properties which likely also underpin its off-target and side effects in clinical use.

346 Prior to the current work, the nature of preferred binding surfaces and molecular
347 determinants of suramin promiscuity were unclear. We reasoned that understanding these features
348 could facilitate structure-based design of new suramin analogs with increased target specificity.
349 Collectively, our analyses of available structures with suramin or suramin analogs revealed the
350 contributions of both the physicochemical nature of the target protein binding pocket and
351 conformational properties of suramin in making it a promiscuous ligand for diverse protein targets.
352 In particular, suramin's high flexibility, conferred by ten rotatable bonds, largely unsubstituted
353 phenyl groups, and numerous non-directional electrostatic interactions with positively charged
354 residues, collectively contribute to an ability to bind to structurally diverse nucleic acid interaction
355 surfaces. Additionally, our analyses revealed an enrichment of aromatic residues in favored
356 suramin binding sites and the role of various types of π-mediated interactions.

357 To date, suramin has been used as a lead in only a limited number of SAR studies. In one
358 case, suramin-derived fragments were designed, synthesized and modified to act as an antagonist
359 of the G protein-coupled receptor P_2Y_2 (38), which mediates microglial inflammation. In another
360 study, the toluene ring in suramin was substituted to improve suramin's binding to falcipain-2 (39),
361 a cysteine protease of the malaria parasite *Plasmodium falciparum*. Although both these
362 experimental studies used suramin as a lead, to our knowledge no high-throughput virtual
363 screening has been performed nor a structure-guided approach taken to improve the affinity and
364 selectivity of suramin based inhibitors. Here, we devised a virtual screening and precision docking

365 strategy using sets of suramin and heparin analogs, the latter of which were then used to guide the
366 redesign of top suramin analogs for enhanced target affinity and specificity. This redesign strategy
367 was fully exemplified using SARS-CoV-2 nucleocapsid and RdRp proteins in a proof-of-concept
368 redesign of suramin analogs as specific inhibitors of these viral proteins with potentially broad
369 anti-coronaviral activity.

370 Suramin has been explored for its inhibitory activity against both bacterial and viral
371 pathogens. Here, we demonstrated the potential utility of our virtual screening and ligand docking
372 enabled identification and redesign of suramin analogs to improve their affinity and specificity for
373 both known (RecA and HU) and new antibacterial targets (Fis and TsnR). For example, our study
374 produced suramin analogs which bind more tightly to HU, based on computational docking scores,
375 than those previously identified by screening inhibitors of HU (33). Similarly, our computational
376 strategy generated docking models for suramin and its analogs that can explain their previously
377 identified capacity to block the function of diverse viral proteins including HIV-1 gp120, norovirus
378 RdRp and bunyavirus nucleocapsid.

379 SARS-CoV-2 has also been shown to be inhibited by suramin (7) and, during preparation
380 of this manuscript, cryo-EM structural analysis and biochemical evidence revealed the molecular
381 basis of suramin binding to SARS-CoV-2 RdRp (21). Our docking study is consistent with these
382 structural insights in which multiple binding poses of suramin were found in the RNA binding
383 channel, one blocking the RNA template strand and the other binding to the primer strand in the
384 catalytic site. Suramin and its analogs have been shown to be 20-fold more potent than remdesivir
385 binding to the same site, and our results suggest that significant further improvement in suramin
386 analog design is possible. Specifically, we further computationally identified suramin analogs
387 which could potentially be more specific to this target and our docking guided elaboration of

388 suramin analogs showed improvement in the drug scaffold to make it a tighter binder.
389 Additionally, our studies identified the potential to target SARS-CoV-2 nucleocapsid with suramin
390 analogs and proposed analogs possessing enhanced affinity and specificity. This approach could
391 also be applied to the 3CL protease which was recently identified as another possible suramin
392 binding target (40). Our phylogenetic and evolutionary analyses also suggest that the suramin
393 binding pockets are evolutionary conserved in related coronavirus RdRp and nucleocapsid proteins
394 which could make these suramin analogs more broadly applicable therapeutics against this group
395 of viral pathogens.

396 In summary, this study has defined the basis of suramin promiscuity and the nature of its
397 preferred protein target binding sites and established a computational pipeline (**Fig. 2**) to rapidly
398 identify new suramin analogs targeting diverse viral and bacterial targets for novel antimicrobial
399 development. Suramin analogs screened in this study (e.g. Sur-8, Sur-23, Sur-25 and Sur-33) are
400 built from modular aromatic fragments which can be decorated with different polar or charged
401 substitutions and connected by various linkers (e.g. amide, sulfonamide and diazene; **Fig S8**).
402 These properties can be readily exploited for future fragment-based designs of suramin-like
403 inhibitors of viral and bacterial nucleic acid binding proteins. As such, our strategy offers the
404 possibility for expedient design of novel suramin analogs with improved properties and efficacy
405 for synthesis and experimental confirmation of their potential as novel lead antimicrobials.

406

407 **MATERIALS AND METHODS**

408 **Virtual screening, docking and conformational analysis**

409 Suramin and heparin analog sets were created using chemoinformatic guided searches in PubChem
410 with parameters: Tanimoto similarity of >90%, <40 rotatable bonds, and molecular weight < 2 kD.

411 The structures of compounds retrieved were processed with the LigPrep module in Schrödinger
412 Software. For conformational analysis of the ligands, the ConfGen module (41) of Schrödinger
413 Software was used with the "mix torsional low mode sampling" method, which is a combination
414 of *Monte Carlo* and systematic sampling of rotatable bonds to find low energy ligand
415 conformations.

