1 **A molecular docking study of human STEAP2 for the discovery of new**
2 **potential anti-prostate cancer chemotherapeutic candidates**
3 **Docking of human STEAP2 for new anti- prostate cancer**
4 **chemotherapeutics**

5 Author list.

6 Timothy Ongaba^{1¶*}, Christian Ndekezi^{12&}, Nana Jaqueline Nakiddu^{45&}

7 ¹Department of Biomolecular Resources and Biolaboratory Sciences, College of
8 Veterinary Medicine, Animal Resources and Biosecurity (CoVAB), Makerere
9 University, Kampala, Uganda.

10 ²Uganda Virus Research Institute, Entebbe, Uganda.

11 ³Joint Clinical Research Centre, Kampala Uganda

12 ⁴College of Health Sciences (CHS), Makerere University

13 Email: dextertim5@gmail.com

14 **Abstract**

15 Prostate cancer refers to uncontrolled abnormal cell growth (Cancer) within the prostate
16 gland. The disease is a rising health concern and accounts for 3.8% of all cancer deaths
17 globally. Uganda has one of the highest incidence rates of the disease in Africa being 5.2%
18 and many of the diagnosed patients are found to have advanced stage prostate cancer. This
19 study aimed to use STEAP2 protein (prostate cancer specific biomarker) for the discovery of
20 new lead compounds against prostate cancer.

21 To determine the most likely compound that can bind to STEAP2 protein, we docked the
22 modelled STEAP2 3 Dimension structure against 2466 FDA (Food and Drug Administration)
23 approved drug candidates using Autodock vina. Protein Basic Local Alignment Search Tool
24 (BLASTp) search, Multiple Sequence Alignment (MSA), and phylogenetics were further
25 carried out to analyse the diversity of this marker and determine its conserved domains as
26 suitable target regions.

27 Six promising drug candidates (ligands) were identified of which Triptorelin had the highest
28 binding energy (-12.1 kcal/mol). Leuprolide was the second most promising candidate with a
29 docking energy of -11.2 kcal/mol. All the top 3 of 6 drugs interacted with highly conserved
30 residues Ser-372 and Gly-369 in close proximity with the iron binding domain that was
31 shown to be important for catalysis of metal reduction.

32 The two drugs had earlier been approved for treatment of advanced prostate cancer but with
33 an elusive mode of action. Thanks to this study we now have an insight on how their
34 interaction with STEAP2 might be important during treatment.

35 **Author summary**

36 Previous studies on prostate cancer prevalence have shown that Uganda among other low to
 37 middle income countries has a high prostate cancer incidence rate. Majority of the patients at
 38 the time of diagnosis have advanced stage disease. While there has been great improvement
 39 in therapeutic options for prostate cancer over the last decade, the drugs currently used are
 40 still limited and not 100% effective making cure elusive. This is further compounded by
 41 undesirable side-effects from some of the treatments. Altogether this creates a need for the
 42 discovery of novel, safe, and efficacious chemotherapeutic agents with minimal or no side
 43 effects. In this study, we used a prostate cancer specific protein biomarker; STEAP2 that is
 44 highly expressed at all disease stages and is androgen independent and FDA approved drugs
 45 from the drugbank to help improve drug specificity, efficacy, and reduce undesirable side
 46 effects through *in-silico* methods. The results from the study give an insight on mechanisms
 47 of action of current therapy for advanced disease and suggest these very drugs be used even
 48 at early stage disease.

49

50 **Introduction**

51 Cancer is a disease in which abnormal cells divide uncontrollably and in the process destroy
 52 body tissue. It is the second leading cause of death before age 70 years in 91 of 172 countries
 53 according to the World Health Organisation in 2015 (1). It is responsible for an estimated 9.6
 54 million deaths in 2018 (1). One in 5 men and 1 in 6 women develop cancer in their lifetime
 55 while 1 in 8 men and 1 in 11 women die from the disease (1). As of 2018, the global cancer
 56 burden had risen to 18.1 million new cases (1). Worldwide, the total number of people alive
 57 with a 5-year cancer diagnosis, called the 5-year prevalence, was estimated to be 43.8 million
 58 (2). About 70% to 80% of the avoidable cancer mortality occurs in Low to Middle Income
 59 Countries (LMICs) (3). Global cancer incidence is on the rise and it is predicted that by 2050,
 60 70% of the 24 million annual cancer diagnoses will be individuals residing in LMICs (4).

61 Prostate cancer is the second most frequently diagnosed cancer in men globally and accounts
 62 for 3.8% of all deaths caused by cancer in men as of 2018 (1,5). In Uganda for a 16-year
 63 period from 1991 to 2006 there was an increase in cancer risk particularly for breast and
 64 prostate cancer (4.5% annually) (6). The incidence of prostate cancer in Uganda is among the
 65 highest recorded in Africa at 39.6 per 100,000 after age-standardizing (6). A recent study on
 66 prostate cancer burden puts its annual incidence at 5.2% (7).

67 Prostate cancer occurs within the prostate. It can be classified into Localized and metastatic
 68 disease depending on absence or presence of spread respectively. Localized prostate cancer is
 69 classified as high risk based on clinical staging, prostate specific antigen (PSA) levels, and/or
 70 the Gleason score (8). Patients with metastatic prostate cancer are at higher risk of disease
 71 and death. The disease can be categorized as castration-sensitive (mCSPC) when surgical
 72 removal of the testicles halts cancer advancement due to decreased blood testosterone

73 androgen or castration-resistant (mCRPC) when the cancer continues to progress even in the
74 absence of testosterone androgen (8). Majority of mCSPC patients have a high risk of
75 progressing to mCRPC when initial hormonal treatment fails due to resistance (8).

76 Prostate cancer can be treated in several ways. However, overall disease management
77 depends on stage/severity and type of cancer. Surgery, nanotechnology for controlled drug
78 delivery, monoclonal antibody therapy, hormonal therapy, radiotherapy, and chemotherapy
79 are some of the treatment methods (9,10). Chemotherapy (especially when hormonal
80 treatment fails), surgery and radiotherapy are the most commonly used modalities of therapy.

81 Six Transmembrane Antigen of the Prostate 2 (STEAP2) also known as the Six
82 Transmembrane Protein of prostate 1 (STAMP1) is a member of the metalloredutase family
83 that are important in metal reduction of copper and iron. *In vitro* and *in vivo* studies show that
84 STEAP2 plays a key role in prostate cell ability to advance from localised cancer to
85 metastasis (11). STEAP2 is located on the plasma membrane of prostate cells and Golgi
86 complex. It increases prostate cancer progression, controls cell proliferation, differentiation
87 and decreases apoptosis (12). Immunohistochemical staining significantly demonstrate its
88 expression at the cell-cell junctions of prostate cancer cells (13). It is differentially expressed
89 in normal and cancerous tissue making it a potential target for new therapeutic strategies for
90 disease treatment (14). STEAP2 is expressed more than 10 times in normal prostate than
91 other tissues and is exponentially expressed in malignant prostate cancer cells (15,16). It is
92 expressed in low amounts in other tissues such as the brain, liver, skeletal muscle, mammary
93 glands, kidney, lung, uterus, and trachea (15,16). Its levels in these tissues is so low
94 compared to that in the prostate that it is unlikely to have any functional significance (16).

