How good are AlphaFold models for docking-based virtual screening?

Valeria Scardino[1,2], Juan I. Di Filippo [2,3], Claudio N. Cavasotto[2,3,4,*]

[1]Meton AI, Inc., Wilmington, DE, 19801
[2]Austral Institute for Applied Artificial Intelligence, Universidad Austral, Pilar, Buenos Aires, Argentina
[3]Computational Drug Design and Biomedical Informatics Laboratory, Translational Medicine Research Institute (IIMT), Universidad Austral-CONICET, Pilar, Buenos Aires, Argentina
[4]Facultad de Ciencias Biomédicas, and Facultad de Ingeniería, Universidad Austral, Pilar, Buenos Aires, Argentina

*****Corresponding Author: CCavasotto@austral.edu.ar; cnc@cavasotto-lab.net

ABSTRACT

A crucial component in structure-based drug discovery is the availability of high-quality three-dimensional structures of the protein target. In molecular docking, whenever experimental structures were not available, *in silico* structural homology modeling has been the method of choice. However, using computationally predicted structures adds a further degree of uncertainty to the docking process. Recently, AlphaFold (AF), an artificial intelligence-based modeling tool, has shown impressive results in terms of model accuracy within the field of *ab initio* protein structure prediction. This outstanding success prompted us to evaluate how accurate AF models are from the perspective of docking-based drug discovery. We compared the performance of AF models in high-throughput docking (HTD) to their corresponding experimental PDB structures using a benchmark set of 16 targets spanning different protein families and binding site properties. Four docking programs and two consensus techniques were used to evaluate the HTD performance. The AF models showed consistently worse performance than their corresponding PDB structures, with zero enrichment factor values in several cases. While AlphaFold shows a remarkable ability to predict protein architecture and binding site anatomy, we conclude that this is not enough to guarantee that AF models can be reliably used for HTD purposes. Moreover, we show that very small variations at the side chain level of essential ligand-binding residues have a large impact on the outcome of HTD, what suggests that post-modeling refinement strategies might be key to increase the chance of success of AF models in prospective HTD campaigns.

INTRODUCTION

A crucial component in molecular docking is the availability of three-dimensional (3D) structures of the protein target. While the number of deposited structures in the PDB is continuously increasing (~194,000 in July 2022), the gap between non-redundant protein sequences and experimental structures is steadily widening. For the last 20 years, the structural genomics consortia initiatives (Levitt, 2007; Lundstrom, 2007) have been accelerating the characterization of representative structures, mainly from families poorly represented in the PDB.

Whenever experimental structures were not available, or easily obtainable, *in silico* homology modeling has been widely used to obtain a reliable 3D representation of the target (or at least, of the binding site) for docking-based drug discovery endeavors (Cavasotto, 2011). Homology modeling is a computational methodology to characterize an unknown protein structure (the target) using a related homologous protein whose experimental structure (the template) is known (Fiser, 2004). This methodology is based on the underlying assumption that proteins with similar sequences should display similar structures (Cavasotto and Phatak, 2009). The use of homology models in docking projects is already consolidated with a performance

comparable to experimental structures (Novoa et al., 2010; Spyrakis and Cavasotto, 2015; Tuccinardi, 2009; Vilar et al., 2011).

While the quality of homology models depends on several aspects, such as target-template sequence similarity, accuracy of the alignment, and the choice and resolution of the template, it is acknowledged that the post-modeling refining process is critical to obtain a reliable 3D representation of the binding site (BS) (Cavasotto et al., 2019; Kufareva et al., 2014; Kufareva et al., 2011; Michino et al., 2009). This can be understood in view of the dependence of the BS structure on the bound ligand, what highlights the importance of accounting for protein flexibility, at least at a BS level, in the homology modeling process (Bordogna et al., 2011; Phatak et al., 2010; Thomas et al., 2014). Thus, it is natural to incorporate information about existing ligands in co-modeling the binding site, such as in the ligand-steered homology method (Cavasotto et al., 2008; Phatak *et al.*, 2010), in which the six rigid coordinates of the ligand, the conformational space of the ligand torsional angles, and the BS sidechains are optimized through flexible-ligand—flexible-receptor Monte Carlo-based docking (Cavasotto and Abagyan, 2004). Similar approaches have been published, showing that refined models display an enhanced performance in high-throughput docking (HTD) (Cavasotto et al., 2005; Dalton and Jackson, 2010; Moro et al., 2006; Pala et al., 2013).

Recently, the implementation of DeepMind's artificial intelligence model, AlphaFold (AF) (Jumper et al., 2021a), set a milestone within the field of *ab initio* protein structure prediction, that is, *in silico* structural characterization just from protein sequence. The astonishing and outperforming results within the 14th Critical Assessment of protein Structure Prediction (CASP14) (Jumper et al., 2021b; Lupas et al., 2021) set AlphaFold as the breakthrough of the year by Science (doi.org/10.1126/science.acx9810) and method of the year by Nature (Method of the Year 2021: Protein structure prediction, 2022). AlphaFold predictions have gained a notorious importance; not only the structure prediction of the entire human proteome has been already carried out (Tunyasuvunakool et al., 2021), but a collaboration between DeepMind and the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) led to the creation of the AlphaFold Protein Structure Database (David et al., 2022; Varadi et al., 2022), which, at the time of writing (July 2022), contains over 200 million predicted structures. Evidently, the great excitement driven by AF is leading to a paradigm shift in the field of structural biology (Subramaniam and Kleywegt, 2022). Even the PDBsum database, which contains experimentally determined structures, has incorporated AF predictions (Laskowski and Thornton, 2022). Furthermore, developments implementing AF model predictions are emerging at a fast pace (Akdel et al., 2021; Jones and Thornton, 2022), including coupling AlphaFold with cryogenic electron microscopy maps for structure determination (Gupta et al., 2021), molecular replacement (McCoy et al., 2022; Pereira et al., 2021), NMR structural refinements (Fowler and Williamson, 2022), prediction of protein-DNA binding sites (Yuan et al., 2022), protein design (Jendrusch et al., 2021; Moffat et al., 2021), and the prediction of protein-protein interactions (Bryant et al., 2022), among others.

The public and impressive success of AF in terms of model accuracy prompted us to evaluate how accurate and useful AF models are from the perspective of docking-based drug discovery. On 16 diverse proteins we compared the performance of AF models vs PDB structures in HTD, using four different docking programs, and two different consensus ranking methods (ECR and PRC). We conclude that despite an overall very good accuracy in reproducing protein topology and modeling the binding site, HTD on AF models exhibits a consistent worse performance compared to experimental structures, with zero enrichment factors in several proteins.