416 To prepare each protein ("receptor") for ligand docking, hydrogen atoms were added, and
417 the structure energy minimized using the Protein Preparation Wizard in Schrödinger Software. The
418 docking grid location was determined by analysis of the nucleic acid binding region of each
419 protein, either by prediction using SiteMap (Schrödinger Software) or from published
420 experimental evidence for a given target protein. Docking grid size was determined based on the
421 nucleic acid binding pocket volume and designed to completely cover all residues and centered
422 symmetrically. Docking was performed using the virtual screening workflow of GLIDE (42) in
423 Schrödinger Software, in HTVS mode and then for precision docking in SP mode as used in our
424 previous studies (33, 43).

425 Ligand physicochemical properties and chemical similarity were calculated in the Canvas
426 module of Schrödinger Software (44). Ligand-protein nonbonded interactions were calculated
427 using BIOVIA DS Viewer for salt bridges, hydrogen bonds and other π mediated interactions.
428 PyMOL was used for the visual inspection of structures and generation of figures.

429

430 **Phylogenetic analysis and conservation study**

431 Homologous sequences of coronavirus RdRp and nucleocapsid were collected by BLAST search
432 in NCBI with ~10,000 and 5,000 sequences identified, respectively. The large number of
433 sequences arises due to the number of sequenced SARS-CoV-2 and related viruses that have >99%

434 identity in the dataset. Therefore, the CD-HIT server (45) was next used to remove redundant
435 sequences from the dataset using a sequence identity cut-off of 98%. Multiple sequence alignment
436 and evolutionary analyses were then performed using Clustal Omega and MEGA 7 (46),
437 respectively, with evolutionary history inferred using the Maximum Likelihood method with
438 bootstrap consensus tree inferred from 100 replicates. The fraction of replicate trees in which the
439 associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the
440 branches. Evolutionary distances were computed using the JTT matrix-based method and are in
441 the units of number of amino acid substitutions per site. The rate variation among sites was
442 modeled with a gamma distribution (shape parameter 4). Finally, the phylogenetic tree was
443 visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Conserved residues were
444 plotted on the SARS-CoV-2 RdRp (PDB 7BV1, chain A) and nucleocapsid (PDB 6M3M, chain
445 A) using ConSurf (47) from the pre-calculated multiple sequence alignment.
446

447 **Acknowledgements**
448 We thank Dr. William Wuest for comments on the manuscript and members of the Conn lab for
449 discussions and comment through the course of this work.
450

451 **REFERENCES**

452 1. Guy RK, DiPaola RS, Romanelli F, Dutch RE. 2020. Rapid repurposing of drugs for
453 COVID-19. Science 368:829-830.
454 2. Choy KT, Wong AY, Kaewpreedee P, Sia SF, Chen D, Hui KPY, Chu DKW, Chan MCW,
455 Cheung PP, Huang X, Peiris M, Yen HL. 2020. Remdesivir, lopinavir, emetine, and
456 homoharringtonine inhibit SARS-CoV-2 replication in vitro. Antiviral Res 178:104786.
457 3. Costanzo M, De Giglio MAR, Roviello GN. 2020. SARS-CoV-2: Recent Reports on
458 Antiviral Therapies Based on Lopinavir/Ritonavir, Darunavir/Umifenovir,

459 Hydroxychloroquine, Remdesivir, Favipiravir and other Drugs for the Treatment of the New
460 Coronavirus. Curr Med Chem 27:4536-4541.

461 4. Abraham SM, Lawrence T, Kleiman A, Warden P, Medghalchi M, Tuckermann J,
462 Saklatvala J, Clark AR. 2006. Antiinflammatory effects of dexamethasone are partly
463 dependent on induction of dual specificity phosphatase 1. J Exp Med 203:1883-9.

464 5. Solinas C, Perra L, Aiello M, Migliori E, Petrosillo N. 2020. A critical evaluation of
465 glucocorticoids in the management of severe COVID-19. Cytokine Growth Factor Rev
466 doi:10.1016/j.cytogfr.2020.06.012.

467 6. Heidary F, Gharebaghi R. 2020. Ivermectin: a systematic review from antiviral effects to
468 COVID-19 complementary regimen. J Antibiot (Tokyo) doi:10.1038/s41429-020-0336-z.

469 7. da Silva CSB, Thaler M, Tas A, Ogando NS, Bredenbeek PJ, Ninaber DK, Wang Y,
470 Hiemstra PS, Snijder EJ, van Hemert MJ. 2020. Suramin inhibits SARS-CoV-2 infection in
471 cell culture by interfering with early steps of the replication cycle. Antimicrob Agents
472 Chemother doi:10.1128/AAC.00900-20.

473 8. Brun R, Blum J, Chappuis F, Burri C. 2010. Human African trypanosomiasis. Lancet
474 375:148-59.

475 9. Hawking F. 1978. Suramin: with special reference to onchocerciasis. Adv Pharmacol
476 Chemother 15:289-322.

477 10. Morgan HP, McNae IW, Nowicki MW, Zhong W, Michels PA, Auld DS, Fothergill-
478 Gilmore LA, Walkinshaw MD. 2011. The trypanocidal drug suramin and other trypan blue
479 mimetics are inhibitors of pyruvate kinases and bind to the adenosine site. J Biol Chem
480 286:31232-40.

481 11. Fleck SL, Birdsall B, Babon J, Dluzewski AR, Martin SR, Morgan WD, Angov E,
482 Kettleborough CA, Feeney J, Blackman MJ, Holder AA. 2003. Suramin and suramin
483 analogues inhibit merozoite surface protein-1 secondary processing and erythrocyte invasion
484 by the malaria parasite Plasmodium falciparum. J Biol Chem 278:47670-7.

485 12. Wiedemar N, Hauser DA, Maser P. 2020. 100 Years of Suramin. Antimicrob Agents
486 Chemother 64.