95 This study aimed at using the STEAP2 prostate cancer biomarker as a target for improving
96 current disease treatment regimen using *in-silico* methods and promising results were
97 achieved.

98 Results

99 STEAP2 relative tissue expression.

100 The bar graphs figs.1 to 4 show the differential expression of STEAP2 in different tissues as
 101 identified by different tissue gene expression software from the Human Protein Atlas.
 102 (<https://www.proteinatlas.org/>). Messenger Ribonucleic Acid (mRNA) expression analysis
 103 of the FANTOM5 dataset showed that STEAP2 was most abundantly expressed in prostate
 104 tissue followed by ovarian and vaginal tissue (Fig. 2) (17,18). Fig. 2 showed that prostate
 105 tissue had the most STEAP2 in mean protein coding transcripts per million as compared to all
 106 other listed body tissues using the HPA dataset (18,19).

107

108 **Fig. 1:** A bar graph showing the differential tissue expression of STEAP2 RNA tissue data reported as scaled
 109 tags per million by FANTOM5 dataset (source: <https://www.proteinatlas.org/ENG00000157214-STEAP2>). The
 110 graph shows that STEAP2 mRNA is most abundantly expressed in prostate tissue followed by ovarian and
 111 vaginal tissue

112

113 **Fig. 2:** A bar graph showing the differential tissue expression of STEAP2 in mean protein coding transcripts per
 114 million by HPA dataset (source: <https://www.proteinatlas.org/ENG00000157214-STEAP2>). The prostate tissue
 115 had a significantly high STEAP2 in mean protein coding transcripts per million as compared to all other listed
 116 body tissues followed by the appendix and parathyroid gland.

117 Fig. 3 further showed prostate tissue to have the highest quantity of STEAP2 mRNA
 118 transcripts from different tissue in mean protein coding transcripts per million in comparison
 119 to 33 other listed tissues using the GTex dataset (18,20). Fig. 4 combined data from the three
 120 datasets by a normalised eXpression level and showed the same trend. (17–20).

121

122 **Fig. 3:** A bar graph showing the differential tissue expression of STEAP2 by RNA sequence tissue reported in
 123 mean protein coding transcripts per million shown by GTex dataset (source:
 124 <https://www.proteinatlas.org/ENSG00000157214-STEAP2/tissue>). Prostate tissue has a highly significant
 125 content of STEAP2 RNA compared to all other tissues followed by lung and pituitary glands.

Fig. 4: Consensus Normalized eXpression (NX) levels for 55 tissue types and 6 blood cell types, created by combining the data from the three transcriptomics datasets (HPA, GTEx and FANTOM5) using the internal normalization pipeline. (Source: <https://www.proteinatlas.org/ENSG00000157214-STEAP2/tissue>). The graph maintains the initial observation that prostate tissue contains significantly high amounts of STEAP2 that any other body tissue of the 55 considered.

Data retrieval of STEAP2 sequence.

The protein *H. sapiens* STEAP2's (Uniprot accession number: Q8NFT2) in FAST All (FASTA) format from UniProt (Bateman *et al.*, 2017);

Multiple Sequence alignment.

A dendrogram generated using Molecular Evolution Genetic Analysis (MEGA) showed that STEAP2 was conserved mostly in species under phylum Chordata(Fig. 5) (21). The tree was a result of 1000 bootstrap iterations with the values at each node indicating the percentage times of 1000 that particular node was redrawn as a result of genetic relationship of its taxons.

The phylogenetic tree showed the evolutionary relationship between the 30 selected species in relationship to STEAP2. The highlighted clade shows species containing STEAP2 that also

happen to be under phylum Chordata. The species go on differentiating into sub taxa till the mammalian class and primates order where our query species; *H. sapiens* falls together with its most similar species; *G.g.gorilla*. as shown in Fig. 5. The rest of the clades contain other STEAP proteins and one NADP from other phyla.

147

Fig. 5 The Phylogenetic tree of *H.sapiens* STEAP2 shows that it is most related to *G.gorilla* gorilla since it is redrawn 78% of 1000 times by the bootstrap method. This therefore, indicates that these three species' have the most conserved regions (21,22)

Having clustered the STEAP2s alone, a second phylogenetic analysis of these was carried out to have a more accurate and precise evolutionary relationship between the different species' proteins (Fig. 6).

154 It is noteworthy that the mammalian species baring STEAP2 clustered together, followed by
155 those of reptilia, pisces, and vertebrata species; *A. radiata* with the tree having been rooted to
156 a yeast species; *S. cerevisiae*.

157
158 **Fig. 6:** From the phylogenetic tree, which is a result of 1000 bootstraps, it is shown that the STEAP2 protein
159 from G.g.gorilla (mammal) is most related to human STEAP2 from 99% of the 1000 redrawn trees and the least
160 related is A.radiata. The first four listed species from top belong to class Mammalia, followed by an alligator
161 (Reptilia), birds (Aves) fish (Pices) and a turtle (vertebrata) all under phylum Chordata.

162 **Analysis of highly conserved residues from the MSA that might**
163 **be responsible for the phylogenetic relationships of the different**
164 **species**

165 The MSA image depicts the conserved residues across 15 selected sequences of STEAP2
166 baring species and *S. cerevisiae* from the original 30 from the protein BLAST search. The
167 multiple sequence alignment (MSA) was done using MUSCLE in Jalview to determine the
168 most conserved residues among different species and these conserved residues were

169 highlighted in columns (23,24) The first MSA image of the 30 sequences is attached in
 170 SI_Fig.17. The colouring was done by clustalx in both images and the length of sequences in
 171 the wrap was adjusted to first 100 residue window. The sequences that were used for the
 172 MSA are detailed in SI_Table1.

173 Some of the conserved residues as highlighted in colour in the MSA with respective
 174 Weblogos to the left were found to be associated to functional motifs such as the Rossman
 175 fold that is closely associated with NADP domain (25) and some of the transmembrane
 176 domains were associated with iron binding domains.

