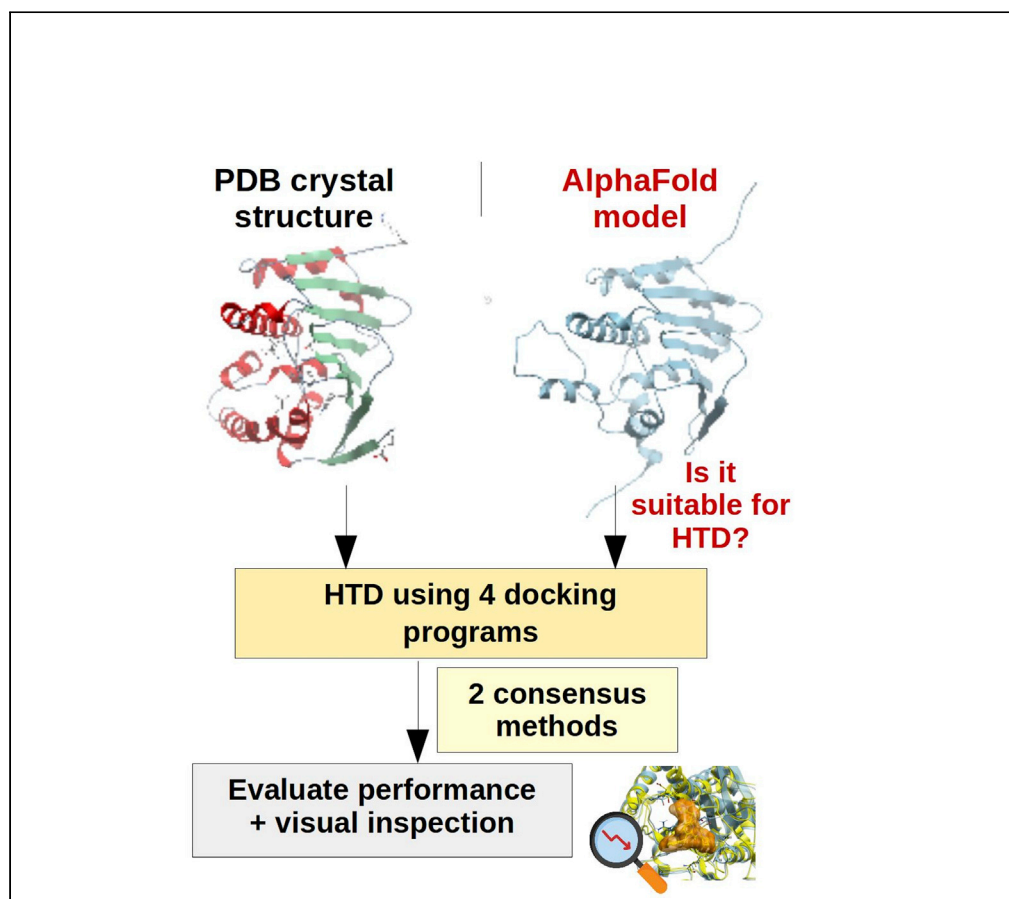**Article**

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

Valeria Scardino, Juan I. Di Filippo, Claudio N. Cavasotto

ccavasotto@austral.edu.ar

Highlights

Well-known AF models are evaluated for their HTD capability using 4 docking programs

The performance of as-is AF models is significantly lower compared with PDB structures

Even on very accurate models, small side-chain variations impact the performance

A refinement of AF models might be crucial to maximize the chances of success in HTD

Article

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

Valeria Scardino,[1,2,5] Juan I. Di Filippo,[2,3,5] and Claudio N. Cavasotto[2,3,4,6,*]

SUMMARY

A crucial component in structure-based drug discovery is the availability of high-quality three-dimensional structures of the protein target. Whenever experimental structures were not available, homology modeling has been, so far, the method of choice. Recently, AlphaFold (AF), an artificial-intelligence-based protein structure prediction method, has shown impressive results in terms of model accuracy. This outstanding success prompted us to evaluate how accurate AF models are from the perspective of docking-based drug discovery. We compared the high-throughput docking (HTD) performance of AF models with their corresponding experimental PDB structures using a benchmark set of 22 targets. The AF models showed consistently worse performance using four docking programs and two consensus techniques. Although AlphaFold shows a remarkable ability to predict protein architecture, this might not be enough to guarantee that AF models can be reliably used for HTD, and post-modeling refinement strategies might be key to increase the chances of success.

INTRODUCTION

A crucial component in molecular docking is the availability of three-dimensional (3D) structures of the protein target. Although the number of deposited structures in the PDB[1] is continuously increasing (~199,000 in November 2022), the gap between non-redundant protein sequences and experimental structures is steadily widening. For the last 20 years, the structural genomics consortia initiatives[2,3] have been accelerating the characterization of representative protein 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.[4] 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.[5] This methodology is based on the underlying assumption that proteins with similar sequences should display similar structures.[6] The use of homology models in docking projects is already consolidated with a performance comparable to experimental structures.[7–10]

Although 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).[11–14] This can be understood in view of the dependence of the binding site structure on the bound ligand, what highlights the importance of accounting for protein flexibility, at least at a binding site level, in the homology modeling process.[15–17] Thus, it is natural to incorporate information about existing ligands in co-modeling the binding site, such as in the ligand-steered homology method,[16,18] in which the six rigid coordinates of the ligand, the conformational space of the ligand torsional angles, and the binding site sidechains are optimized through flexible-ligand—flexible-receptor Monte-Carlo-based docking.[19] Similar approaches have been published, showing that refined models display an enhanced performance in high-throughput docking (HTD).[20–23]