487 13. Fantini J, Hammache D, Delezay O, Yahi N, Andre-Barres C, Rico-Lattes I, Lattes A. 1997.
488 Synthetic soluble analogs of galactosylceramide (GalCer) bind to the V3 domain of HIV-1
489 gp120 and inhibit HIV-1-induced fusion and entry. J Biol Chem 272:7245-52.

490 14. Aguilar JS, Rice M, Wagner EK. 1999. The polysulfonated compound suramin blocks
491 adsorption and lateral difusion of herpes simplex virus type-1 in vero cells. Virology
492 258:141-51.

493 15. Garson JA, Lubach D, Passas J, Whitby K, Grant PR. 1999. Suramin blocks hepatitis C
494 binding to human hepatoma cells in vitro. J Med Virol 57:238-42.

495 16. Chen Y, Maguire T, Hileman RE, Fromm JR, Esko JD, Linhardt RJ, Marks RM. 1997.
496 Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate.
497 Nat Med 3:866-71.

498 17. Ren P, Zheng Y, Wang W, Hong L, Delpeyroux F, Arenzana-Seisdedos F, Altmeyer R.
499 2017. Suramin interacts with the positively charged region surrounding the 5-fold axis of the
500 EV-A71 capsid and inhibits multiple enterovirus A. Sci Rep 7:42902.

501 18. Albulescu IC, Kovacikova K, Tas A, Snijder EJ, van Hemert MJ. 2017. Suramin inhibits
502 Zika virus replication by interfering with virus attachment and release of infectious particles.
503 Antiviral Res 143:230-236.

504 19. Mastrangelo E, Pezzullo M, Tarantino D, Petazzi R, Germani F, Kramer D, Robel I,
505 Rohayem J, Bolognesi M, Milani M. 2012. Structure-based inhibition of Norovirus RNA-
506 dependent RNA polymerases. J Mol Biol 419:198-210.

507 20. Albulescu IC, van Hoolwerff M, Wolters LA, Bottaro E, Nastruzzi C, Yang SC, Tsay SC,
508 Hwu JR, Snijder EJ, van Hemert MJ. 2015. Suramin inhibits chikungunya virus replication
509 through multiple mechanisms. Antiviral Res 121:39-46.

510 21. Yin W, Luan X, Li Z, Zhou Z, Wang Q, Gao M, Wang X, Zhou F, Shi J, You E, Liu M,
511 Wang Q, Jiang Y, Jiang H, Xiao G, Zhang L, Yu X, Zhang S, Eric Xu H. 2021. Structural
512 basis for inhibition of the SARS-CoV-2 RNA polymerase by suramin. Nat Struct Mol Biol
513 28:319-325.

514 22. Ullmann H, Meis S, Hongwiset D, Marzian C, Wiese M, Nickel P, Communi D, Boeynaems
515 JM, Wolf C, Hausmann R, Schmalzing G, Kassack MU. 2005. Synthesis and structure-
516 activity relationships of suramin-derived P2Y11 receptor antagonists with nanomolar
517 potency. J Med Chem 48:7040-8.

518 23. Trapp J, Meier R, Hongwiset D, Kassack MU, Sippl W, Jung M. 2007. Structure-activity
519 studies on suramin analogues as inhibitors of NAD+-dependent histone deacetylases
520 (sirtuins). ChemMedChem 2:1419-31.

521 24. Kaina B. 2021. On the Origin of SARS-CoV-2: Did Cell Culture Experiments Lead to
522 Increased Virulence of the Progenitor Virus for Humans? In Vivo 35:1313-1326.

523 25. Murakami MT, Arruda EZ, Melo PA, Martinez AB, Calil-Elias S, Tomaz MA, Lomonte B,
524 Gutierrez JM, Arni RK. 2005. Inhibition of myotoxic activity of Bothrops asper myotoxin II
525 by the anti-trypanosomal drug suramin. J Mol Biol 350:416-26.

526 26. Schuetz A, Min J, Antoshenko T, Wang CL, Allali-Hassani A, Dong A, Loppnau P, Vedadi
527 M, Bochkarev A, Sternglanz R, Plotnikov AN. 2007. Structural basis of inhibition of the
528 human NAD+-dependent deacetylase SIRT5 by suramin. Structure 15:377-89.

529 27. Ren C, Morohashi K, Plotnikov AN, Jakoncic J, Smith SG, Li J, Zeng L, Rodriguez Y,
530 Stojanoff V, Walsh M, Zhou MM. 2015. Small-molecule modulators of methyl-lysine
531 binding for the CBX7 chromodomain. Chem Biol 22:161-8.

532 28. Jiao L, Ouyang S, Liang M, Niu F, Shaw N, Wu W, Ding W, Jin C, Peng Y, Zhu Y, Zhang
533 F, Wang T, Li C, Zuo X, Luan CH, Li D, Liu ZJ. 2013. Structure of severe fever with
534 thrombocytopenia syndrome virus nucleocapsid protein in complex with suramin reveals
535 therapeutic potential. J Virol 87:6829-39.

536 29. Kang S, Yang M, Hong Z, Zhang L, Huang Z, Chen X, He S, Zhou Z, Zhou Z, Chen Q, Yan
537 Y, Zhang C, Shan H, Chen S. 2020. Crystal structure of SARS-CoV-2 nucleocapsid protein
538 RNA binding domain reveals potential unique drug targeting sites. Acta Pharm Sin B
539 doi:10.1016/j.apsb.2020.04.009.

540 30. Yahi N, Sabatier JM, Nickel P, Mabrouk K, Gonzalez-Scarano F, Fantini J. 1994. Suramin
541 inhibits binding of the V3 region of HIV-1 envelope glycoprotein gp120 to

542 galactosylceramide, the receptor for HIV-1 gp120 on human colon epithelial cells. J Biol
543 Chem 269:24349-53.