Fig. 7: An MSA of STEAP2 proteins from 14 species from 4 different classes of chordates. The most conserved residues are associated with functional roles of the protein. From top, the Rossman fold, an alpha beta alpha secondary structure with sequence GxGxxG is closely associated with the dinucleotide phosphate group of NADP followed by the first trans- membrane domain (TM1) (25). The second, third, fourth and fifth TM domains (TM4,5,6 and 12 respectively) in close contact with the iron binding TM domain and finally the sixth TM domain

Homology modelling

The query sequence of STEAP2 was retrieved from UniProt in FASTA format (UniProt accession number: Q8NFT2). The template search results by the homology modelling engines together with parameters of similarity to STEAP2 are shown (table 1). Template 6HCY/6HDY (human STEAP4) was the most common top hit.

Since the engines search the PDB database and with different principles, there was a difference in top hits.

192 **Table 1: Top homology modelling templates from selected model engines and respective parameter of the hits.**

ENGINE	SEQUENCE IDENTITY (%)	COVERAGE	ORGANISM	PDB ID	STOICHIOMETRY	RESOLUTION (Å)	STRUCTURAL DETERMINATION
HHPRED	46.00	18-171	<i>H.sapiens</i>	6HCY_A	Homo-3-mer	3.1	EM
SWISS MODEL	47.61	30-466	<i>H.sapiens</i>	6HCY.1.A	Homo-3-mer	3.1	EM
PRIMO	53.00	33-215	<i>H.sapiens</i>	2VNS	Homo-2-mer	2.0	X-RAY DIFFRACTION
RAPTORX		31-466	<i>H.sapiens</i>	6HD1C	Homo-3-mer	3.8	EM
ROBETTA	47.83	1-490	<i>H.sapiens</i>	6HCYA	Homo-3-mer	3.1	EM
I-TASSER	47.90	0.890	<i>H.sapiens</i>	6HCYC	Homo-3-mer	3.1	EM

193 The common hit among 5 of 6 engines was human 6HCY/ 6HD1 (STEAP4) with a better
194 coverage than 2VNS identified by PRIMO (Table 1). In all but PRIMO, the top hit was
195 6HCY with better coverage than 2VNS as a template. This suggests 6HCY as the best
196 template for homology modelling.

197 The selected templates were used to model human STEAP2 and the homology models are
198 displayed in Fig. 8 in cartoon format and coloured by chain. The homology model from
199 PRIMO was not satisfactory because its best hit template (2VNS) had a very small template
200 coverage of 183 amino acids out of the 490 amino acid query sequence. PRIMO only
201 provides templates that have been resolved by X-ray crystallography as such, it was not
202 possible to select the 6HCY/ 6HD1 (STEAP4) template. X-ray crystallography is also limited
203 by size in terms of structure determination. It is ideal for small size proteins and not large
204 sized proteins like STEAP4 a homo-3-mer (26).

205 Homology models from the different engines (images generated in cartoon format using
206 PyMol).

Fig. 8: The images of the homology models were shown in cartoon format and coloured by chain. Image A shows the homology model by SWISSMODEL, B: The top homology model by I-Tasser, C: The homology model by HHPRED, D: The homology model by RaptorX, E: The homology model by Robetta and. F: The homology model by Primo. Only the homology model by SWISSMODEL was a homo trimer, the rest were single chain models and the used template; human STEAP4 is a homotrimer

The homology model from SWISSMODEL (Fig. 8-A) was a homo-3-mer made up of 437 amino acids per chain. The template used was 6HCY and the homology model had the least RMSD of 0.086 when superimposed with the 6HCY template (Table 2). Additionally, on average it had the best score by the used evaluation tools. For example, the modeller normalised DOPE score of -5.95, the Ramachandran plot having 94.4 of its residues lying within the favoured region a -3.06 and -6.33 Z-score from QMEAN and ProSA respectively. Altogether, these results show that the SWISSMODEL homology model was most similar model to the used template (27).

The I-TASSER model (Fig.8-B) was a monomer of 490 amino acids had a RMSD of 0.339, the second best after that by SWISSMODEL. It had the best ProSA Z score of -6.95 QMEAN Z score of -6.36 but the worst percentage of residues (87.5%) in the favoured region on a Ramachandran plot when compared to other homology models. It also had the second best Modeller normalised DOPE score of -.0238. It had no long ending loops and was therefore of satisfactory quality.

The HHPRED model (Fig.8-C) was a monomer of 448 amino acids. The used template was 6HCY which on superimposition with the homology model had a 4th best of 5 RMSD of 1.445. It had a second best ProSA Z score of -6.33 same as that for SWISSMODEL, a third best -3.03 QMEAN Z score, the best percentage of residues falling in the favoured region on a Ramachandran plot (96.7) but had the second last Modeller normalised DOPE score of 0.454. It had long loops which are indicators of non-satisfactory regional model quality.

233 RAPTORX model (Fig.8-D) was a monomer of 487 amino acids, had a third best RMSD of
234 0.367. It had the best percentage of residues within the favourable region on a Ramachandran
235 plot together with that from HHPRED. It however, had the second last of ProSA and
236 QMEAN Z scores, -6.06 and -2.91 respectively and the last Modeller normalised DOPE score
237 of 0.867. It also had noticeable long loops at the extremes of its upper and lower domains.

238 ROBETTA model (Fig.8-E) was a monomer of 28 amino acids. It was too short, a product
239 for a sequence of 490 residues. It had a 3rd best Modeller normalised DOPE score of 0.285, a
240 4th best 88% percentage of residues falling within the favoured region on a Ramachandran
241 plot. It however, had the last of 5 for the RMSD, ProSA and QMEAN Z scores that were
242 3.789, -0.59 and -0.19 respectively. It was, therefore, not a satisfactory homology model.

243 The model by PRIMO (Fig.8-F) was a monomer of 183 amino acids. This was short a
244 homology model for a query sequence of 490 amino acids. It was not considered for further
245 evaluation as the other 5 because its template was already different from them.

246 The above stated results are summarised in table 2.

247 **Table 2: Structural evaluation of homology models by different validation tools.**

MODELLING ENGINE	MODEL STOICHIOMETRY	ProSA (Z-Score)	QMEAN (Z- Score)	RAMPAGE			MODELER (NORMALISED DOPE SCORE)	AA MODELLED Z	RMSD WITH 6HCY
				FAVOURED	ALLOWED	OUTLIER			
HHRED	MONOMER	-6.33	-3.03	96.7	2.2	1.1	0.454	448	1.445
SWISSMODEL	HOMO-3-MER	-6.33	-3.06	94.4	4.8	0.2	- 0.595	437	0.086
I-TASSER	MONOMER	-6.95	-6.36	87.5	9.8	2.7	- 0.238	490	0.339
RAPTORX	MONOMER	-6.06	-2.91	96.7	2.7	0.6	0.867	487	0.367
ROBETTA	MONOMER	-0.59	-0.19	88.0	12.0	0.0	0.285	28	3.789

248 From the table, SWISS MODEL on average had the best homology model after validation by the listed validation tool

249 Fig.9 showed the ProSA evaluation of SWISSMODEL which was the best homology model
 250 together with its Ramachandran plot on the right. The z-score of -6.33 refers to the standard
 251 deviation of the homology model from an ensemble of misfolded proteins whose structures
 252 were determined by NMR and X-ray crystallography (28). The Ramachandran plot shows a
 253 plot of the homology model's residue in which those in black, dark grey and light grey
 254 regions represent highly preferred conformations (Favored), white with black grid region
 255 represents preferred conformations (Allowed) and white with grey grid lines region represents
 256 questionable conformations (outlier).