RESULTS

In order to evaluate the impact of AF models in HTD, we selected a benchmark set of 16 targets of diverse protein families that we had used in a previous work (Scardino et al., 2021)

(Table 1). Due to the limitations of AF in modeling co-factors and dimerization, we chose only monomer systems containing no small-molecule co-factors; systems including water molecules were considered, but HTD on both PDB and AF structures was performed without water molecules. AF-modeled structures were obtained from the AlphaFold Protein Structure Database (Varadi *et al.*, 2022) and were compared with the corresponding crystal structures present in PDB. Four docking programs were used, AutoDock 4, ICM, rDock and PLANTS, which have different search algorithms and scoring functions. We evaluated the HTD performance of AF models using two proven effective consensus techniques, ECR (Palacio-Rodriguez et al., 2019) and PRC (Scardino *et al.*, 2021). While the ECR is a ranking-based consensus method, PRC is a combination of both ranking and docking-based consensus, which has shown a remarkable performance improvement over previous consensus methods and individual docking programs. Additionally, we docked native ligands present in crystal structures to compare with their poses on AF models.

Table 1. Target proteins used for HTD

Receptor	Receptor code	PDB	Resolution (Å)
β_2 adrenergic receptor	ADRB2	4LDO	3.2
Androgen Receptor	ANDR	2AM9	1.6
Cyclin-dependent kinase 2	CDK2	1FVV	2.8
Cyclooxygenase-1	COX1	2OYU	2.7
Estrogen receptor α	ESR1	3ERD	2.0
Fatty acid-binding protein 4	FABP4	2NNQ	1.8
Heat shock protein 90 α	HSP90a	1UYG	2.0
Insulin-like growth factor I receptor	IGF1R	2OJ9	2.0
Leukocyte Function Associated Antigen-1	LFA1	2ICA	1.6
Progesterone receptor	PRGR	3KBA	2.0
Protein kinase C β	KPCB	2I0E	2.6
Protein-tyrosine phosphatase 1B	PTN1	2AZR	2.0
Purine nucleoside phosphorylase	PNPH	3BGS	2.1
Renin	RENI	3G6Z	2.0
Tyrosine-protein kinase ABL	ABL1	2HZI	1.7
Urokinase-type plasminogen activator	UROK	1SQT	1.9

AF limitations must be carefully assessed when using models for HTD

The comparison of AF models to PDB structures is shown in Table 2. The pLDDT metric, as well as the RMSD values between backbones of the entire structure and within the BS residues are displayed. Most AF models show very good overlap to their corresponding PDB structures measured using backbone RMSD for the complete protein and within the BS (cf. columns 3-5 from Table 2). Some targets show subtle differences in certain secondary structure elements that interfere with the BS, and a few of them show structural differences that directly impede carrying out docking within the BS; for example, in RENI, where the BS in the AF structure is blocked by the N-terminal loop, which adopts a completely disordered conformation compared to their corresponding residues in the crystal structure (see Figure 1).

Table 2. Analysis of AF structural models and comparison to their corresponding experimental structures
The pLDDT metric is reported for residues within the BS, as a measure of model confidence. The RMSD values calculated at the backbone level are also displayed.

Receptor	pLDDT[a]	Backbone[b] RMSD (Å)	Backbone[c] RMSD (Å)	BS backbone RMSD (Å)	General comments
ABL1	>70 for all res in BS	1.43	0.47	0.79	The Gly-rich loop is pulled towards the BS
PNPH	H257<70	1.69	0.50	0.85	The N55:G66 loop is modeled towards the interior of the protein, near but not in contact the BS
ADRB2	H1296<70	2.53	2.06	0.81	PDB has missing residues K1232:S1262, which are included in the AF model
IGF1R	Y1161, Q1007, G1008 <70	1.84	1.29	1.64	The Gly-rich loop is in a conserved position, while the DFG loop (D1123:E1132) is more in the out position
CDK2	>70 for all res in BS	3.73	2.04	0.71	Large differences in the activation loop and C-helix
COX1	>90 for all res in BS	0.42	1.26	0.32	Mutations at 164 and 193, which have no effect on the BS; the AF model and the PDB structure lack the heme-group near BS, which does not affect docking
PRGR	>90 for all res in BS	0.61	0.52	0.47	-
ANDR	>90 for all res in BS	0.61	0.44	0.16	-
LFA1	L302, K305 and I306 <70	0.73	0.68	1.52	Helix α7 (D297:I306) is pulled inside protein
PTN1	>70 for all res in BS (>90 for all res except 2)	0.34	0.27	0.22	-
UROK	<70 for 10 res in BS (10 out of 19)	1.32	0.46	0.95	PDB has M36I mutation (far from pocket). PDB has crystal waters important for ligand binding.
FABP4	>90 for all res in BS	0.46	0.39	0.47	PDB has crystal waters important for ligand binding
KPCB	>70 for all res in BS	2.71	2.50	1.4	PDB mutation in RES 500, far from BS. Sequence difference within the C-terminal region: large difference in C622:H636, pulled towards BS
HSP90	>70 for all res in BS	9.23	4.91	4.56	High backbone RMSD. Large difference in residues N106:G137, near BS. PDB has crystal waters important for ligand binding
ESR1	H524<70	1.36	0.38	0.29	The AF Model is in the agonist-bound conformation
RENI	I77-N80 < 70	7.76	0.59	10.24	Disordered N-terminal loop, which blocks the BS and prevents using the AF structure for docking

[a]per residue Local Distance Difference Test for residues in the BS (see Methods)
[b]Considering all protein amino acids
[c]Considering only amino acids involved in secondary structure motifs

Nuclear receptors ESR1, ANDR and PRGR could be found in two structurally different biological conformations (agonist and antagonist-bound) in the PDB. In the case of ESR1, from visual inspection of the AF model, we found that helix 12 (H12) was pulled towards binding site, with a topology that corresponds best to an agonist-bound conformation. Thus, the agonist-bound PDB structure 3ERD had a more adequate backbone superposition than the corresponding antagonist-bound (PDB 3ERT), as shown in Figure 2, and therefore it was

chosen for comparison. AF models of ANDR and PRGR were also in the agonist-bound conformation.