544 31. Dillon SC, Dorman CJ. 2010. Bacterial nucleoid-associated proteins, nucleoid structure and
545 gene expression. Nat Rev Microbiol 8:185-95.

546 32. Stella S, Cascio D, Johnson RC. 2010. The shape of the DNA minor groove directs binding
547 by the DNA-bending protein Fis. Genes Dev 24:814-26.

548 33. Bhowmick T, Ghosh S, Dixit K, Ganesan V, Ramagopal UA, Dey D, Sarma SP,
549 Ramakumar S, Nagaraja V. 2014. Targeting Mycobacterium tuberculosis nucleoid-
550 associated protein HU with structure-based inhibitors. Nat Commun 5:4124.

551 34. Dey D, Nagaraja V, Ramakumar S. 2017. Structural and evolutionary analyses reveal
552 determinants of DNA binding specificities of nucleoid-associated proteins HU and IHF. Mol
553 Phylogenet Evol 107:356-366.

554 35. Nautiyal A, Patil KN, Muniyappa K. 2014. Suramin is a potent and selective inhibitor of
555 Mycobacterium tuberculosis RecA protein and the SOS response: RecA as a potential target
556 for antibacterial drug discovery. J Antimicrob Chemother 69:1834-43.

557 36. Dunstan MS, Hang PC, Zelinskaya NV, Honek JF, Conn GL. 2009. Structure of the
558 thiostrepton resistance methyltransferase.S-adenosyl-L-methionine complex and its
559 interaction with ribosomal RNA. J Biol Chem 284:17013-20.

560 37. Peng Q, Peng R, Yuan B, Zhao J, Wang M, Wang X, Wang Q, Sun Y, Fan Z, Qi J, Gao GF,
561 Shi Y. 2020. Structural and Biochemical Characterization of the nsp12-nsp7-nsp8 Core
562 Polymerase Complex from SARS-CoV-2. Cell Rep 31:107774.

563 38. Pillaiyar T, Funke M, Al-Hroub H, Weyler S, Ivanova S, Schlegel J, Abdelrahman A,
564 Muller CE. 2020. Design, synthesis and biological evaluation of suramin-derived dual
565 antagonists of the proinflammatory G protein-coupled receptors P2Y2 and GPR17. Eur J
566 Med Chem 186:111789.

567 39. Marques AF, Esser D, Rosenthal PJ, Kassack MU, Lima LM. 2013. Falcipain-2 inhibition
568 by suramin and suramin analogues. Bioorg Med Chem 21:3667-73.

569 40. Zhu W, Xu M, Chen CZ, Guo H, Shen M, Hu X, Shinn P, Klumpp-Thomas C, Michael SG,
570 Zheng W. 2020. Identification of SARS-CoV-2 3CL Protease Inhibitors by a Quantitative
571 High-throughput Screening. bioRxiv doi:10.1101/2020.07.17.207019.

572 41. Watts KS, Dalal P, Murphy RB, Sherman W, Friesner RA, Shelley JC. 2010. ConfGen: a
573 conformational search method for efficient generation of bioactive conformers. J Chem Inf
574 Model 50:534-46.

575 42. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll
576 EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. 2004. Glide: a new approach for
577 rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med
578 Chem 47:1739-49.

579 43. Dey D, Kavanaugh LG, Conn GL. 2020. Antibiotic Substrate Selectivity of Pseudomonas
580 aeruginosa MexY and MexB Efflux Systems Is Determined by a Goldilocks Affinity.
581 Antimicrob Agents Chemother 64.

582 44. Sastry M, Lowrie JF, Dixon SL, Sherman W. 2010. Large-scale systematic analysis of 2D
583 fingerprint methods and parameters to improve virtual screening enrichments. J Chem Inf
584 Model 50:771-84.

585 45. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-
586 generation sequencing data. Bioinformatics 28:3150-2.

587 46. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis
588 Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870-4.