Recently, the implementation of DeepMind's artificial intelligence model, AlphaFold (AF),[24] set a milestone within the field of protein structure prediction. The astonishing and outperforming results within the 14th Critical Assessment of protein Structure Prediction (CASP14)[25,26] set AlphaFold as the breakthrough of the

[1]Meton AI, Inc, Wilmington, DE 19801, USA

[2]Austral Institute for Applied Artificial Intelligence, Universidad Austral, Pilar, Buenos Aires, Argentina

[3]Computational Drug Design and Biomedical Informatics Laboratory, Instituto de Investigaciones en Medicina Traslacional (IIMT), Universidad Austral-CONICET, Pilar, Buenos Aires, Argentina

[4]Facultad de Ciencias Biomédicas, and Facultad de Ingeniería, Universidad Austral, Pilar, Buenos Aires, Argentina

[5]These authors contributed equally

[6]Lead contact

*Correspondence: ccavasotto@austral.edu.ar

https://doi.org/10.1016/j.isci.2022.105920

year by Science (doi.org/10.1126/science.acx9810) and method of the year by Nature.[27] AlphaFold predictions have gained a notorious importance; not only the structure prediction of the entire human proteome has been already carried out[28] 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,[29,30] which, at the time of writing (November 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.[31] Even the PDB, which contains experimentally determined structures, has incorporated AF predictions.[32] Furthermore, not only different implementations of AF with specific refinements are being actively developed[33,34] but also developments implementing AF model predictions are emerging at a fast pace,[35,36] including coupling AlphaFold with cryogenic electron microscopy maps for structure determination,[37] molecular replacement,[38,39] NMR structural refinements,[40] prediction of protein-DNA binding sites,[41] protein design,[42,43] and the prediction of protein-protein interactions,[44] among others.

Remarkably, "AlphaFold is trained to predict the structure of proteins *as they might appear in the PDB*" (https://alphafold.ebi.ac.uk/faq); moreover, "backbone and side chain coordinates are frequently consistent with the expected structure in the presence of ions (e.g., for zinc-binding sites) or co-factors (e.g., side chain geometry consistent with heme binding)" (https://alphafold.ebi.ac.uk/faq). These facts, and the public and impressive success of AF in terms of overall model accuracy, prompted us to evaluate how accurate and useful as-is AF models are in the context of docking-based drug discovery, as an alternative to using PDB structures. On 22 diverse proteins we compared the performance of AF models (extracted from the AlphaFold Protein Structure Database) versus PDB structures in HTD. We conclude that despite an overall very good accuracy in reproducing protein topology and the binding site, HTD on AF models exhibits a consistent worse performance compared with experimental structures, with zero enrichment factors in several proteins.

RESULTS

We selected a benchmark set of 22 targets from diverse protein families used in an earlier work[45] (Table 1). Considering what has been said earlier of AF models in terms of their representativity of ligand-bound complexes, to evaluate the performance of as-is AF models in HTD we chose to compare with holo PDB structures. Because AlphaFold does not predict the positions of co-factors, metals, ligands, ions, or water molecules, to compare structures on an equal standing, we stripped PDB structures from water molecules, ions, co-factors, etc.; we also avoided any co-refinement of the PDB structure with the native or other ligands, what would also have enhanced the outcome. AF-modeled structures were obtained from the AlphaFold Protein Structure Database.[30] 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[46] and PRC.[45] Although 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. In addition, we docked native ligands present in crystal structures to compare with their poses on AF models.

The topology of AF models is analyzed to assess whether they are suitable 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 binding site residues are displayed. Most AF models show very good overlap to their corresponding PDB structures measured using backbone RMSD for the complete protein and also for binding site residues (cf. columns 3–5 from Table 2). Some targets show subtle differences in certain secondary structure elements that interfere with the binding site, and a few of them show structural differences that directly impede carrying out docking within the binding site; for example, in RENI, where the pocket in the AF structure is blocked by the N-terminal loop, which adopts a completely disordered conformation compared with their corresponding residues in the crystal structure (see Figure 1).

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 toward 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

Table 1. Target proteins used for HTD

Receptor	Receptor code	PDB	Resolution (Å)
β₂ 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 1 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
Dopamine D₃ receptor	DRD3	3PBL	2.8
Thymidine kinase	KITH	2UZ3	2.5
Phosphodiesterase 5A	PDE5A	1UDT	2.3
Coagulation factor VII	FA7	1W7X	1.8
Hexokinase type IV	HXK4	3F9M	1.5
Dihydroorotate dehydrogenase	PYRD	1D3G	1.6

Figure 2, and therefore it was chosen for comparison. AF models of ANDR and PRGR were also in the agonist-bound conformation.

In the case of KPCB, where the AF model and the PDB structure had differences at the sequence level 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 toward the inside of protein, making near contact to the binding site and modifying its topology. In this case, however, because the binding pocket is not blocked, we still used the modeled AF structure for HTD to evaluate its performance.

Protein kinases CDK2, IGFR1 and ABL1 show, on average, very good RMSD compared with their PDB structures. The AF model of CDK2 has large differences within the activation loop (containing the DFG motif) and the C-helix (compared with PDB 1FVV). In the case of ABL1, the Gly-rich loop is modeled toward binding site (compared with PDB 2HZI). In KITH, two possible conformations of the flexible loop formed by K49:S68 can be found depending on the ligand bound, as stated by Kosinska and co-workers.[47] We found that although PDB 2B8T has a high backbone superposition to AF model of 4.11 Å in the binding site, PDB 2UZ3 has a better overlap showing an RMSD of 0.69 Å (cf. Table 2). Therefore, the latter PDB structure was used to compare AF model performance.