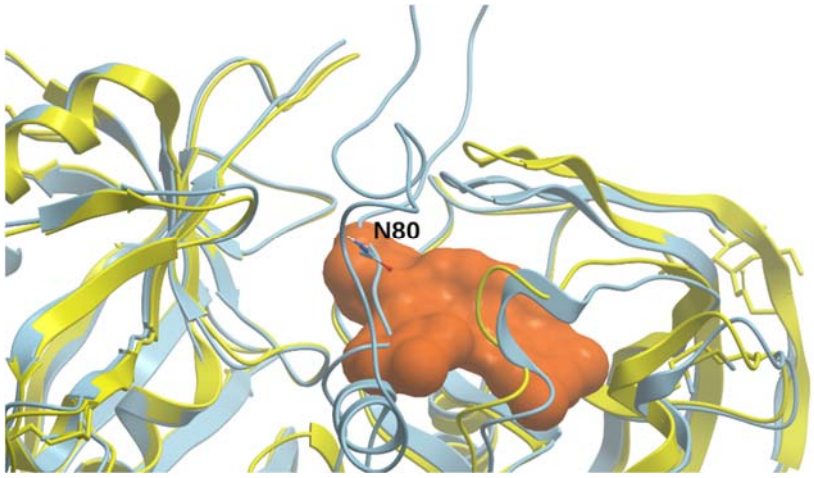

Figure 1. AF model of RENI receptor (cyan) showing an obstructed binding site
The N-terminal loop containing residue N80 is blocking the ligand-binding space. The corresponding PDB structure 3G6Z is displayed in yellow for comparison.

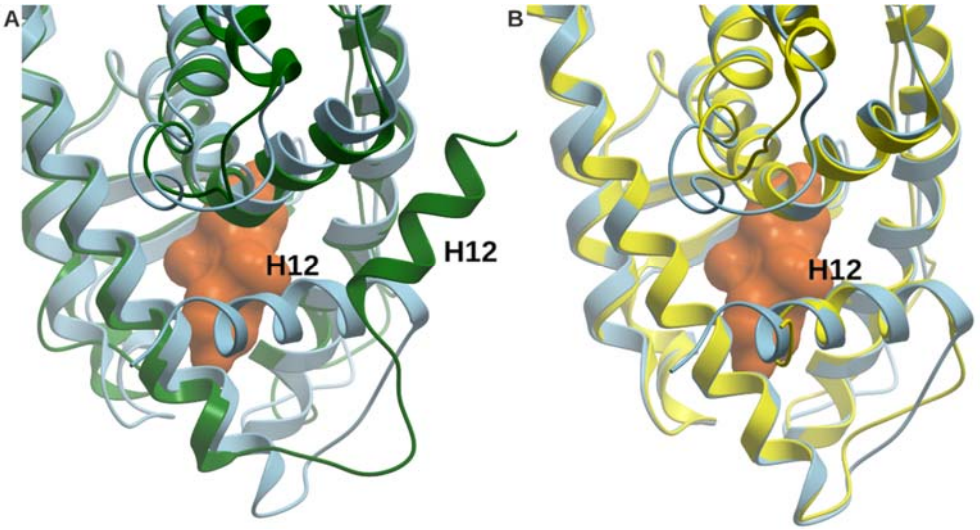

Figure 2. AF modeling of the estrogen receptor
ESR1 AF model (cyan) superimposed to the (a) antagonist-bound conformation (PDB 3ERT), and (b), and agonist-bound conformation (PDB 3ERD)

In the case of KPCB, where the AF model and the PDB structure had sequence differences in the C-terminal section, we generated the modeled structure with the available AF Colab Notebook (https://github.com/deepmind/alphafold) using the PDB 2I0E sequence as input. However, almost no differences were observed between our generated model and the AF Protein Structure Database model. In both AF structures the C-terminal loop (C622:H636) is pulled towards the inside of protein, making near contact to the BS and modifying its topology. In this case, however, since the BS is not blocked, we still used the modeled AF structure for HTD to evaluate the impact of this variation.

Protein kinases CDK2, IGFR1 and ABL1 show, on average, very good RMDS compared to the PDB structure. The AF model of CDK2 has large differences within the activation loop (containing the DFG motif) and the C-helix (compared to PDB 1FVV). In the case of ABL1, the Gly-rich loop is modeled towards to BS (compared to PDB 2HZI).

For the rest of the targets, very subtle differences were observed from the backbone superposition that are detailed in Table 2.

Small variations in the AF-modeled side chains could have a very large impact on the results obtained in molecular docking

Table 3 shows the results of HTD using AF structures. The EF at 1% (EF1) is displayed for ICM, which on average was the best performing program, and for ECR. The EF and HR results of PRC consensus method as well as the RMSD values of native ligand docking are also displayed. It can be rapidly seen that the AF models had a very low performance. On average, EF1 values of 6.5 and 5.6 were obtained with ICM and ECR respectively. The same trend is observed with PRC, where an average EF of 7.9 was obtained, with a low average HR of 0.14. Many targets had EF results less than 1.0, and even 0.0 in some cases. It should be noted that the PRC method provided better EFs on AF models than single docking programs, and the consensus ECR, what constitutes a small-scale validation of the PRC on protein models.

Table 3. Docking results using AF structural models

Receptor	ICM EF1	ECR EF1	PRC			Native ligand RMSD (Å)
			A/S*	EF	HR	
ABL1	15.5	9.5	21/65	19.5	0.32	0.66
PNPH	9.8	9.8	18/69	17.9	0.26	1.2
ADRB2	6.3	3.4	1/16	2.5	0.06	2.03
IGF1R	9.5	7.5	3/19	10.1	0.16	5.01
CDK2	4.1	6.1	3/10	10.9	0.30	8.3
COX1	1.9	1.3	4/74	2.5	0.05	>10
PRGR	9.6	7.9	36/107	18.3	0.34	0.93
ANDR	0.8	0.0	0/169	0.0	0.00	6.5
LFA1	1.5	2.9	0/14	0.0	0.00	7.7
PTN1	13.2	16.3	15/40	21.3	0.38	1.6
UROK	17.3	2.5	1/25	2.5	0.04	2.01
FABP4	0.0	0.0	0/11	0.0	0.00	5.2
KPCB	2.2	8.9	1/35	1.9	0.03	6.3
HSP90	4.6	0.0	0/32	0.0	0.00	4.5
ESR1	1.1	7.8	36/206	10.2	0.17	2.5
Average	**6.5**	**5.6**	**-**	**7.9**	**0.14**	**-**

**Active/Selected*

Table 4 shows a comparison of the results obtained in AF models vs PDB structures using the two consensus methods. It can be seen that AF models greatly worsen the HTD performance compared to their corresponding crystal structures. PRGR modeled structure was the only one that obtained similar results to the PDB. For COX1, while the ECR EF1 was low, the same was true for the PDB structure. UROK, KPCB, ANDR, FABP4 and ADRB2 show the largest ECR EF1 decrease compared to docking on PDB structures, followed by PNPH and LFA1. Consistent with this, Table 5 shows that while most PDB structures achieved very low native ligand docking RMSD values, the opposite trend was found for AF models.