257
 258 **Fig.9** A ProSA z-score and Ramachandran plot of the SWISSMODEL homology model. The z-score of -6.33
 259 suggests that the homology model is properly folded and 96.7% of the protein residues being under the favoured
 260 region shows that almost all the protein residues are properly sterically placed in 3D space.

261 Molecular Docking

262 Figs 11 to 16 show the interaction of six different residues and part of the receptor molecule
 263 STEAP2. The types of bonds between interacting residue and ligand atoms such as hydrogen
 264 bonds and other forces of attraction between different atoms are shown in the 2 dimension
 265 ligplots as shown by the key. To the right, the 3-dimension image shows the ligand with the

266 respective interacting receptor residues that form hydrogen bonds with the respective ligands
267 labelled and shown as sticks.

269

270 **Fig.10:** A protein-ligand complex of STEAP2 and Drugbank1543 shows residue Ser-372 from the 6th TM
271 domain according to the MSA of STEAP2s and this forms a hydrogen bond with an atom of the ligand
272 (Drugbank1543). The rest of the interacting receptor residues are shown as crescents and these form other bonds
273 with the receptors. These interacting residues are many compared to other complexes and contribute a
274 significant binding energy of the complex.

275 Ser-372 with the hydrogen bond lies in the 6th TM domain which is part of the iron binding
276 domain and this is shown to be highly conserved from the MSA of only STEAP2 containing
277 species (Fig. 7). It together with many other shown protein amino acids bind with Triptorelin
278 a drug approved for treatment of advanced prostate cancer.

279

281 **Fig.11:** A protein-ligand complex of STEAP2 and Drugbank2 shows residues Glu-390, Gly-369, Ser-372 and
 282 Gly-369 within the 6th and 12th TM domain form hydrogen bonds with ligand Drugbank2.

283 The interacting protein amino acids lie within the TM domains that are highly conserved as
 284 shown in the MSA (Fig. 7). These are associated with the iron binding domain and bind to
 285 the ligand Leuprolide which is also approved for treatment of advanced prostate cancer.

286

287

288 **Fig.12:** In the above complex, receptor residues Thr-210 (NADP domain), Glu-390, and Lys-303 (TM5) form
 289 hydrogen bonds with ligand Drugbank2328.

290 Some of the listed residues such as Thr-210 are conserved under the NADP domain, the 5th
 291 TM domain having residues Glu-390 and Lys-303. The residues together with other form
 292 bonds with ligands Drugbank2328 which is 1,2 icosapentoyl-sn-glycero-3-phosphoserine an
 293 approved experimental drug for Corona virus-19 disease treatment.

294

295
296 **Fig.13:** In the above complex, receptor residues Glu-390, Ala-161,386, Asn-186,384, Lue-386, Tyr-166, Trp-
297 212, Ser-383, Lys-303, Arg-183 and Asp160 all form hydrden bonds with ligand DrugBank2154. The residues
298 are majorly from the NADP binding dominan and the 12th TM domain

299 The listed amino acids from the protein that form hydrogen bonds are all conserved highly in
300 the NADP binding domain and the 12th TM domain which is associated with iron binding.
301 The ligand Drugbank2154 is Etelcalcetide a calcimimetic drug used for treatment of
302 hyperparathyroidism.

Fig.14: In the above complex, receptor residues Thr-397, Ser-396 (TM12) and Lys-446 (TM17) form hydrogen bonds with the ligand Drugbank138

The mentioned residues in TM12 are associated with the iron binding domain of STEAP2 proteins in the MSA. The ligand Drugbank138 is Lymecycline a second generation tetracycline antibacterial used for broad-spectrum treatment of various bacterial infections.

310

311

Fig.15: In the above complex (A), residues Asp-160, Ser-162,150 and Ala-161 all within the NADP binding domain form hydrogen bonds with the ligand DrugBank1423.

From the images above, the protein ligand complex of ligand 1543 had the least number of hydrogen bonds between the ligand and protein but had the most number of hydrophobic contacts within the receptor. The summation of these bonds was therefore responsible for the high binding energy associated with the complex. The ligand Drugbank 1423 is Droxidopa a drug used in treatment of Parkinsonism.

319 Table 3 shows the protein-ligand complexes that met the thresholds of the binding energy.

320 Table 3: Protein-ligand complexes that met the binding energy threshold with respect to the
321 different chains of the protein.

Complex of 01_swissmodel 3D model with different drug compound	Compound name	Homologous chain	Binding energy (kcal/mol)
drugbank1543	Triptorelin	A	-12.1
drugbank2	Leuprolide		-11.2
drugbank2328	1,2 icosapentoyl-sn-glycero- 3-phosphoserine		-8.3
drugbank2154	Etelcalcetide		-8.2
drugbank1543	Triptorelin	B	-12.1
drugbank2	Leuprolide		-11.2
drugbank2328	1,2 icosapentoyl-sn-glycero- 3-phosphoserine		-8.3
drugbank138	Lymecycline		-8.1
drugbank1423	Droxidopa		-7.1
drugbank1543	Triptorelin	C	-12.1
drugbank2	Leuprolide		-11.2
drugbank2328	1,2 icosapentoyl-sn-glycero- 3-phosphoserine		-8.3
drugbank138	Lymecycline		-8.1

drugbank1423 Droxidopa -7.1

The complex; 01_swissmodel and drugbank1543 had the least binding energy which means it was the most stable of all complexes.

Docking Validation using independent docking engines.

The top 6 mentioned ligands' binding energy order was re-evaluated using an independent docking engine ; Patchdock and it produced the same order in terms of scores based on the geometric shape complementarity of solutions (29). The scores were ranked in the same order as the binding energy measured by autodock vina

Table 4 shows the top 3 solution of each complex ranked in order of their scores by Patchdock. The area stands for approximate interface area of the complex and atomic desolvation energy is measured according to Zhang *et al.* (30).