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. Column 2 shows the results obtained with the ECR consensus

Table 2. Analysis of AF structural models and comparison to their corresponding experimental structures

Receptor	pLDDT[a]	Backbone[b] RMSD (Å)	Backbone[c] RMSD (Å)	Binding site backbone RMSD (Å)	General comments
ABL1	92 ± 5	1.43	0.47	0.79	The Gly-rich loop is pulled toward the binding pocket.
PNPH	95 ± 3	1.69	0.50	0.85	The N55:G66 loop is modeled toward the interior of the protein, near but not in contact with the ligands.
ADRB2	97 ± 2	2.53	2.06	0.81	PDB has missing residues K1232:S1262, which are included in the AF model.
IGF1R	82 ± 16	1.84	1.29	1.64	The Gly-rich loop is in a conserved position, whereas the DFG loop (D1123:E1132) is pulled toward the outside of the protein.
CDK2	92 ± 4	3.73	2.04	0.71	Large backbone differences in the activation loop and C-helix.
COX1	96 ± 1	0.59	0.49	0.61	PDB has D164G and S193G mutations, which have no effect on the binding site; the AF model and the PDB structure lack the heme group near the pocket, which does not affect docking.
PRGR	95 ± 1	0.61	0.52	0.47	—
ANDR	95 ± 1	0.61	0.44	0.16	—
LFA1	85 ± 12	0.73	0.68	1.52	Helix α7 (D297:I306) is pulled toward the inside of the protein, narrowing the binding cavity space.
PTN1	96 ± 6	0.34	0.27	0.22	—
UROK	72 ± 17	1.32	0.46	0.95	PDB has M36I mutation (far from pocket). PDB has crystal waters important for ligand binding.
FABP4	96 ± 3	0.46	0.39	0.47	PDB has crystal waters important for ligand binding.
KPCB	92 ± 5	2.71	2.50	1.4	Residues T500 and S660 are phosphorylated in the PDB but are far from binding site. There is a sequence difference within the C-terminal region (C622:H636), and the backbone is pulled toward the inside of the binding site.
HSP90	94 ± 5	9.23	4.91	4.56	High backbone RMSD of the whole protein. There is a large difference in the position of residues N106:G137, near binding site. PDB has crystal waters important for ligand binding.
ESR1	96 ± 2	1.36	0.38	0.29	The AF model is in the agonist-bound conformation.

(Continued on next page)

Table 2. *Continued*

Receptor	pLDDT[a]	Backbone[b] RMSD (Å)	Backbone[c] RMSD (Å)	Binding site backbone RMSD (Å)	General comments
RENI	84 ± 13	7.76	0.59	10.24	AF model shows a disordered N-terminal loop, which blocks the binding cavity and prevents using the AF structure for docking.
DRD3	93 ± 3	1.09	0.51	0.35	Big difference in the modeled structure between residues R219:G320, far from the binding site.
KITH	94 ± 6	0.75	0.63	0.69	—
PDE5A	95 ± 3	1.45	1.02	0.43	PDB has a gap between residues Y664:Y676. AF model shows a difference in the position of those two residues, which are pulled toward the outside of the protein expanding the binding site. PDB has crystal waters important for ligand binding.
FA7	73 ± 16	1.53	0.71	1.02	—
HXK4	90 ± 6	1.38	0.95	1.70	V62:G71 loop is pulled toward the inside of the binding site, narrowing the space available for ligand binding.
PYRD	98 ± 1	0.55	0.37	0.40	—

The pLDDT metric is reported for residues within the binding site as a measure of model confidence: pLDDT >90: highly confident prediction; 70 < pLDDT<90: confident prediction; 50 < pLDDT<70: low confident prediction; pLDDT<50: should not be interpreted. Reported values correspond to mean and SD. The RMSD values calculated at the backbone level are also displayed.
[a]Per residue Local Distance Difference Test (pLDDT) for residues in the BS (see STAR Methods).
[b]Considering all protein amino acids.
[c]Considering only amino acids involved in secondary structure motifs.

method. Moreover, 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 readily seen that the AF models had a very low performance. On average, EF1 values of 8.4 and 8.8 were obtained with ICM and ECR, respectively. The same trend is observed with the PRC, where an average EF of 8.9 was obtained, with a low average HR of 0.16. Many targets had EF results less than 3.0 and even 0.0 in some cases. It should be noted that the PRC method provided, on average, 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 4 shows a comparison of the results obtained in AF models versus PDB structures using the two consensus methods. It can be seen that, in general, AF models greatly worsen the HTD performance compared with their corresponding crystal structures. The same is also true for the four docking programs individually as seen in Table S1. PRGR, PTN1, DRD3, and KITH were the cases that obtained similar results to the PDB structures. UROK, KPCB, ANDR, FABP4, ADRB2, and PYRD show the largest ECR EF1 decrease compared with docking on PDB structures, followed by PNPH and LFA1. Consistent with this, Table 5 shows that although most PDB structures achieved very low native ligand docking RMSD values, the opposite trend was found for AF models.

Although the AF models used to perform HTD exhibit, in general, an adequate backbone superposition in the binding site to their corresponding PDB structures (cf. RMSD values in Table 2), some striking variations at the side-chain level within the binding site can be observed (cf. Column 6 in Table 4).