Table 4. Comparison of VS results between AF models and PDB structures
Results of the two consensus methods ECR and PRC are displayed.

Receptor	ECR EF1		PRC EF		Visual inspection comments on BS comparison.
	PDB	**AF**	PDB	**AF**	
ABL1	25.3	**9.5**	26.4	**19.5**	D381 is pulled towards BS. Small difference in Gly-rich loop.
PNPH	37.1	**9.8**	34.9	**17.9**	S33 has a 2.66 Å difference in OH group.
ADRB2	24.5	**6.3**	23.4	**2.5**	Small variation in N1293 and S1203 side chains.
IGF1R	18.3	**7.5**	38.6	**10.1**	G1125 has a difference of 4 Å. DFG loop is more on an out position.
CDK2	12.8	**6.1**	16.3	**10.9**	K89 and F80 side chains slightly pulled inside BS.
COX1	3.4	**1.3**	5.8	**2.5**	F518 side chain slightly pulled inside BS.
PRGR	9.2	**7.9**	17.3	**18.3**	W755 is inverted. Difference in Q725 side chain: OH at 2.45 Å distance.
ANDR	9.0	**0.0**	13.5	**0.0**	Differences in Q711 and T877 side chains.
LFA1	10.9	**2.9**	11.6	**0.0**	Helix α7 (D297:I306) is pulled inside protein, shrinking BS.
PTN1	29.5	**16.3**	23.9	**21.3**	D48 and D181 side chains rotated towards BS.
UROK	25.9	**2.5**	47.0	**2.5**	N322, S323 and T324 are pulled towards binding site with an average backbone RMSD of 2.28 Å.
FABP4	22.1	**0.0**	26.4	**0.0**	F57 is pulled outwards BS with an RMSD of 1.6 Å.
KPCB	45.3	**8.9**	53.8	**1.9**	C-terminal residues C622:H636 are greatly pulled towards BS, modifying its topology. F353 is more in an out position.
HSP90	0.0	**0.0**	0.0	**0.0**	Big difference in structure in N106:G137, near BS. Important crystal waters missing, which might be critical for ligand binding.
ESR1	30.8	**7.8**	29.7	**10.2**	Small difference in M421 and H524 side chains, slightly pulled inside BS.

Table 5. Native Ligand RMSD comparison with PDB structures using ICM docking poses

Receptor	PDB (Å)	AF (Å)
ABL1	0.15	**0.66**
PNPH	0.59	**1.2**
ADRB2	0.35	**2.0**
IGF1R	1.06	**5.0**
CDK2	1.5	**8.3**
COX1	1.8	**>10.0**
PRGR	1.03	**0.93**
ANDR	0.17	**6.5**
LFA1	1.9	**7.7**
PTN1	0.53	**1.6**
UROK	0.24	**2.0**
FABP4	0.54	**5.2**
KPCB	1.2	**6.3**
HSP90	6.3	**4.5**
ESR1	0.2	**2.5**

While the AF models used to perform HTD exhibit an adequate backbone superposition in the BS to their corresponding PDB structures (cf. RMSD values in Table 2), some striking variations at the side chain level within the BS can be observed.

In UROK, differences can be observed at the backbone level for BS residues N143, S144 and T145, which are pulled further into the BS in the AF model with a backbone RMSD value of 2.3 Å, thus shrinking the available space for ligand binding. Moreover, deviations are also observed in side chains of Q194 and S192, as shown in Figure 3A. Regarding KPCB, the BS of the AF model is also modified at the backbone level, with residues from C-terminal region C622:H636 pulled inside the protein, interfering with the BS. As expected, this had a huge

impact on HTD results. For ANDR, variations can be noticed in Q711 and T877 side chains, shown in Figure 3B. While for Q711 it was shown by Pereira de Jesús et al. (Pereira de Jésus-Tran et al., 2006) that it can appear in both conformations, T877 is essential for ligand binding, making important interactions with the native ligand in the crystallized PDB structure. In HSP90, a very poor performance was obtained, using both the AF model and the PDB structure without crystallized waters. It should be noted that the PDB structure with waters had a PRC EF of 15.4 in a previous work (Scardino *et al.*, 2021), which shows how critical it is to include them for HTD.

In the case of FABP4, although most of the side chains are correctly modeled, F57 is pulled further back, thus opening more space within the BS. This residue participates in important hydrophobic interactions with the native ligand in the PDB. For PNPH, almost only one significant difference is found in the OH group from S33, which is pulled 2.7 Å further into the BS in the AF model, as shown in Figure 3C. This might be critical as serine residues are often involved in important interactions for ligand binding. Figure 3D shows LFA1 BS where it can be observed a notable difference at the backbone level in helix α7 containing residues L302:I306. This helix is pulled inside the BS in the AF model, thus modifying the space available for ligand-binding. Small variations in the side chains of residues E284 and K287 are also observed.