Table 4: Patchdock docking evaluation of docking scores by Autodock Vina

STEAP2 com- plex with;	Results' number	Score	Area	Atomic Desolvation Energy	Binding energy measured by Autodock vi- na(Kcal/mol)	Rank
Drugbank1543	1	11346	1452.70	-272.92	-12.1	1
	2	11170	1468.30	-254.68		
	3	1114	1512.50	-262.71		
Drugbank2	1	11274	1460.40	-454.04	-11.2	2
	2	11088	1490.30	-502.35		
	3	11086	1469.60	-501		
Drigbank2328	1	11224	1375.80	-421.80	-8.3	3

	2	10568	1383.10	-501.26		
	3	10506	1321.70	-462.1		
Drugbank2154	1	9240	1222.30	-769.57	-8.2	4
	2	9096	1187.60	-734730		
	3	9094	1132.30	-453.96		
Drugbank138	1	7074	814.80	-421.80	-8.1	5
	2	7016	797.00	-382.80		
	3	6970	841.00	-373.47		
Drugbank1423	1	3300	358.90	-133.28	-7.1	6
	2	3294	357.70	-125.81		
	3	3294	363.30	-122.28		

333 The exact docking poses from autodock vina could not be reproduced by patchdock despite
334 maintaining the same ranking order. Another docking engine; Autodock tools could not
335 reproduce the exact docking poses.

336 Discussion

337 The development of a more specific drug for the treatment of prostate cancer not only
338 reduces side effects of the current regimen but increases potency and effectiveness of the
339 candidate. With a modelled homologue and docking process of a pool of approved candidates
340 from the drug bank, a number of drug compounds that formed significant interactions with
341 the homology model gave insights of some of the current drugs that could be repurposed for
342 the treatment of prostate cancer.

343 The best homology model of STEAP2 was generated by SwissModel as shown by the results
344 from assessment by the 5 evaluation tools used (Fig.8-A). SwissModel, in a study by Wallnar
345 has been shown to accurately model homologues whose templates are of as low as 40%
346 similarity (31). This suggests that the modelling engine is reliable.

347 On model analysis, a lower score below 100% in the favoured region on a Ramachandran
348 plot shows that there are residues whose phi and psi angles are not sterically possible when
349 modelled (32). A lower percentage of residues in the favoured region therefore indicates a
350 poor homology model as shown for models from Robetta and I-Tasser.

351 From the docking results, ligand Drugbank1543 had the least binding energy and highest
352 affinity from a selected set of 2466 ligands. Having the least binding energy means the
353 complex is more naturally stable due to affinity between the two constituent molecules with
354 no or less external energy applied. The ligand formed its hydrogen bond with a residue
355 Ser372 in the transmembrane domain of the receptor molecule as most of its other non-
356 hydrogen bond interactions.

357 From a key of the drugbank, Drugbank1543, accession number DB06825, generic name
358 Triptorelin is a synthetic decapeptide gonadotrophin releasing hormone (GnRH) agonist. The

359 drug is proven for treatment of advanced stage prostate cancer as palliative care to reduce
360 sequelae of advanced disease such as compression of the spinal cord (33). With the results
361 from this study, this drug could potentially be used to treat early stage disease targeting
362 STEAP2 that is shown to be responsible for disease progression.

363 The study by Ravenna *et al.* demonstrates that Triptorelin has an inhibitory-stimulatory effect
364 on Lymph Node Carcinoma of the Prostate (LNCaP) cells at lower and high doses
365 respectively. In the experiment, in Pc3 cell lines, only low affinity but high capacity receptors
366 were expressed as a biological reaction to the drug. Both low and moderately high affinity
367 with high and low capacity receptors respectively were expressed in LNCaP cells. (34). This
368 shows the potency of Triptorelin and its binding energy in the treatment of prostate cancer at
369 different stages. Since STEAP2 is present in both cell types and given the binding
370 relationship between the protein and drug, this study suggests a mode of action of the latter.

371 Drugbank2 accession number DB00007 generic name Leuprolide is a synthetic 9 residue
372 peptide analogue of GnRH with the same mode of action as Triptorelin and is equally
373 indicated for treatment of advanced prostate cancer (35,36). It has been shown to have a
374 quicker onset of action in achieving serum testosterone levels equivalent to medically
375 castrated men compared to Triptorelin (37). The findings from this study suggest STEAP2 as
376 a target route for action.

377 Data from the protein atlas places brain tissue (pituitary gland, cerebral cortex and basal
378 ganglia) third after the ovary and vagina. In men where the ovary and vagina are absent, brain
379 tissue becomes second after the prostate though with a significant difference in tissue
380 STEAP2 expression. This could account for some of the potency of the GnRH receptor
381 analogue agonists in treating prostate cancer as mentioned above.

Drugbank2328, accession number DB14096, 1,2 icosapentoyl-sn-glycero-3-phosphoserine is an approved experimental drug for the treatment of Corona Virus Disease-19 (COVID-19) (38). Drugbank138, accession number, DB00256, Lymecline is a broad spectrum second generation tetracycline antibacterial used for treatment of susceptible bacterial infections such as acnes (39). Drugbank1423, accession number DB06262, generic name, Droxidopa a precursor for noradrenaline and is in the treatment of Parkinsonism (40). Drugbank2154, accession number DB12865, generic name Etelcalcetide is a calcimimetic drug used for the treatment of secondary hyperparathyroidism in patients having haemodialysis. From the findings of this study, these drugs could treat prostate cancer by binding to and inhibiting the catalytic role of STEAP2 that is responsible for disease progression. Further investigation on their role in treatment of prostate cancer is still warranted.

The exact docking poses could not be reproduced by autodock tools probably because the engine requires use of a single chain of the homo-3-mer for the protein ligand docking vis-à-vis the 3 identical chains used in the docking by autodock vina. The coherence in ranking order of protein ligand complex stability between autodock vina and patchdock, however, suggests the complexes are rightfully ordered.

Due to time and cost limitations we were unable to successfully validate the docking poses of the top ligands using different and independent docking engines. We recommend that future work should include reproducing the docking poses using other commercial and/or free docking engines in addition to molecular docking simulations. The docking engine used (Autodock vina) is, however, documented to be one of the best at the time as shown in the study by Sousa *et al.* (41). Additionally, we also recommend further analysis in terms of *in-vitro* assays to validate the results from this *in-silico* study on growth and viability changes in the presence of these drugs.

406 **Materials and methods**

407 **Study design**

408 The study involved retrieval of STEAP2 related protein sequences from the protein data bank
 409 from a protein BLAST search. The selected sequences were used for an MSA to phylogenetic
 410 analysis to determine the most related species to human STEAP2 protein. An MSA was used
 411 to determine highly conserved residues important in function of STEAP2 protein.