In UROK, differences can be observed at the backbone level for ligand binding residues N143, S144, and T145, which are pulled further into the pocket 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 binding site of the AF model is also modified
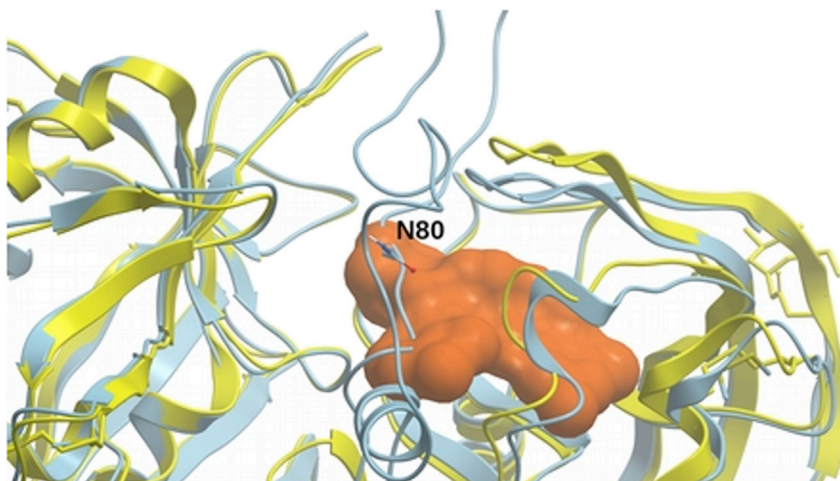
Figure 1. AF model of RENI receptor (cyan) showing an obstructed binding site
The N-terminal loop containing residue N80 is blocking the ligand-binding space (displayed in orange). The corresponding PDB structure 3G6Z is displayed in yellow for comparison.

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. Although for Q711 it was shown by Pereira de Jesús et al.[48] 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 study,[45] which shows how critical it is to include them for HTD. In PYRD, L68 side chain points into the binding pocket, interfering in ligand binding, whereas it points away in the PDB structure. Small variations are also observed in the side chains of residues R136, Y147, H56, and T360.

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 pocket 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
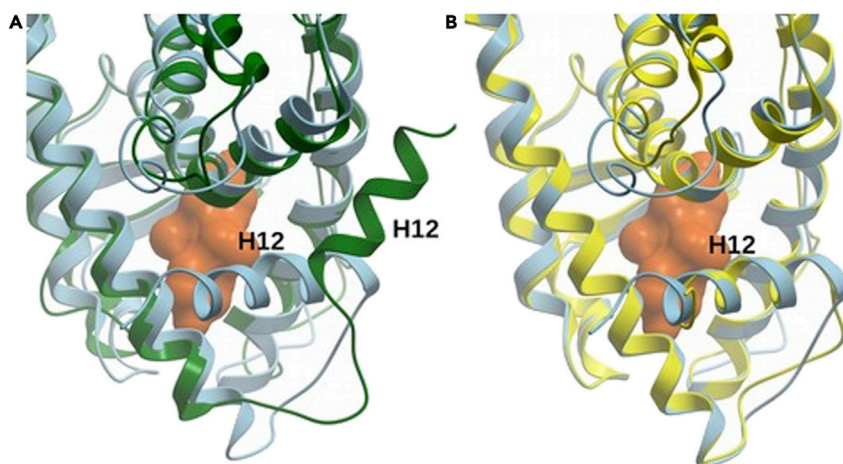

Figure 2. AF modeling of the estrogen receptor
ESR1 AF model (cyan) superimposed to the (A) antagonist-bound conformation (PDB 3ERT) and (B) agonist-bound conformation (PDB 3ERD). The ligand binding space is displayed with orange surfaces.

Table 3. Docking results using AF structural models

Receptor	ICM EF1	ECR EF1	PRC A/S[a]	EF	HR	Native ligand RMSD (Å)
ABL1	24.8	16.0	21/65	19.5	0.32	0.66
PNPH	13.6	18.6	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	8.1	10.2	3/10	10.9	0.30	8.3
COX1	1.9	1.3	4/74	2.5	0.05	>10
PRGR	15.7	12.6	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	24.1	29.5	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	3.7	11.8	1/35	1.9	0.03	6.3
HSP90	4.6	0.0	0/32	0.0	0.00	4.5
ESR1	1.1	8.3	36/206	10.2	0.17	2.5
DRD3	0.6	10.4	7/33	8.5	0.21	7.2
KITH	18.7	22.1	13/32	20.7	0.41	1.0
PDE5A	3.5	10.3	29/141	14.4	0.21	9.32
FA7	9.6	13.1	5/12	23.2	0.42	2.33
HXK4	4.3	1.1	0/5	0	0	9.64
PYRD	7.2	3.6	3/53	3.3	0.06	8.8
Average	8.4	8.8	–	8.9	0.16	–

EF1 is shown for ICM and ECR. The PRC consensus method is evaluated by EF and HR. The corresponding equations can be found in STAR Methods. All these metrics are dimensionless.
[a]Active/Selected.

LFA1 binding site where a notable difference can be observed at the backbone level in helix α7 containing residues L302:I306. This helix is pulled inside the pocket 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.

It can be seen from this analysis that small changes at the side-chain level of essential ligand-binding residues could 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, because overall, those were acceptable. In four out of the five AF models that worsened the HTD performance the most, the pLDDT metric is equal to or greater than 70 for every residue in the binding site (cf. Column 1 in Table 2), indicating high confidence in these modeled structures.

DISCUSSION

In a real-world structure-based drug discovery scenario, most of the researchers would directly use a structure from the PDB, and if not available, it is now possible to select an AlphaFold structure from the AlphaFold Protein Structure Database. The objective of this study is to judge how good are these as-is AlphaFold structures for docking-based virtual screening.