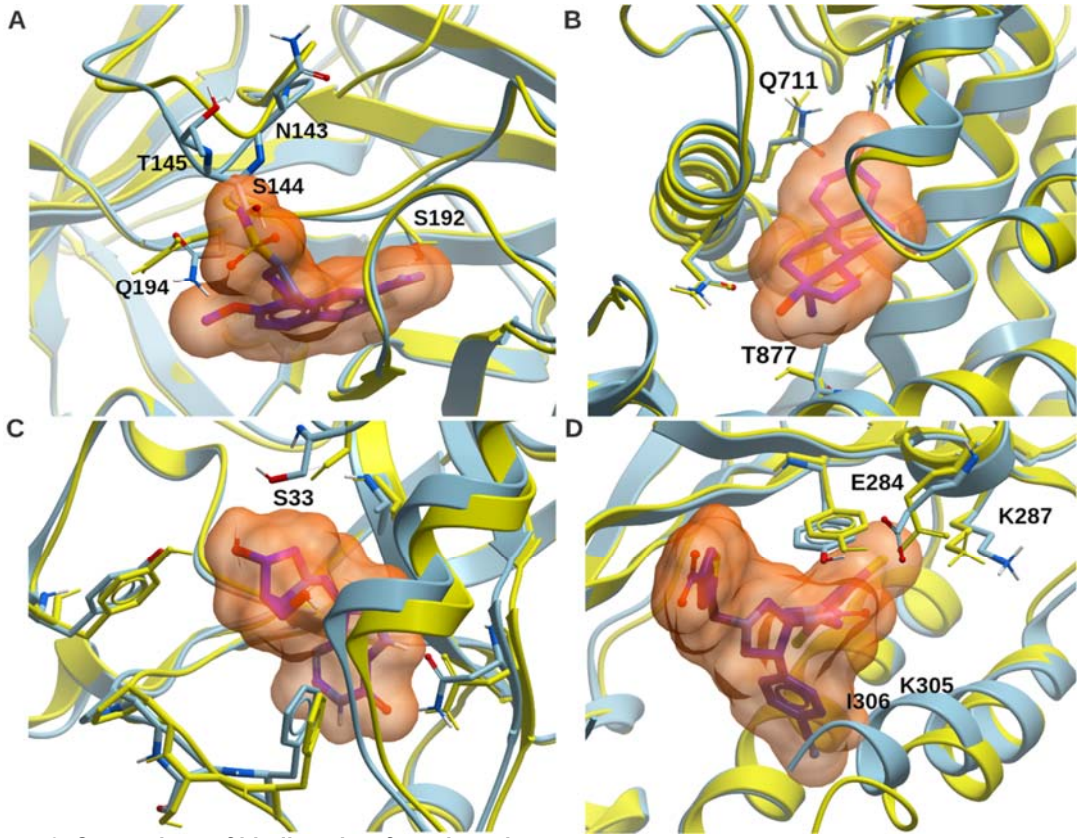

Figure 3. Comparison of binding sites for selected targets
AF models are displayed in cyan and PDB structures in yellow. Native ligands are displayed in stick representation, and the binding sites represented with orange surfaces. (a) UROK binding site: differences in backbone can be observed for N143:T145. (b) ANDR binding site: small variation in T877 side chain can be observed, which makes important interactions for ligand-binding. (c) PNPH binding site: the most notable difference can be seen in S33 side chain. (d) LFA1 binding site: backbone differences in the helix containing K305, and small variations in the side chains of E284 and K287 are observed.

It can be seen from this analysis that small changes at the side chain level of essential ligand-binding residues have a very large impact on the EFs obtained from HTD campaigns, and on the docking of native ligand structures. However, this impact could not have been expected in advance by looking at the backbone RMSD nor at the pLDDT metric, since overall, those were acceptable. For three out of the four AF models that worsened the HTD performance the most, the pLDDT metric is equal to or greater than 70 for every residue in the BS (cf. Column 1 in Table 2), indicating high confidence in these modeled structures.

DISCUSSION

As it can be seen in Table 4, HTD on AF models shows consistently lower EF values assessed with two consensus methods (ECR and PRC) when compared to the HTD on the corresponding PDB structures, also complemented with poor native ligand RMSD values (cf. Table 5); in several cases, the EF on AF models is even zero. Results also deteriorated for each individual docking program. From Tables 2 and 4, it can be inferred that these poor EF values could be due by: i) large differences at the backbone level within the binding site (as in RENI, where even no docking could be performed due to the distortion of the BS); ii) small variations either at the backbone level (UROK, for example), or at the side chain level (ANDR, for example). In several cases, even very subtle differences within the BS could have a huge impact in the EF, such as in ANDR and FABP4. In agreement with what has been shown by others (Jumper *et al.*, 2021a; Jumper *et al.*, 2021b; Stevens and He, 2022), the AF models exhibit low backbone RMSD values compared to PDB structures, thus demonstrating the remarkable ability of AlphaFold to predict protein architecture from sequence alone; moreover, from Table 2, it can be readily seen that our models also show low backbone RMSD and good pLDDT values within the BS. Therefore, we must conclude that the accuracy of AlphaFold in reproducing protein topology and BS anatomy with very good values of the pLDDT metric, is not enough to guarantee that AF models can be reliably used for molecular docking purposes. Thus, crude AF models do not seem to be suitable for HTD without performing post-modeling refinement techniques (Cavasotto *et al.*, 2019).

It should be also highlighted that the single structural model provided by AF from a given sequence cannot represent: i) different biological states of the proteins (such as agonist- and antagonist-bound conformations, as in the case of GPCRs and nuclear receptors, or open vs closed, as in channels); ii) protein dynamics (such as different conformations of the Gly-rich, catalytic and activation loops in protein kinases); iii) structural conformational differences – especially within the BS- associated with ligand binding. In fact, it has been highlighted that modeling a receptor not in the desired biological state is one of the current main limitations of AF (Schauperl and Denny, 2022); while it is probable that the AF model corresponds to the state that is most represented in the training set, an intermediate state conformation could also be observed (Schauperl and Denny, 2022). It should be thus acknowledged that, even with their limitations, different structures of the same protein available in the PDB might indeed sample structural diversity to a certain degree,which right now is not available for AF models.

In this contribution, we compared the AF models to their best PDB match in terms of backbone RMSD. However, in a real-world prospective case, biological and biochemical knowledge should be taken into consideration at the modeling stage in order to ensure that the modeled structure is in the desired biological conformation. It should be noted that this issue is many times avoided in homology modeling, since the structural template from the PDB is chosen taking into consideration the sought biological state of the target (Cavasotto and Phatak, 2009); for example, for modeling a given GPCR in the agonist bound conformation, the templates from the PDB are selected among those exhibiting an agonist-bound conformation (Cavasotto and Palomba, 2015). It should also be noted that efforts extending the use of AlphaFold to predict both active and inactive states of a protein target have been reported recently (Heo and Feig, 2022).

Regarding AlphaFold limitations, which have been discussed elsewhere (Akdel *et al.*, 2021; Jones and Thornton, 2022; Laskowski and Thornton, 2022; Schauperl and Denny, 2022), it should be acknowledged that, from a structure-based drug discovery perspective, AF provides an incomplete structural model due to the lack of water molecules, metal ions, and co-factors. Just to illustrate, in HSP90 a very poor performance was obtained using both the AF model and the PDB structure omitting crystallized waters (cf. Table 4); however, using the PDB structure with waters, we previously obtained a PRC EF of 15.4 (Scardino *et al.*, 2021); the ligand RMSD values with and with no water molecules (Table 5) were 0.8 Å and 6.3 Å, respectively, which highlights the importance of including water molecules for HTD in some targets. As routinely done with PDB structures, AF models should be also carefully checked for correct histidine tautomers, asparagine and glutamine flipping, protonation states (especially acidic residues, histidine, and cysteine eventually involved in metal binding), and polar hydrogens conformation.