412 For molecular docking, a set of FDA approved drugs were retrieved from the drug bank and
 413 these were used as ligands for protein-ligand docking. The protein (STEAP2) 3 Dimension
 414 structure was acquired from homology modelling since it did not exist as an approved
 415 structure in the protein data bank. The homology model, after being evaluated was used for
 416 molecular docking with the drugs from the drug bank and the results of each were ranked.

417 The Protein-ligand interactions were characterised using ligpots and protein interacting
 418 residues were correlated with the those earmarked for protein function from the MSA and
 419 evaluated using independent docking engines. This entire methodology is summarised in
 420 Fig.16.

Fig.16: A study design of the methodology. The relationship of data from multiple sequence alignment of related species to human STEAP2 gave insight on key protein residues important in function of the protein to look out for in ligand interaction with the protein during protein-ligand docking.

STEAP2 relative tissue expression.

To determine the expression of STEAP2 protein in different tissues so as to ascertain its specificity as a drug target, an expression analysis was done using the Human Protein atlas (<https://www.proteinatlas.org/ENSG00000157214-STEAP2>) that showed its relative expression in over 50 different body tissue in addition to the prostate (18).

434 Data retrieval

435 The amino acid sequence of the query protein *H. sapiens* STEAP2 (Uniprot accession
436 number: Q8NFT2) was obtained in FASTA format from UniProt (42). This sequence was
437 used in the protein BLAST search for homologous sequences in the PDB.

438 STEAP2 Multiple Sequence Alignment and phylogenetic analysis

439 The MSA was done to determine the essential residues for the metalloredutase activity of
440 STEAP proteins. A protein BLAST (BLASTp) search of the reference proteins in the PDB
441 was done with the algorithm parameter of maximum target sequences to return adjusted to
442 5000. 30 sequences of the 5000 returned were selected for MSA. These were selected basing
443 on percentage similarity in bins of 0-40, 41-60, 61-80 and 81-100%. A number of
444 sequences, was selected from each range to have a uniform selection and also minimise bias
445 in the phylogenetic analysis. (Table 5).

446 **Table 5: Sequences selected for phylogenetic analysis of STEAP2 by MSA.**

SIMILARITY RANGE (%)	SEQUENCES SELECTED
0-40	8
41-60	8
61-80	7
81-100	7

447

448 From the table, the 0-40 and 41-60 similarity ranges took an extra hit so as to ensure that
449 more distant hits from the query sequence were considered for confidence purposes.

450 All selected sequence files were downloaded in FASTA format and used in an MSA utilizing
451 Muscle in MEGAX so as to determine the conserved regions between the different
452 metalloredutases from 30 different species with different similarities to STEAP2 (21).

453 In MEGAX, a new protein alignment was created. In phylogeny, the bootstrap method was
454 used, an LG+G+I model selected and the number of bootstraps set to 1000. The rest of the
455 settings were left as default. A dendrogram showing a phylogenetic relationship of STEAP2
456 was generated (Fig. 6 and 7).

457 **Homology modelling.**

458 Using data from MSA, we were able to identify possible similar sequences from the 80-100%
459 similarity range to the STEAP2 sequence giving direction for hits that might already have a
460 determined 3D structures and would act as templates for homology modelling.

461 **Web-based homology modelling engines**

462 *H. sapiens* STEAP2 (Uniprot accession number: Q8NFT2) FASTA sequence was used as a
463 query in the different modelling engines that were best ranked by the Continuous Automated
464 Model Evaluation (CAMEO) website and one other from the Critical Assessment of Protein
465 Structure Prediction (CASP) website (I-TASSER) to search for similar sequences for
466 homology modelling (43,44). The different homology modelling engines BLAST searched
467 the PDB for similar sequences that have approved 3D structures. The similarity of returned
468 results from the BLAST was shown in terms of coverage of the query sequence, the
469 percentage identity and the E-value. Also, the organism that was the source of the sequence
470 and the PDB identification of each result was shown. The best hit's resolution and method of
471 structural determination were obtained from the protein data bank(45).

The best sequence in all but those of RaptorX and I-TASSER were manually chosen for homology modelling. I-TASSER automatically chose the best nine sequences and made homology models of STEAP2 using each and the best ranked result (homology model 1) was chosen as the final homology model. RaptorX automatically chose its best template and the showed its homology model as a result at the end of the process.

The top templates by the rest of the engines were used to model STEAP2 using Modeller and ProMod3 (46,47). The homology models were downloaded as PDB files and visualised using PyMol and Chimera (48,49)

Model Evaluation of homology models.

The homology models from the different modelling engines were downloaded in PDB format and evaluated using different structural model evaluation tools i.e., ProSa, QMEAN, Rampage, DOPE scores and RMSD. Ramachandran Plot Server (URL: <https://zlab.umassmed.edu/bu/rama/index.pl>) was used to generate Ramachandran plots of the homology models. Using PyMol, the stoichiometry of the homology model was determined by chain coloration and critically looking at structures to identify any non-conformities such as long loops or knots (50). RMSD with the 6HCY (human STEAP4) template was calculated by use of PyMol alignment tool to align the template with the homology models to determine average structural deviations of the homology model from that of the used template.

The DOPE score of each homology model was calculated by Modeller. Z-DOPE scores were calculated for individual homology models using Modeller which was installed using Conda from commands from the site <https://anaconda.org/salilab/modeller> and it was used to calculate the normalised z-DOPE score of each homology model(47,51)

Molecular Docking

With the best model selected from 6 STEAP2 homology models after evaluation, the homology model by Swissmodel was used for flexible ligand, rigid receptor protein-ligand docking using a group of approved drug compounds downloaded from the drug bank. These compounds were downloaded as a SANCDB file of 2466.

The docking was done using AutoDock Vina. The ligand and receptor molecules were prepared using prepare_ligand4.py and prepare_receptor4.py respectively by command line on Linux terminal. The search box in the configuration file was set to dimensions x,y,z (105 by 105 by 105 Å, with a centre at: x=122.224Å, y=118.600Å, and z=102.658Å).

The results of molecular docking were characterised by 3D proximity to a central residue termed as “Euclidian distance”- which is the length of a line between two points in geometrical space which can be 2 or 3D and binding energy of the ligand to the receptor molecule. All results were screened and a binding energy threshold of less than or equal to -7 and Euclidian distance of 50Å were set. The complexes that met the selection criteria were then considered as the candidates for further analysis.

The entire protein was covered by the search box during molecular docking; a method also known as blind docking such that the docking is not biased to any region of the receptor.

Characterising Protein-Ligand Interactions

To determine the nature of the bonds formed between STEAP2 and the different ligands, Ligplot which is a program used to generate 2D representations of the nature of bonds formed between protein receptor residues and ligand atoms was used for this task (52). The program showed the hydrogen bond interactions in green dotted lines with their respective lengths in angstroms and the different interacting protein residues in an arc shape with spokes radiating towards the atoms they are in contact with. The interactions not only accounted for hydrogen bonds but all hydrophobic interactions between the two molecules (52).