To assess the docking performance of these AlphaFold models, we chose to compare with the performance of HTD in holo PDB structures. As AlphaFold structures present no bound ligand, it could be tempting to judge this holo-PDB versus "apo-like" AF comparison as unfair, as it has been shown that holo structures are more suitable for HTD.[49,50] However, this is not the case, because AF was not designed

Table 4. Comparison of VS results between AF models and PDB structures

Receptor	ECR EF1		PRC EF		Visual inspection comments on binding sites comparison to PDB structures.
	PDB	AF	PDB	AF	
ABL1	25.3	16.0	26.4	19.5	D381 is pulled toward the inside of the binding site. Small difference in the position of the Gly-rich loop.
PNPH	37.1	18.6	34.9	17.9	S33 has a difference in the OH group, which is 2.66 Å pulled to the inside of the pocket.
ADRB2	24.5	3.4	23.4	2.5	Small variation in N1293 and S1203 side chains.
IGF1R	18.3	7.5	38.6	10.1	DFG loop is located toward the outside of the protein. G1125 is 4 Å away in the AF model.
CDK2	12.8	10.2	16.3	10.9	K89 and F80 side chains are slightly pulled inside the pocket, narrowing the binding site.
COX1	3.4	1.3	5.8	2.5	F518 side chain slightly pulled inside the binding site.
PRGR	9.2	12.6	17.3	18.3	W755 is inverted. Difference in Q725 side chain: OH is at a 2.45 Å distance.
ANDR	9.0	0.0	13.5	0.0	Differences in Q711 and T877 side chains (see Figure 3C).
LFA1	10.9	2.9	11.6	0.0	Helix α7 (D297:I306) is pulled inside protein, shrinking the binding site.
PTN1	29.5	29.5	23.9	21.3	D48 and D181 side chains are rotated toward the binding site.
UROK	25.9	2.5	47.0	2.5	N322, S323, and T324 are pulled toward binding site with an average backbone RMSD of 2.28 Å.
FABP4	22.1	0.0	26.4	0.0	F57 is pulled outward of the pocket with an RMSD of 1.6 Å.
KPCB	45.3	11.8	53.8	1.9	C-terminal residues C622:H636 are greatly pulled toward the binding site, modifying its topology. F353 is pulled to the out.
HSP90	0.0	0.0	0.0	0.0	Big difference in structure in N106:G137, near binding site. Important crystal waters missing, which might be critical for ligand binding.
ESR1	34.3	8.3	29.7	10.2	Small difference in M421 and H524 side chains, slightly pulled toward the binding site.
DRD3	3.2	10.4	5.0	8.5	S192 is slightly pulled out of the pocket. T369 is inverted.
KITH	22.1	22.1	20.0	20.7	Small differences in the side chains of residues R53 and R61.
PDE5A	17.0	10.3	23.2	14.4	Y664 is noticeably pulled to the outside of the protein, whereas in the PDB it interferes with the binding site. Q817 and M816 side chains are inverted.
FA7	47.1	13.1	48.0	23.2	Differences in the position of residue K189, slightly pulled out of the pocket.

(Continued on next page)

Table 4. Continued

Receptor	ECR EF1		PRC EF		Visual inspection comments on binding sites comparison to PDB structures.
	PDB	AF	PDB	AF	
HXK4	5.5	1.1	15.2	0	Residues S64:P66 are notably pulled into the binding cavity, narrowing the space available for ligand binding. Y214 side chain is also pulled slightly toward the cavity.
PYRD	27.7	3.6	25.5	3.34	Small differences in R136 and Y147 side-chain positions. L68 points into the binding site, whereas it points away in the PDB. H56 and T360 side chains are flipped.
Average	20.5	8.8	24.1	8.9	—

Results of the two consensus methods ECR and PRC are displayed. Comments at the side-chain level of the binding site residues are found in the last column. For single docking programs results see Table S1.

to predict structures in the apo conformation: AF was trained both with apo and holo structures, and as stated in the Introduction, backbone and side chain coordinates are frequently consistent with the expected structure in the presence of non-protein components (https://alphafold.ebi.ac.uk/faq).

Moreover, given that the goal of this study is to assess how fit AF models are for HTD, it is evident that the comparison must be made between the best option from the AF database and the best option from the PDB database. Given a protein target, the AF database offers a single structure; in the case of the PDB, the reasonable option would be to select a holo structure. Then, the comparison made herein is the one that best serves the main goal of the study.

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 with 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 to (i) large differences at the backbone level within the binding site (as in RENI, where no docking could be performed due to the distortion of the binding site) and (ii) small variations either at the backbone level (UROK, for example) or at the side-chain level (ANDR and PYRD, for example). In several cases, even very subtle differences within the binding site could have a huge impact on the EF, such as in ANDR and FABP4. In agreement with what has been shown by others,[24,25,51] the AF models exhibit low backbone RMSD values compared with PDB structures, thus demonstrating the remarkable ability of AlphaFold to predict protein architecture; moreover, from Table 2, it can be readily seen that our models also show low backbone RMSD and good pLDDT values within the binding site. Therefore, we must conclude that the accuracy of AlphaFold in reproducing protein topology and binding site 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.[11] On the one hand, these results are in agreement with two contemporary studies, namely, Zhang et al.,[52] who evaluated AF models for 28 targets extracted from DUD-E with the Glide docking software,[53] and Díaz-Rovira et al.,[54] who evaluated AF models for 10 targets of the DUD-E. Although in the latter study the utilized docking software was also Glide, the assessment was carried out in a "real-world scenario" by developing a customized AF version that excludes all high-sequence identity templates from the training set.[55] In addition to assessing out-of-the-box AF structures, Zhang et al. have shown that refining AF structures using the IFD-MD induced-fit docking method[56] significantly improves enrichment factors. On the other hand, Wong et al.[57] developed a model to predict protein-ligand interactions based on AF structures and molecular docking and indicated, contrary to our results, that "molecular docking using AlphaFold2-predicted structures is similar to using experimentally determined ones." On top of mentioning that the comparison that yields this conclusion was only made with eight experimental structures, it is also worth considering that model performance was weak by using either experimental structures or AlphaFold structures: the mean area under the receiver operating characteristic curve (AUROC) was, approximately, 0.48, which is worse than random. A slight improvement was obtained when using machine learning scoring functions (mean AUROC of 0.63).