From a practical point of view and provided the AF model is in the desired biological state, a co-refinement of the BS together with known ligands (whenever available) in a ligand-steered fashion (Phatak *et al.*, 2010) might be the best strategy to sample BS conformational diversity and maximize the chances of success in a prospective HTD endeavor.

Although the analysis of this study has been focused on the regions of AlphaFold models that superimpose with the crystalized domains of their corresponding PDB structure, it is worth mentioning that, in some cases, the regions that were cut from the AF models seem to exhibit, by simple visual inspection, a high degree of disorder. As expected, these *a priori* disordered regions present low values of pLDDT, but the notorious contrast of the perceived model quality in matching and non-matching regions results striking. Even though low pLDDT regions (pLDDT<50) were suggested to have a high likelihood of being unstructured in isolation, or only structured as part of a complex (Tunyasuvunakool *et al.*, 2021), this issue should be further analyzed.

Our conclusions will help to understand the current limitations of AlphaFold models in HTD, and from this knowledge to develop strategies to circumvent its drawbacks and thus enhance its further application in drug discovery.

METHODS

Target preparation

The 16 protein targets used in this study (Table 1) were downloaded from the PDB. Water molecules and co-factors were deleted in all of them. For each target, an AF model was retrieved from the AlphaFold Protein Structure Database (Varadi *et al.*, 2022) using the corresponding Uniprot identification. An additional AlphaFold structure was utilized for KPCB, which was generated using a slightly simplified version of AF which is publicly available (https://github.com/deepmind/alphafold). In every case, AF models were cut to match their corresponding crystalized domains present in the PDB.

Both PDB structures and AF models were prepared in the same way using the ICM program (Abagyan et al., 1994) (version 3.9-2e; MolSoft, San Diego, CA, May 2022), in a similar fashion as in earlier works (Cavasotto and Aucar, 2020; Scardino *et al.*, 2021). Missing amino acids and hydrogen atoms were added to PDB structures; local energy minimization was performed both on PDB structures and AF models. Polar hydrogens within the binding site were optimized using a Monte Carlo sampling in the dihedral space. Glutamate and aspartate residues were assigned a -1 charge, and lysine and arginine were assigned a +1 charge. For PDB structures, asparagine and glutamine residues were inspected for flipping and corrected whenever, and His tautomers were assigned according to their hydrogen bonding network.

Protein metrics

For comparison with PDB structures, AF models were superimposed to them using backbone atoms (C, C_α, N) considering: i) the complete protein; ii) residues which participate in defined secondary structure elements (α-, π- or 3.10 helices, or β-sheets) (cf. Table 2). RMSD values between backbones were calculated for the whole structure and for the BS residues, which were determined according to their distance to the native ligand in the PDB structures: if a heavy atom is within 4.0 Å of any heavy atom in the ligand, that residue is considered a BS residue. The predicted Local Distance Difference Test (pLDDT) is a per residue metric reported in the AlphaFold Protein Structure Database (Varadi *et al.*, 2022) as an estimate of model confidence on a scale from 0 to 100; the LDDT is a superposition-free score that evaluates local distance differences of all atoms in a model and includes validation of stereochemical plausibility (Mariani et al., 2013). Following this evaluation criterion, we looked at the pLDDT metric especially for BS residues.

Docking libraries

For each target, the corresponding docking chemical libraries consist of a set of active molecules and their corresponding matching decoys according to similar physico-chemical properties and structural dissimilarity, which has been shown to ensure unbiased calculations in docking simulations (Gatica and Cavasotto, 2012; Huang et al., 2006). For all molecules, chirality and protonation states were inherited from the corresponding original databases. Libraries were obtained from the DUD-E database (Mysinger et al., 2012), except for the ESR1 agonists library which was obtained from NRLiSt (Lagarde et al., 2014) database, and the ADRB2 library which was taken from GLL/GDD (Gatica and Cavasotto, 2012). The number of molecules present varies from ~2,200 in CDK2 to ~23,000 in ESR1.

Docking methods

Four docking programs were used in total: ICM (Abagyan *et al.*, 1994), Auto Dock 4 (Morris et al., 2009), rDock (Ruiz-Carmona et al., 2014) and PLANTS (Korb et al., 2009). These programs have different search algorithms and scoring functions as described in previous works (Palacio-Rodriguez *et al.*, 2019; Scardino *et al.*, 2021). Auto Dock Tools utilities (Morris *et al.*, 2009) were used to prepare the input files for Auto Dock 4. For all the HTD runs, the parameters of the programs were set in the same way as in a previous work (Scardino *et al.*, 2021), what allowed direct comparison of AF docking results with earlier calculations. Only when needed, docking boxes on AF models were slightly modified to be accommodated due to small differences in binding sites.

Consensus methods

Two consensus methods were used to combine the results of the docking programs. The Exponential Consensus Ranking (ECR) (Palacio-Rodriguez *et al.*, 2019) combines the ranks of each molecule determined using different scoring functions with an exponential distribution, calculated as

$$ECR(i) = \frac{1}{\sigma}\sum_{j} \exp\left[-\frac{r_j(i)}{\sigma}\right]$$

where $r_j(i)$ is the rank of molecule *i* idetermined using the scoring function of program *j*, and σ is the expected value of the exponential distribution; the ECR was found to be quasi-independent on σ, and we used σ = 10% of the total number of molecules for each docking library. Further details can be found elsewhere (Palacio-Rodriguez *et al.*, 2019; Scardino *et al.*, 2021).

The Pose/Ranking Consensus method (PRC) (Scardino *et al.*, 2021) consists of a hybrid consensus technique that combines ranks and docking poses obtained with different docking programs and selects the molecules that meet the following criteria: if a molecule has a maximum of two matching poses, the corresponding ranks should be within the top 5% of the

corresponding docking programs; with a maximum of three matching poses, those corresponding three ranks should be within the top 10%, and with four matching poses, the four ranks ought to be in the top 20%. Finally, only the molecules that are also in the top 1.5% of ECR are selected. It was shown that this subset of molecules increases the chance of finding real hits, measured through the Enrichment Factor (*EF*) and the hit rate (*HR*).