519 **Docking validation using independent docking engines.**

520 The order of binding energy was validated using patchdock server and autodock tools.
 521 (29,53). The receptor pdb file and ligand pdbqt files were uploaded on patchdock server and
 522 results were downloaded via a link from the webserver. In autodock tools, the protocol in the
 523 tutorial by Huey *et al.* was followed (53)

524 **Acknowledgments.**

525 Special appreciation to the Department of Biomolecular Resources and Biolaboratory
 526 sciences for the guidance and resources used in executing this work.

527 **References**

- 528 1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer
529 statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36
530 cancers in 185 countries. *CA Cancer J Clin.* 2018 Nov;68(6):394–424.
- 531 2. World Health Organisation. Latest global cancer data: Cancer burden rises to 18.1
532 million new cases and 9.6 million cancer deaths in 2018. *Int Agency Res Cancer.*
533 2018;
- 534 3. Knaul FM, Arreola-Ornelas H, Rodriguez NM, Méndez-Carniado O, Kwete XJ,
535 Puentes-Rosas E, et al. Avoidable Mortality: The Core of the Global Cancer Divide. *J*
536 *Glob Oncol.* 2018;
- 537 4. Kingham TP, Alatisse OI, Vanderpuye V, Casper C, Abantanga FA, Kamara TB, et al.
538 Treatment of cancer in sub-Saharan Africa. *Lancet Oncol [Internet].* 2013;14(4):e158–
539 67. Available from: [http://dx.doi.org/10.1016/S1470-2045\(12\)70472-2](http://dx.doi.org/10.1016/S1470-2045(12)70472-2)
- 540 5. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Global Cancer
541 Observatory: Cancer Today. Lyon, Fr *Int Agency Res Cancer.* 2018;
- 542 6. Parkin DM, Nambooz S, Wabwire-Mangen F, Wabinga HR. Changing cancer
543 incidence in Kampala, Uganda, 1991-2006. *Int J Cancer.* 2010;126(5):1187–95.
- 544 7. Cooney MM, Okuku F, Orem J. Prostate Cancer Burden at the Uganda Cancer
545 Institute. 2019;2(4):181–5.
- 546 8. Caster JM, Falchook AD, Hendrix LH, Chen RC. Risk of pathologic upgrading or
547 locally advanced disease in early prostate cancer patients based on biopsy gleason
548 score and PSA: A population-based study of modern patients. *Int J Radiat Oncol Biol*
549 *Phys.* 2015;

- 550 9. Held-Warmkessel J. Treatment of advanced prostate cancer. *Semin Oncol Nurs*.
551 2001;17(2):118–28.
- 552 10. Nanus DM, Milowsky MI, Kostakoglu L, Smith-Jones PM, Vallabahajosula S,
553 Goldsmith SJ, et al. Clinical use of monoclonal antibody HuJ591 therapy: Targeting
554 prostate specific membrane antigen. In: *Journal of Urology*. 2003.
- 555 11. Whiteland H, Claire SS. RESEARCH PAPER A role for STEAP2 in prostate cancer
556 progression. 2014;909–20.
- 557 12. Grunewald TGP, Bach H, Cossarizza A, Matsumoto I. The STEAP protein family□:
558 Versatile oxidoreductases and targets for cancer immunotherapy with overlapping and
559 distinct cellular functions. 2012;641–57.
- 560 13. Hubert RS, Vivanco I, Chen E, Rastegar S, Leong K, Mitchell SC, et al. STEAP: A
561 prostate-specific cell-surface antigen highly expressed in human prostate tumors
562 [Internet]. 1999. Available from: www.tuat.ac.jp/mitakuadvsoisusubmit.html
- 563 14. Gomes IM, Maia CJ, Santos CR. STEAP proteins: From structure to applications in
564 cancer therapy. Vol. 10, *Molecular Cancer Research*. 2012. p. 573–87.
- 565 15. Korkmaz CG, Korkmaz KS, Kurys P, Elbi C, Wang L, Klok TI, et al. Molecular
566 cloning and characterization of STAMP2 , an androgen-regulated six transmembrane
567 protein that is overexpressed in prostate cancer. 2005;2(5):4934–45.
- 568 16. Porkka KP, Helenius MA, Visakorpi T. Cloning and characterization of a novel six-
569 transmembrane protein STEAP2, expressed in normal and malignant prostate. *Lab*
570 *Investig*. 2002 Nov 1;82(11):1573–82.
- 571 17. Abugessaisa I, Shimoji H, Sahin S, Kondo A, Harshbarger J, Lizio M, et al.
572 FANTOM5 transcriptome catalog of cellular states based on Semantic MediaWiki.

- 573 Database (Oxford). 2016;
- 574 18. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al.
575 Tissue-based map of the human proteome. *Science* (80-). 2015;
- 576 19. Uhlén M, Hallström BM, Lindskog C, Mardinoglu A, Pontén F, Nielsen J.
577 Transcriptomics resources of human tissues and organs. *Mol Syst Biol*. 2016;
- 578 20. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-
579 Tissue Expression (GTEx) project. *Nature Genetics*. 2013.
- 580 21. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary
581 Genetics Analysis across Computing Platforms. 2018;35(6):1547–9.
- 582 22. Edgar RC, Drive RM, Valley M. MUSCLE: multiple sequence alignment with high
583 accuracy and high throughput. 2004;32(5):1792–7.
- 584 23. Clamp M, Cuff J, Searle SM, Barton GJ. The Jalview Java alignment editor.
585 2004;20(3):426–7.
- 586 24. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and
587 space complexity. 2004;19:1–19.
- 588 25. Hanukoglu I. Multimedia in Biochemistry and Molecular Biology Education
589 Proteopedia: Rossmann Fold: A Beta-Alpha- Beta Fold at Dinucleotide Binding
590 Sites. *Wiley Online Libr*. 2015;206–9.
- 591 26. Murata K, Wolf M. BBA - General Subjects Cryo-electron microscopy for structural
592 analysis of dynamic biological. *BBA - Gen Subj* [Internet]. 2018;1862(2):324–34.
593 Available from: <https://doi.org/10.1016/j.bbagen.2017.07.020>
- 594 27. Rohl CA, Strauss CEM, Chivian D, Baker D. Modeling Structurally Variable Regions