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
DRD3	0.65	7.2
KITH	0.51	1.0
PDE5A	3.37	9.32
FA7	3.13	2.33
HXK4	0.92	9.64
PYRD	0.23	8.8

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 versus 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 binding site 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;[58] although 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.[58] It should be thus acknowledged that different structures of the same protein available in the PDB might indeed represent structural diversity to a certain degree, which right now is not available for AF models.

In this contribution, we compared the AF models with 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 to ensure that the modeled structure is in the desired biological conformation. It should be noted that this issue is many times avoided by using homology modeling, where the structural template from the PDB is chosen taking into consideration the sought biological state of the target;[6] 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.[59] 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 recently reported.[60]

Regarding AlphaFold limitations, which have been discussed elsewhere,[32,35,36,58] it is observed that, from a structure-based drug discovery perspective, AF also provides an incomplete structural model due to the lack of water molecules, metal ions, and co-factors. Just to further illustrate this issue, in HSP90 a very poor performance was obtained using both the AF model and the PDB structure omitting crystallized waters (cf. Table 4), whereas by including water molecules in docking a PRC EF of 15.4 is obtained[45] (the ligand RMSD values with and without 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
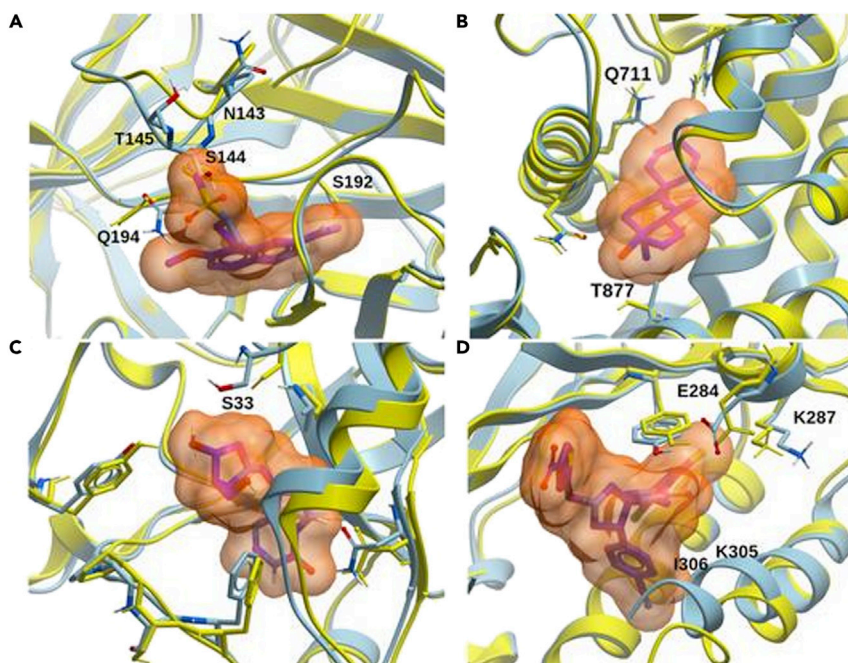
Figure 3. Comparison of binding sites for selected targets

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.

glutamine flipping, protonation states (especially acidic residues, histidines, and cysteines 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 binding pocket together with known ligands (whenever available) in a ligand-steered fashion[16] might be the best strategy to sample binding site 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 out 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 is 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,[28] this issue clearly deserves further analysis.

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.

Limitations of the study

The conclusions drawn from this study to assess the impact of AF models on HTD enrichments are based on a benchmark of 22 different proteins; although this benchmark could be extended, we expect the conclusions drawn in that case to be qualitative like the ones outlined earlier. This study utilizes AlphaFold

structures reported in the AlphaFold Database (accessed November 2022). Although updates in the AlphaFold database or structures generated with the latest version of AlphaFold may lead to slightly different results, we do not expect significant modifications of the results obtained nor the conclusions drawn from them.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Target preparation
  - Protein metrics
  - Docking libraries
  - Docking methods
  - Consensus methods

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105920.

AUTHOR CONTRIBUTIONS

Conceptualization, C.N.C.; Methodology, V.S., J.I.DF., and C.N.C.; Software, V.S., J.I.DF., and C.N.C.; Validation, V.S. and J.I.DF.; Formal Analysis, V.S. and J.I.DF.; Investigation, V.S., J.I.DF., and C.N.C.; Resources, V.S., J.I.DF., and C.N.C.; Writing—Original Draft, V.S., J.I.DF., and C.N.C.; Writing—Review & Editing, V.S., J.I.DF., and C.N.C.; Visualization, V.S.; Supervision, C.N.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., et al. (2002). The protein data bank. Acta Crystallogr. D Biol. Crystallogr. *58*, 899–907.

2. Levitt, M. (2007). Growth of novel protein structural data. Proc. Natl. Acad. Sci. USA *104*, 3183–3188. https://doi.org/10.1073/pnas.0611678104.

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

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

5. Fiser, A. (2004). Protein structure modeling in the proteomics era. Expert Rev. Proteomics *1*, 97–110. https://doi.org/10.1586/14789450.1.1.97.