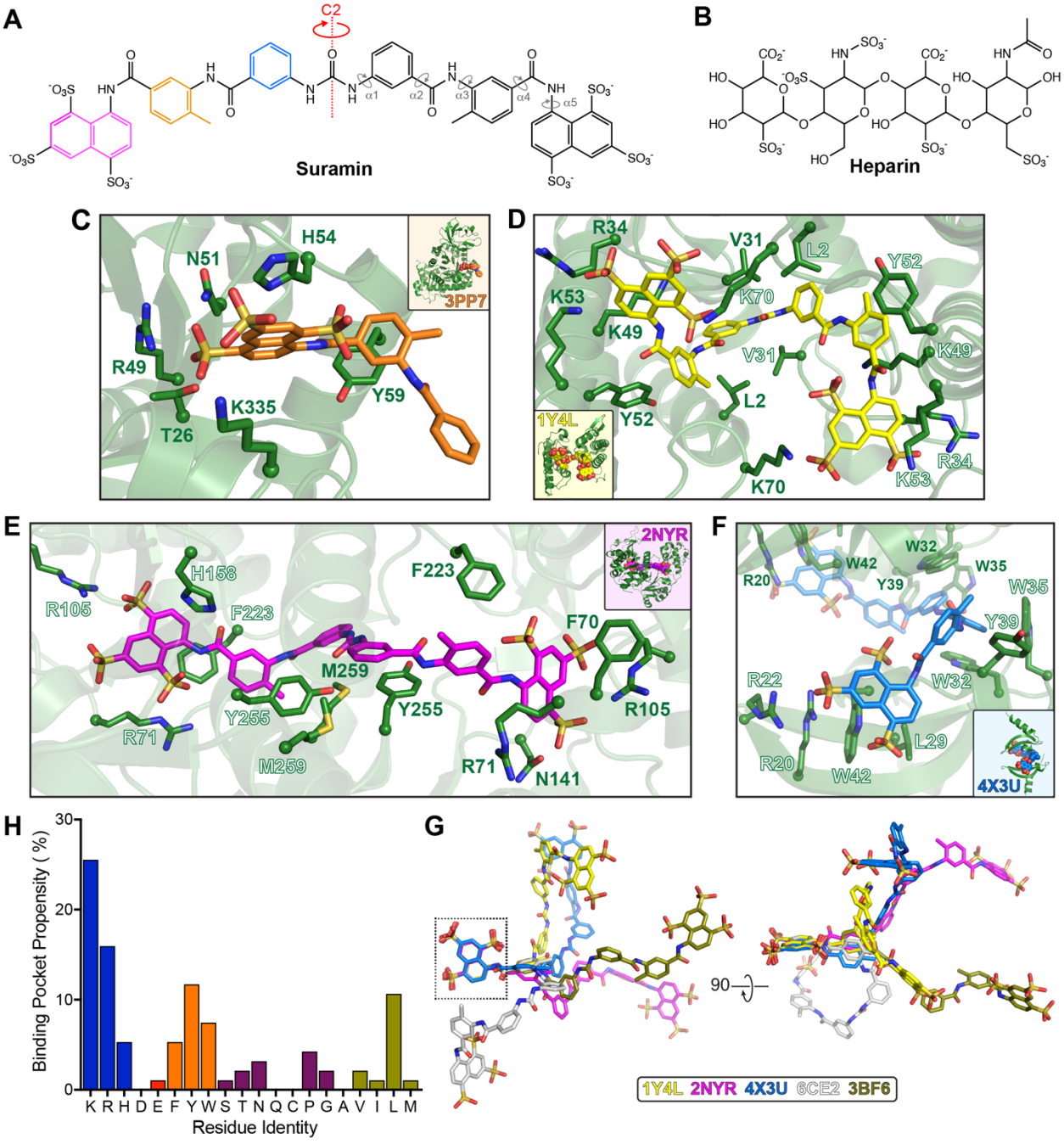
589 47. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. 2016. ConSurf
590 2016: an improved methodology to estimate and visualize evolutionary conservation in
591 macromolecules. Nucleic Acids Res 44:W344-50.

592

593

594

595 **FIGURE LEGENDS**

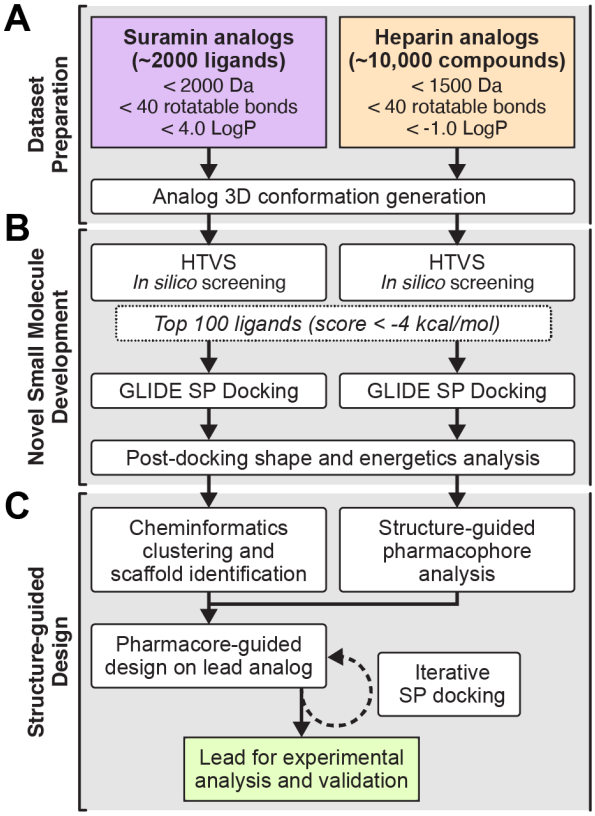


596

597 **Figure 1. Structural basis for suramin protein target site diversity.** Chemical structures of ***A***,
598 suramin and ***B***, heparin. Crystal structures of suramin (or partial fragment) bound to: ***C***,
599 *Leishmania mexicana* pyruvate kinase (PDB 3PP7), ***D***, snake venom protein myotoxin (PDB
600 1Y4L), ***E***, NAD^+-dependent protein deacetylase SIRT5 (PDB 2NYR), and ***F***, epigenetic reader
601 chromobox homolog 7 (PDB 4X3U). ***H***, Residue propensities in the suramin binding pockets in