- 595 in Homologous Proteins with Rosetta. *Proteins Struct Funct Genet.* 2004;55(3):656–
596 77.
- 597 28. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of
598 errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007;
- 599 29. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and
600 SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res [Internet].*
601 2005 Jul 1 [cited 2021 Jul 5];33(SUPPL. 2):W363–7. Available from:
602 https://academic.oup.com/nar/article/33/suppl_2/W363/2505698
- 603 30. Zhang C, Vasmatzis G, Cornette JL, DeLisi C. Determination of atomic desolvation
604 energies from the structures of crystallized proteins. *J Mol Biol [Internet].* 1997 Apr 4
605 [cited 2021 Jul 5];267(3):707–26. Available from:
606 <https://pubmed.ncbi.nlm.nih.gov/9126848/>
- 607 31. Wallner B, Elofsson A. All are not equal: A benchmark of different homology
608 modeling programs. *Protein Sci.* 2005;14(5):1315–27.
- 609 32. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide
610 chain configurations. *Journal of Molecular Biology.* 1963.
- 611 33. N A, K N, R N, Cancer JA. Triptorelin in the management of prostate cancer.
612 *Abeloff's Clin Oncol 5/e [Internet].* 2015;8(2):938–44. Available from:
613 <http://dx.doi.org/10.1016/B978-1-4557-2865-7.00084-9>
- 614 34. Ravenna L, Salvatori L, Morrone S, Lubrano C, Cardillo MR, Sciarra F. Effects of
615 Triptorelin , a Gonadotropin-Releasing Hormone. *Cell.* 2000;21(4):549–57.
- 616 35. Sharifi R, Soloway M, Correa RJ, Glass AG, Guinan PD, Garnick MB, et al. Clinical
617 study of leuprolide depot formulation in the treatment of advanced prostate cancer. *J*

- 618 Urol [Internet]. 1990;143(1):68–71. Available from: <http://dx.doi.org/10.1016/S0022->
619 5347(17)39868-3
- 620 36. Conn PMi, F WC. GONADOTROPIN-RELEASING HORMONE AND ITS
621 ANALOGUES. N Engl J Med [Internet]. 1991;329(14):977–86. Available from:
622 [http://content.nejm.org/cgi/content/abstract/329/14/977%5Cnhttp://www.nejm.org/doi/](http://content.nejm.org/cgi/content/abstract/329/14/977%5Cnhttp://www.nejm.org/doi/abs/10.1056/NEJM199309303291401)
623 [abs/10.1056/NEJM199309303291401](http://content.nejm.org/cgi/content/abstract/329/14/977%5Cnhttp://www.nejm.org/doi/abs/10.1056/NEJM199309303291401)
- 624 37. Heyns CF, Simonin MP, Groscurin P, Schall R, Porchet HC. Comparative efficacy of
625 triptorelin pamoate and leuprolide acetate in men with advanced prostate cancer. BJU
626 Int. 2003;92(3):226–31.
- 627 38. Kouznetsova VL, Huang D, Tsigelny IF. Potential COVID-19 protease inhibitors□:
628 Repurposing FDA- approved drugs. 2020;1–8.
- 629 39. Zouboulis CC, Piquero-Martin J. Update and future of systemic acne treatment.
630 Dermatology. 2003;206(1):37–53.
- 631 40. Longardner K, Dickson DW, Litvan I. Orthostatic hypotension preceding dementia
632 with Lewy bodies by over 15 years□: a clinicopathologic case report. Clin Auton Res
633 [Internet]. 2020;2–4. Available from: <https://doi.org/10.1007/s10286-020-00730-0>
- 634 41. Sousa SF, Ribeiro AJM, Coimbra JTS, Neves RPP, Martins SA, Moorthy NSHN, et al.
635 Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years in the
636 Field Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years
637 in the Field. Curr Med Chem. 2013;(March).
- 638 42. Bateman A, Martin MJ, O’Donovan C, Magrane M, Alpi E, Antunes R, et al. UniProt:
639 The universal protein knowledgebase. Nucleic Acids Res. 2017;
- 640 43. Haas J, Gumienny R, Barbato A, Ackermann F, Tauriello G, Bertoni M, et al.

641 Introducing “best single template” models as reference baseline for the Continuous
642 Automated Model Evaluation (CAMEO). *Proteins Struct Funct Bioinforma.* 2019;

643 44. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. Critical assessment of
644 methods of protein structure prediction (CASP)—Round XIII. *Proteins: Structure,*
645 *Function and Bioinformatics.* 2019.

646 45. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, et al.
647 The protein data bank: A computer-based archival file for macromolecular structures. *J*
648 *Mol Biol.* 1977;

649 46. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al.
650 SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic*
651 *Acids Res.* 2018;46(W1):W296–303.

652 47. Webb B, Sali A. Protein structure modeling with MODELLER. In: *Methods in*
653 *Molecular Biology.* 2017.

654 48. DeLano WL. The PyMOL Molecular Graphics System, Version 1.1. Schrödinger
655 LLC. 2002;

656 49. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al.
657 UCSF Chimera. *Journal of Computational Chemistry.* 2004.

658 50. Delano WL, Ph D. PyMOL: An Open-Source Molecular Graphics Tool. 2002;

659 51. Anaconda Inc. Anaconda Python/R Distribution - Free Download. Anaconda
660 Distribution. 2019.

661 52. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate
662 schematic diagrams of protein-ligand interactions *Clean up structure.* 1995;8(2):127–
663 34.

664 53. Huey R, Morris GM. Using AutoDock with AutoDockTools□: A Tutorial. 2005;

665

[illegible]

Data retrieval (sequences)

Data retrieval (drugs)

Phylogenetic analysis

Multiple sequence alignment

Homology modelling

Model evaluation

Protein-ligand docking

Characterise protein-ligand
interactions

Validate docking poses

A

B

- Ligand bond
- Non-ligand bond
- Hydrogen bond and its length

His 53 Non-ligand residues involved in hydrophobic contact(s)

Corresponding atoms involved in hydrophobic contact(s)

A

B

Ligand bond

Non-ligand bond

Hydrogen bond and its length

Non-ligand residues involved in hydrophobic contact(s)

Corresponding atoms involved in hydrophobic contact(s)

A

Ligand bond

Non-ligand bond

Hydrogen bond and its length

His 53 Non-ligand residues involved in hydrophobic contact(s)

Corresponding atoms involved in hydrophobic contact(s)

A

B

Ligand bond

Non-ligand bond

Hydrogen bond and its length

His 53 Non-ligand residues involved in hydrophobic contact(s)

Corresponding atoms involved in hydrophobic contact(s)

A

B

Ligand bond

Non-ligand bond

Hydrogen bond and its length

His 53 Non-ligand residues involved in hydrophobic contact(s)

Corresponding atoms involved in hydrophobic contact(s)

—|—|—|

0.50

Consensus dataset

RNA tissue specificity: Tissue enriched (prostate)

Organ

Expression

Alphabetical

Alphabetical