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

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

8. 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. https://doi.org/10.1016/j.abb.2015.08.002.

9. 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.

10. 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. https://doi.org/10.1016/j.jmgm.2010.11.005.

11. 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.

12. Kufareva, I., Katritch, V., Participants of GPCR Dock 2013, 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. https://doi.org/10.1016/j.str.2014.06.012.

13. Kufareva, I., Rueda, M., Katritch, V., Stevens, R.C., and Abagyan, R.; GPCR Dock 2010 participants (2011). Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. Structure 19, 1108–1126. https://doi.org/10.1016/j.str.2011.05.012.

14. Michino, M., Abola, E., GPCR Dock 2008 participants, 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. https://doi.org/10.1038/nrd2877.

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

16. 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. https://doi.org/10.1021/ci100285f.

17. 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. https://doi.org/10.1021/ci400502u.

18. Cavasotto, C.N., Orry, A.J.W., 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.

19. 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.

20. 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. https://doi.org/10.1021/ja042260c.

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

22. 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.

23. 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. https://doi.org/10.1021/ci4000147.

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

25. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Applying and improving AlphaFold at CASP14. Proteins 89, 1711–1721. https://doi.org/10.1002/prot.26257.

26. 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. https://doi.org/10.1042/BCJ20200963.

27. Marx, V. (2022). Method of the year 2021: protein structure prediction. Nat. Methods 19, 5–10. https://doi.org/10.1038/s41592-021-01380-4.

28. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, 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. https://doi.org/10.1038/s41586-021-03828-1.

29. 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. https://doi.org/10.1016/j.jmb.2021.167336.

30. 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. https://doi.org/10.1093/nar/gkab1061.

31. Subramaniam, S., and Kleywegt, G.J. (2022). A paradigm shift in structural biology. Nat. Methods 19, 20–23. https://doi.org/10.1038/s41592-021-01361-7.

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

33. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2022). Protein complex prediction with AlphaFold-Multimer. Preprint at bioRxiv. https://doi.org/10.1101/2021.10.04.463034.

34. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods 19, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

35. 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. Preprint at bioRxiv. https://doi.org/10.1101/2021.09.26.461876.

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

37. 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. Preprint at bioRxiv. https://doi.org/10.1101/2021.05.10.443524.

38. McCoy, A.J., Sammito, M.D., and Read, R.J. (2022). Implications of AlphaFold2 for crystallographic phasing by molecular replacement. Acta Crystallogr. D Struct. Biol. 78, 1–13. https://doi.org/10.1107/S2059798321012122.

39. 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. https://doi.org/10.1002/prot.26171.

40. Fowler, N.J., and Williamson, M.P. (2022). The accuracy of protein structures in solution determined by AlphaFold and NMR. Structure 30, 925–933.e2. https://doi.org/10.1016/j.str.2022.04.005.

41. 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, bbab564. https://doi.org/10.1093/bib/bbab564.

42. Jendrusch, M., Korbel, J.O., and Sadiq, S.K. (2021). AlphaDesign: a de novo protein design framework based on AlphaFold. Preprint at bioRxiv. https://doi.org/10.1101/2021.10.11.463937.

43. Moffat, L., Greener, J.G., and Jones, D.T. (2021). Using AlphaFold for rapid and accurate fixed backbone protein design. Preprint at bioRxiv. https://doi.org/10.1101/2021.08.24.457549.

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

45. 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 Adv. *11*, 35383–35391. https://doi.org/10.1039/d1ra05785e.

46. Palacio-Rodríguez, 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. https://doi.org/10.1038/s41598-019-41594-3.

47. Kosinska, U., Carnrot, C., Eriksson, S., Wang, L., and Eklund, H. (2005). Structure of the substrate complex of thymidine kinase from Ureaplasma urealyticum and investigations of possible drug targets for the enzyme. FEBS J. *272*, 6365–6372. https://doi.org/10.1111/j.1742-4658.2005.05030.x.

48. 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.

49. An, X., Lu, S., Song, K., Shen, Q., Huang, M., Yao, X., Liu, H., and Zhang, J. (2019). Are the apo proteins suitable for the rational discovery of allosteric drugs? J. Chem. Inf. Model. *59*, 597–604. https://doi.org/10.1021/acs.jcim.8b00735.

50. Guterres, H., Park, S.J., Jiang, W., and Im, W. (2021). Ligand-binding-site refinement to generate reliable holo protein structure conformations from apo structures. J. Chem. Inf. Model. *61*, 535–546. https://doi.org/10.1021/acs.jcim.0c01354.

51. Stevens, A.O., and He, Y. (2022). Benchmarking the accuracy of AlphaFold 2 in loop structure prediction. Biomolecules *12*, 985. https://doi.org/10.3390/biom12070985.

52. Zhang, Y., Vaaa, M., Shi, D., Abualrous, E., Chambers, J., Chopra, N., Higgs, C., Kasavajhala, K., Li, H., Nandekar, P., et al. (2022). Benchmarking refined and unrefined AlphaFold2 structures for hit discovery. Preprint at ChemRxiv. https://doi.org/10.26434/chemrxiv-2022-kcn0d-v2.

53. Friesner, R.A., Murphy, R.B., Repasky, M.P., Frye, L.L., Greenwood, J.R., Halgren, T.A., Sanschagrin, P.C., and Mainz, D.T. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J. Med. Chem. *49*, 6177–6196. https://doi.org/10.1021/jm051256o.