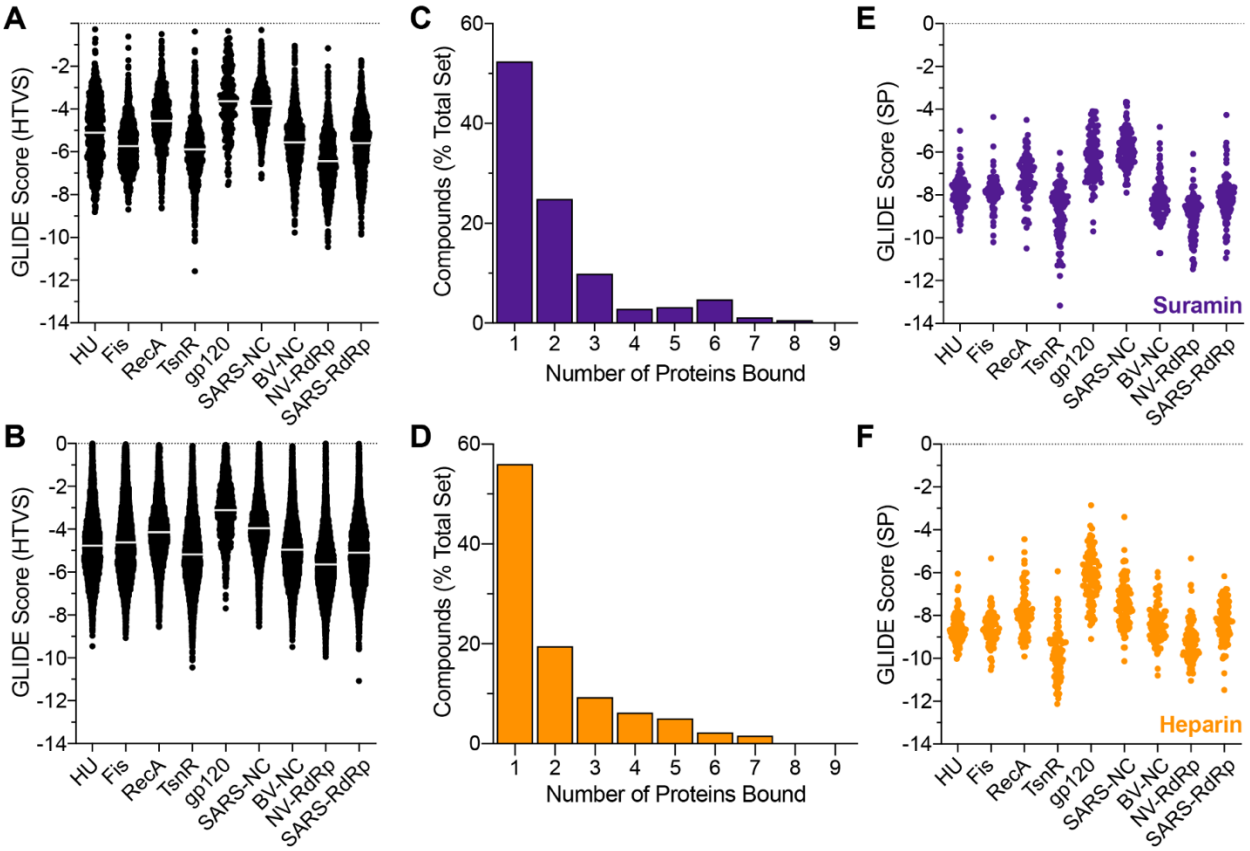
602 targets shown in *panels C-F* and **Figure S1**. ***G***, Superposition via the boxed sulfo-naphthyl ring of
603 suramin molecules from co-crystal structures with diverse targets (PDB 1Y4L, 2NYR, 4X3U,
604 6CE2 and 3BF6).

605



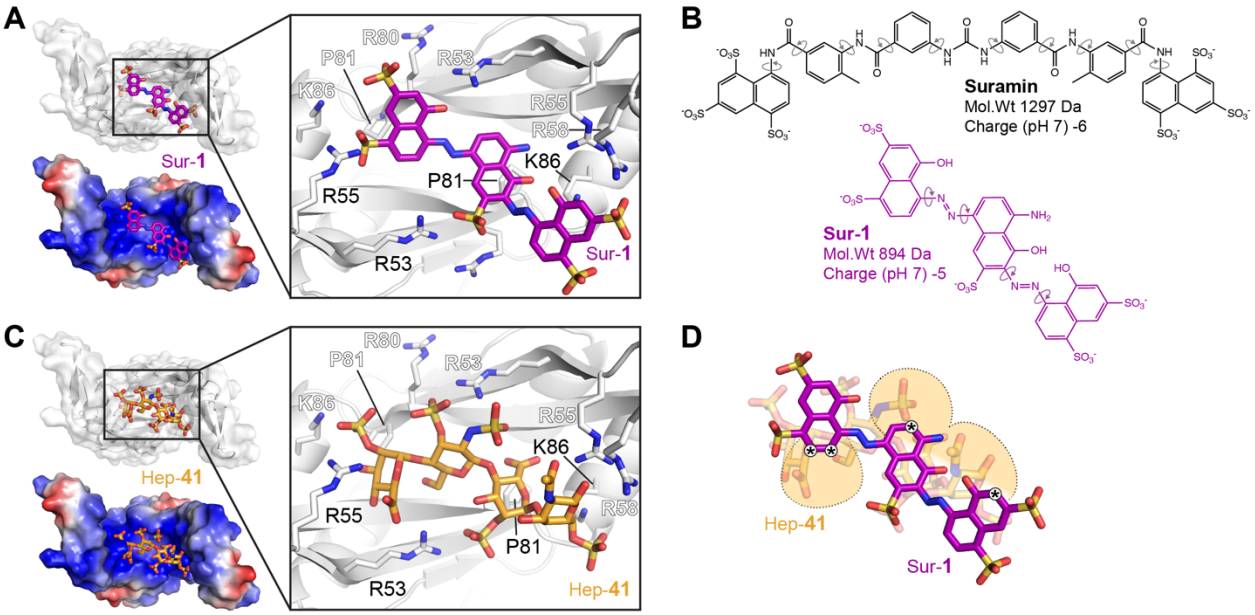
606

607 **Figure 2. Computational workflow used in this study.** ***A***, Suramin and heparin analog sets were
608 retrieved from PubChem with the indicated parameter filters and subject to conformer generation
609 in LigPrep (Schrödinger Software). ***B*****,** Suramin and heparin analog sets were used for separate
610 HTVS and the resulting top 100 ligands then subject to higher precision (SP) docking in GLIDE
611 (Schrödinger Software). ***C***, Following further shape and energetic analysis, chemoinformatics was
612 used to determine different fragments of top scoring suramin analogs. For heparin analogs, the top
613 poses corresponding to the top selected suramin analog for each target was next used for
614 pharmacophore-based identification of regions in the suramin analog which could be substituted
615 to improve its predicted affinity and target specificity. Identified novel analogs would next be used
616 as leads for experimental analysis and validation.