The *EF* is defined as

$$EF(x) = \frac{Hits_x}{N_x} / \frac{Hits_{total}}{N_{total}}$$

where $Hits_x$ represents the number of actives present in a subset *x* of the docked library, N_x the number of molecules in subset *x*, $Hits_{total}$ is the total number of ligands within the entire chemical library, and N_{total} its total number of molecules. When subset x is a percentage of the total number of molecules, for example the top 1%, we call it the *EF* at 1% (*EF1*).
The hit rate (*HR*) is calculated as

$$HR(x) = \frac{Hits_x}{N_x}$$

and is a measure between 0 and 1 which represents the probability of finding an actual ligand within the subset *x*.

ACKNOWLEDGEMENTS
CNC thanks Molsoft LLC (San Diego, CA) for providing an academic license for the ICM program. The authors thank the Centro de Cálculo de Alto Desempeño (Universidad Nacional de Córdoba) for granting the use of their computational resources.

DECLARATION OF INTERESTS
The authors declare no competing interests

REFERENCES
Abagyan, R., Totrov, M., and Kuznetsov, D. (1994). ICM - a New Method For Protein Modeling and Design - Applications to Docking and Structure Prediction From the Distorted Native Conformation. J. Comput. Chem. *15*, 488-506.

Akdel, M., Pires, D.E.V., Porta Pardo, E., Jänes, J., Zalevsky, A.O., Mészáros, B., Bryant, P., Good, L.L., Laskowski, R.A., Pozzati, G., et al. (2021). A structural biology community assessment of AlphaFold 2 applications. bioRxiv, 2021.2009.2026.461876. 10.1101/2021.09.26.461876.

Bordogna, A., Pandini, A., and Bonati, L. (2011). Predicting the accuracy of protein-ligand docking on homology models. J. Comput. Chem. *32*, 81-98. 10.1002/jcc.21601.

Bryant, P., Pozzati, G., and Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. Nat. Commun. *13*, 1265. 10.1038/s41467-022-28865-w.

Cavasotto, C.N. (2011). Homology models in docking and high-throughput docking. Curr. Top. Med. Chem. *11*, 1528-1534. 10.2174/156802611795860951.

Cavasotto, C.N., and Abagyan, R.A. (2004). Protein flexibility in ligand docking and virtual screening to protein kinases. J. Mol. Biol. *337*, 209-225.

Cavasotto, C.N., and Aucar, M.G. (2020). High-Throughput Docking Using Quantum Mechanical Scoring. Front. Chem. *8*, 246. 10.3389/fchem.2020.00246.

Cavasotto, C.N., Aucar, M.G., and Adler, N.S. (2019). Computational chemistry in drug lead discovery and design. Int. J. Quantum Chem. *119*, e25678. https://doi.org/10.1002/qua.25678.

Cavasotto, C.N., Kovacs, J.A., and Abagyan, R.A. (2005). Representing Receptor Flexibility in Ligand Docking through Relevant Normal Modes. J. Am. Chem. Soc. *127*, 9632-9640. 10.1021/ja042260c.

Cavasotto, C.N., Orry, A.J., Murgolo, N.J., Czarniecki, M.F., Kocsi, S.A., Hawes, B.E., O'Neill, K.A., Hine, H., Burton, M.S., Voigt, J.H., et al. (2008). Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. J. Med. Chem. *51*, 581-588.

Cavasotto, C.N., and Palomba, D. (2015). Expanding the horizons of G protein-coupled receptor structure-based ligand discovery and optimization using homology models. Chem. Commun. (Cambridge, U. K.) *51*, 13576-13594. 10.1039/c5cc05050b.

Cavasotto, C.N., and Phatak, S.S. (2009). Homology modeling in drug discovery: current trends and applications. Drug Discov. Today *14*, 676-683.

Dalton, J.A., and Jackson, R.M. (2010). Homology-modelling protein-ligand interactions: allowing for ligand-induced conformational change. J. Mol. Biol. *399*, 645-661. 10.1016/j.jmb.2010.04.047.

David, A., Islam, S., Tankhilevich, E., and Sternberg, M.J.E. (2022). The AlphaFold Database of Protein Structures: A Biologist's Guide. J. Mol. Biol. *434*, 167336. 10.1016/j.jmb.2021.167336.

Fiser, A. (2004). Protein structure modeling in the proteomics era. Expert Rev. Proteomics *1*, 97-110. 10.1586/14789450.1.1.97.

Fowler, N.J., and Williamson, M.P. (2022). The accuracy of protein structures in solution determined by AlphaFold and NMR. Structure *30*, 925-933 e922. 10.1016/j.str.2022.04.005.

Gatica, E.A., and Cavasotto, C.N. (2012). Ligand and Decoy Sets for Docking to G Protein-Coupled Receptors. J. Chem. Inf. Model. *52*, 1-6. 10.1021/ci200412p.

Gupta, M., Azumaya, C.M., Moritz, M., Pourmal, S., Diallo, A., Merz, G.E., Jang, G., Bouhaddou, M., Fossati, A., Brilot, A.F., et al. (2021). CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. bioRxiv, 2021.2005.2010.443524. 10.1101/2021.05.10.443524.

Heo, L., and Feig, M. (2022). Multi-state modeling of G-protein coupled receptors at experimental accuracy. Proteins. 10.1002/prot.26382.

Huang, N., Shoichet, B.K., and Irwin, J.J. (2006). Benchmarking sets for molecular docking. J. Med. Chem. *49*, 6789-6801. 10.1021/jm0608356.

Jendrusch, M., Korbel, J.O., and Sadiq, S.K. (2021). AlphaDesign: A de novo protein design framework based on AlphaFold. bioRxiv, 2021.2010.2011.463937. 10.1101/2021.10.11.463937.

Jones, D.T., and Thornton, J.M. (2022). The impact of AlphaFold2 one year on. Nat. Methods *19*, 15-20. 10.1038/s41592-021-01365-3.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021a). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589. 10.1038/s41586-021-03819-2.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021b). Applying and improving AlphaFold at CASP14. Proteins *89*, 1711-1721. 10.1002/prot.26257.

Korb, O., Stutzle, T., and Exner, T.E. (2009). Empirical scoring functions for advanced protein-ligand docking with PLANTS. J. Chem. Inf. Model. *49*, 84-96. 10.1021/ci800298z.

Kufareva, I., Katritch, V., Stevens, R.C., and Abagyan, R. (2014). Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. Structure *22*, 1120-1139. 10.1016/j.str.2014.06.012.

Kufareva, I., Rueda, M., Katritch, V., Stevens, R.C., and Abagyan, R. (2011). Status of GPCR Modeling and Docking as Reflected by Community-wide GPCR Dock 2010 Assessment. Structure *19*, 1108-1126. 10.1016/j.str.2011.05.012.