54. Díaz-Rovira, A.M., Martín, H., Beuming, T., Díaz, L., Guallar, V., and Ray, S.S. (2022). Are deep learning structural models sufficiently accurate for virtual screening? Application of docking algorithms to AlphaFold2 predicted structures. Preprint at bioRxiv. https://doi.org/10.1101/2022.08.18.504412.

55. Beuming, T., Martín, H., Díaz-Rovira, A.M., Díaz, L., Guallar, V., and Ray, S.S. (2022). Are deep learning structural models sufficiently accurate for free-energy calculations? Application of FEP+ to AlphaFold2-predicted structures. J. Chem. Inf. Model. *62*, 4351–4360. https://doi.org/10.1021/acs.jcim.2c00796.

56. Miller, E.B., Murphy, R.B., Sindhikara, D., Borrelli, K.W., Grieswood, M.J., Ranalli, F., Dixon, S.L., Jerome, S., Boyles, N.A., Day, T., et al. (2021). Reliable and accurate solution to the induced fit docking problem for protein-ligand binding. J. Chem. Theor. Comput. *17*, 2630–2639. https://doi.org/10.1021/acs.jctc.1c00136.

57. Wong, F., Krishnan, A., Zheng, E.J., Stärk, H., Manson, A.L., Earl, A.M., Jaakkola, T., and Collins, J.J. (2022). Benchmarking AlphaFold-enabled molecular docking predictions for antibiotic discovery. Mol. Syst. Biol. *18*, e11081. https://doi.org/10.15252/msb.202211081.

58. 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. https://doi.org/10.1021/acs.jcim.2c00026.

59. 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. https://doi.org/10.1039/c5cc05050b.

60. Heo, L., and Feig, M. (2022). Multi-state modeling of G-protein coupled receptors at experimental accuracy. Proteins *90*, 1873–1885. https://doi.org/10.1002/prot.26382.

61. 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. https://doi.org/10.1021/jm300687e.

62. Lagarde, N., Ben Nasr, N., Jérémie, 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. https://doi.org/10.1021/jm500132p.

63. 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. https://doi.org/10.1021/ci200412p.

64. 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.

65. 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. https://doi.org/10.1002/jcc.21256.

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

67. 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. https://doi.org/10.1371/journal.pcbi.1003571.

68. Cavasotto, C.N., and Aucar, M.G. (2020). High-throughput docking using quantum mechanical scoring. Front. Chem. *8*, 246. https://doi.org/10.3389/fchem.2020.00246.

69. 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. https://doi.org/10.1093/bioinformatics/btt473.

70. Huang, N., Shoichet, B.K., and Irwin, J.J. (2006). Benchmarking sets for molecular docking. J. Med. Chem. *49*, 6789–6801. https://doi.org/10.1021/jm0608356.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
PDB	(Berman et al.,[1] 2002)	https://www.rcsb.org
DUD-E	(Mysinger et al.,[61] 2012)	http://dude.docking.org
NRLiSt	(Lagarde et al.,[62] 2014)	http://nrlist.drugdesign.fr
GLL/GDD	(Gatica and Cavasotto,[63] 2012)	https://cavasotto-lab.net
Alpha-Fold Database	(Jumper et al.,[24] 2021; Varadi et al.,[30] 2022)	https://alphafold.ebi.ac.uk
Alpha-Fold (Colab version)	(Jumper et al.,[24] 2021)	https://github.com/deepmind/alphafold
ICM	(Abagyan et al.,[64] 1994)	https://www.molsoft.com
Auto Dock 4	(Morris et al.,[65] 2009)	https://autodock.scripps.edu
PLANTS	(Korb et al.[66] 2009)	www.tcd.uni-konstanz.de
rDock	(Ruiz-Carmona et al.,[67] 2014)	https://rdock.sourceforge.net

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Claudio Cavasotto (CCavasotto@austral.edu.ar; cnc@cavasotto-lab.net).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This paper analyzes existing, publicly available data. Databases are listed in the key resources table.

- This paper does not report original code.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Target preparation

The 22 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 Alpha-Fold Protein Structure Database[30] using the corresponding Uniprot identification. An additional Alpha-Fold 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[64] (version 3.9-2e; MolSoft, San Diego, CA, May 2022), in a similar fashion as in earlier works.[45,68] 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$_\alpha$, 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 ligand-binding 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 binding site residue. The predicted Local Distance Difference Test (pLDDT) is a per residue metric reported in the Alpha-Fold Protein Structure Database[30] 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.[69] Following this evaluation criterion, we looked at the pLDDT metric especially for binding site 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.[63,70] For all molecules, chirality and protonation states were inherited from the corresponding original databases. Libraries were obtained from the DUD-E database,[61] except for the ESR1 agonists library which was obtained from NRLiSt[62] database, and the ADRB2 library which was taken from GLL/GDD.[63] 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,[64] Auto Dock 4,[65] rDock[67] and PLANTS.[66] These programs have different search algorithms and scoring functions as described in previous studies.[45,46] Auto Dock Tools utilities[65] were used to prepare the input files for Auto Dock 4. The Lamarckian genetic algorithm was used for a 20-run search for each compound using 1.75 million energy evaluation. For ICM, a thoroughness of 2 was used for the search algorithm. The ChemPLP scoring function was used in PLANTS and speed 1 was set as search speed. For rDock, a radius of 8.0 Å $\pm$ 2.0 Å from a reference ligand binding mode was used to represent the cavity. For Vina, an exhaustiveness value of 8 was set. All the other parameters for every software remained at their default values. This parameter setting is the same used in a previous study,[45] 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)[46] 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 and establishes the number of molecules for each scoring function that will be considered; 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.

The Pose/Ranking Consensus method (PRC)[45] 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 consensus method described above 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

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

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.