617

618 **Figure 3. Virtual screening and precision docking identify suramin analogs with increased**
619 **target specificity.** HTVS docking scores for ***A,*** suramin and ***B,*** heparin analogs against bacterial
620 targets HU, Fis, RecA and TsnR; and, viral targets gp120, SARS-CoV-2 and norovirus RdRp, and
621 SARS-CoV-2 and bunyavirus nucleocapsid (SARS-NC and BV-NC, respectively). A high
622 percentage of both ***C***, suramin and ***D***, heparin analogs bind to only one or two targets within the
623 protein panel used, suggesting that virtual screening has identified more specific leads as compared
624 to the parent molecules, suramin and heparin. ***E***,***F***, SP GLIDE docking scores for the top 100
625 ligands from HTVS for suramin and heparin analogs, respectively.



626

627 **Figure 4. Docking-guided elaboration of the top identified suramin analog against HU.** ***A*****,** SP
628 GLIDE docking pose of suramin analog Sur-1 (purple) bound to the DNA interaction surface of
629 HU (top, white surface; bottom, electrostatic potential surface). ***B***, Comparison of the chemical
630 structures of suramin and Sur-1, highlighting the reduced number of rotatable bonds in the analog.
631 ***C*****,** SP GLIDE docking pose of heparin analog Hep-1 (orange) bound to the same surface of HU.
632 ***D***, The overlapping binding poses of Hep-1 and Sur-1 which were used to identify sites in Sur-1
633 for redesign (indicated *) to improve binding affinity and specificity.

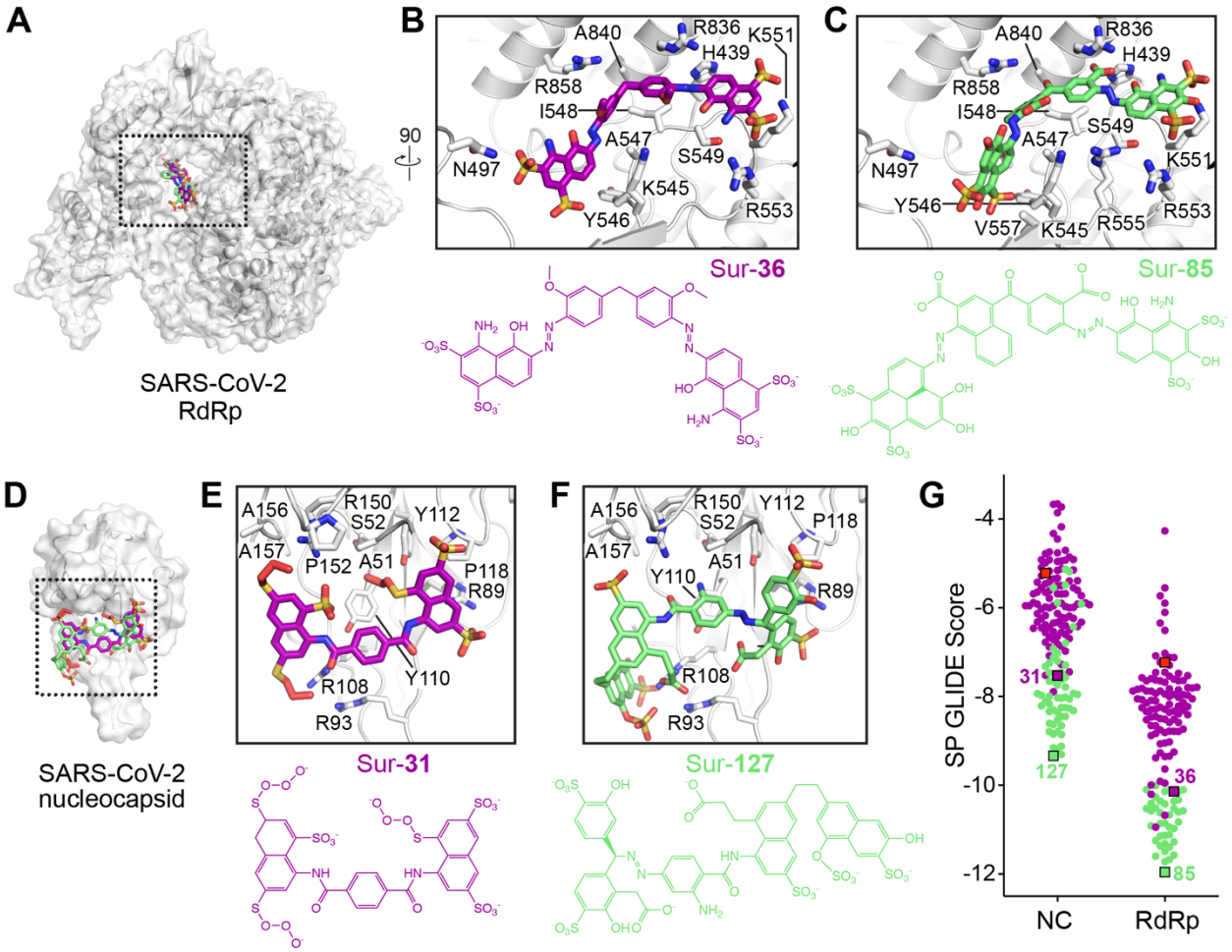
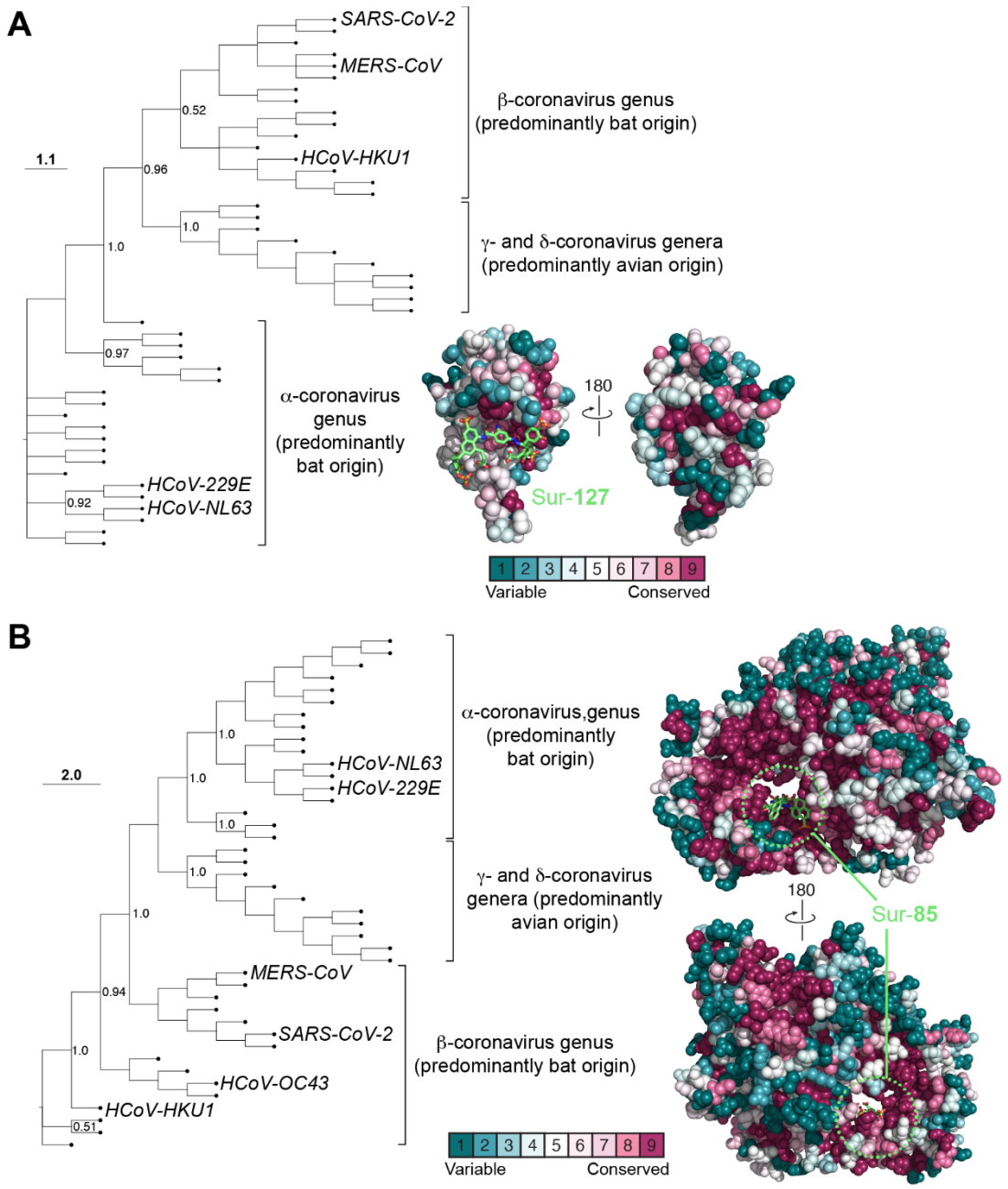