Lagarde, N., Ben Nasr, N., Jeremie, A., Guillemain, H., Laville, V., Labib, T., Zagury, J.F., and Montes, M. (2014). NRLiSt BDB, the manually curated nuclear receptors ligands and structures benchmarking database. J. Med. Chem. *57*, 3117-3125. 10.1021/jm500132p.

Laskowski, R.A., and Thornton, J.M. (2022). PDBsum extras: SARS-CoV-2 and AlphaFold models. Protein Sci. *31*, 283-289. 10.1002/pro.4238.

Levitt, M. (2007). Growth of novel protein structural data. Proc. Natl. Acad. Sci. U.S.A. *104*, 3183-3188. 10.1073/pnas.0611678104.

Lundstrom, K. (2007). Structural genomics and drug discovery. J. Cell. Mol. Med. *11*, 224-238. 10.1111/j.1582-4934.2007.00028.x.

Lupas, A.N., Pereira, J., Alva, V., Merino, F., Coles, M., and Hartmann, M.D. (2021). The breakthrough in protein structure prediction. Biochem. J. *478*, 1885-1890. 10.1042/BCJ20200963.

Mariani, V., Biasini, M., Barbato, A., and Schwede, T. (2013). lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics *29*, 2722-2728. 10.1093/bioinformatics/btt473.

McCoy, A.J., Sammito, M.D., and Read, R.J. (2022). Implications of AlphaFold2 for crystallographic phasing by molecular replacement. Acta Crystallographica Section D *78*, 1-13. doi:10.1107/S2059798321012122.

Method of the Year 2021: Protein structure prediction. (2022). Nature Methods *19*, 1-1. 10.1038/s41592-021-01380-4.

Michino, M., Abola, E., Brooks, C.L., 3rd, Dixon, J.S., Moult, J., and Stevens, R.C. (2009). Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. Nat. Rev. Drug Discov. *8*, 455-463. 10.1038/nrd2877.

Moffat, L., Greener, J.G., and Jones, D.T. (2021). Using AlphaFold for Rapid and Accurate Fixed Backbone Protein Design. bioRxiv, 2021.2008.2024.457549. 10.1101/2021.08.24.457549.

Moro, S., Deflorian, F., Bacilieri, M., and Spalluto, G. (2006). Ligand-based homology modeling as attractive tool to inspect GPCR structural plasticity. Curr. Pharm. Des. *12*, 2175-2185.

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., and Olson, A.J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J. Comput. Chem. *30*, 2785-2791. 10.1002/jcc.21256.

Mysinger, M.M., Carchia, M., Irwin, J.J., and Shoichet, B.K. (2012). Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. J. Med. Chem. *55*, 6582-6594. 10.1021/jm300687e.

Novoa, E.M., Ribas de Pouplana, L., Barril, X., and Orozco, M. (2010). Ensemble docking from homology models. J. Chem. Theory Comput. *6*, 2547–2557.

Pala, D., Beuming, T., Sherman, W., Lodola, A., Rivara, S., and Mor, M. (2013). Structure-based virtual screening of MT2 melatonin receptor: influence of template choice and structural refinement. J. Chem. Inf. Model. *53*, 821-835. 10.1021/ci4000147.

Palacio-Rodriguez, K., Lans, I., Cavasotto, C.N., and Cossio, P. (2019). Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. Sci. Rep. *9*, 5142. 10.1038/s41598-019-41594-3.

Pereira de Jésus-Tran, K., Côté, P.-L., Cantin, L., Blanchet, J., Labrie, F., and Breton, R. (2006). Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. Protein Sci. *15*, 987-999. https://doi.org/10.1110/ps.051905906.

Pereira, J., Simpkin, A.J., Hartmann, M.D., Rigden, D.J., Keegan, R.M., and Lupas, A.N. (2021). High-accuracy protein structure prediction in CASP14. Proteins *89*, 1687-1699. 10.1002/prot.26171.

Phatak, S.S., Gatica, E.A., and Cavasotto, C.N. (2010). Ligand-steered modeling and docking: A benchmarking study in Class A G-Protein-Coupled Receptors. J. Chem. Inf. Model. *50*, 2119-2128. 10.1021/ci100285f.

Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., Garmendia-Doval, A.B., Juhos, S., Schmidtke, P., Barril, X., Hubbard, R.E., and Morley, S.D. (2014). rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. PLoS Comput. Biol. *10*, e1003571. 10.1371/journal.pcbi.1003571.

Scardino, V., Bollini, M., and Cavasotto, C.N. (2021). Combination of pose and rank consensus in docking-based virtual screening: the best of both worlds. RSC Advances *11*, 35383-35391. 10.1039/d1ra05785e.

Schauperl, M., and Denny, R.A. (2022). AI-Based Protein Structure Prediction in Drug Discovery: Impacts and Challenges. J. Chem. Inf. Model. *62*, 3142-3156. 10.1021/acs.jcim.2c00026.

Spyrakis, F., and Cavasotto, C.N. (2015). Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description. Arch. Biochem. Biophys. *583*, 105-119. 10.1016/j.abb.2015.08.002.

Stevens, A.O., and He, Y. (2022). Benchmarking the Accuracy of AlphaFold 2 in Loop Structure Prediction. Biomolecules *12*. 10.3390/biom12070985.

Subramaniam, S., and Kleywegt, G.J. (2022). A paradigm shift in structural biology. Nature Methods *19*, 20-23. 10.1038/s41592-021-01361-7.

Thomas, T., McLean, K.C., McRobb, F.M., Manallack, D.T., Chalmers, D.K., and Yuriev, E. (2014). Homology modeling of human muscarinic acetylcholine receptors. J. Chem. Inf. Model. *54*, 243-253. 10.1021/ci400502u.

Tuccinardi, T. (2009). Docking-based virtual screening: recent developments. Comb Chem High Throughput Screen *12*, 303-314.

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Zidek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. Nature *596*, 590-596. 10.1038/s41586-021-03828-1.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. *50*, D439-D444. 10.1093/nar/gkab1061.

Vilar, S., Ferino, G., Phatak, S.S., Berk, B., Cavasotto, C.N., and Costanzi, S. (2011). Docking-based virtual screening for ligands of G protein-coupled receptors: not only crystal structures but also in silico models. J. Mol. Graph. Model. *29*, 614-623. 10.1016/j.jmgm.2010.11.005.

Yuan, Q., Chen, S., Rao, J., Zheng, S., Zhao, H., and Yang, Y. (2022). AlphaFold2-aware protein–DNA binding site prediction using graph transformer. Brief. Bioinform. *23*. 10.1093/bib/bbab564.