Figure 5. Design of suramin analogs specific to SARS-CoV-2 RdRp and nucleocapsid. ***A***, Location of suramin analogs Sur-36 and Sur-84 in the RNA binding tunnel of SARS-CoV-2 RdRp. ***B***, Zoomed in views of the boxed region in *panel A* of the SP GLIDE docking pose of Sur-36, an existing analog of suramin, bound to the RNA binding tunnel. ***C***, Iterative rounds of chemical enumeration (**Figure S6**) and docking, resulted in the designed suramin analog Sur-84 shown in the same binding pocket. ***D***, Location of suramin analogs Sur-31 and Sur-127 on the RNA binding surface of SARS-CoV-2 nucleocapsid. ***E***,***F***, Zoomed in views of the boxed region in *panel D* of the SP GLIDE docking poses of Sur-31 and Sur-127 respectively. ***G***, Comparison of SP GLIDE docking scores for the initial top 100 suramin analogs (purple) and analogs tested during the iterative chemical enumeration process (green), showing the improvement in predicted affinity. Suramin (red square), starting analogs (Sur-31 and Sur-36 for nucleocapsid (NC) and RdRp, respectively) and final redesigned analogs (Sur-127 and Sur-85) are highlighted.



647

648 **Figure 6. Maximum likelihood phylogenetic tree of coronavirus RdRp and nucleocapsid.** ***A***,
649 Phylogenetic tree of coronavirus capsid protein showing a separate clade for the β-coronavirus
650 genus, with SARS-CoV-2 and MERS-CoV, which are mostly of bat origin. This clade is closely
651 related to γ coronaviruses of avian origin while another bat origin α-coronavirus family forms a
652 separate clade. Conservation of residues in coronavirus capsid proteins shows the binding pocket
653 targeted by the designed suramin analog Sur-127 is evolutionarily conserved. ***B***, As *panel A* but
654 for SARS-CoV-2 RdRp and Sur-85