



Communication

AI-Based Protein Interaction Screening and Identification (AISID)

Zheng-Qing Fu ^{1,2,*}, Hansen L. Sha ³ and Bingdong Sha ^{3,*}

¹ SER-CAT, Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60439, USA

² Department of Biochemistry & Molecular Biology, University of Georgia, Athens, GA 30602, USA

³ Department of Cell, Developmental and Integrative Biology (CDIB), University of Alabama at Birmingham, Birmingham, AL 35294, USA

* Correspondence: fuzq@anl.gov (Z.-Q.F.); bdsha@uab.edu (B.S.)

Abstract: In this study, we presented an AISID method extending AlphaFold-Multimer's success in structure prediction towards identifying specific protein interactions with an optimized AISIDscore. The method was tested to identify the binding proteins in 18 human TNFSF (Tumor Necrosis Factor superfamily) members for each of 27 human TNFRSF (TNF receptor superfamily) members. For each TNFRSF member, we ranked the AISIDscore among the 18 TNFSF members. The correct pairing resulted in the highest AISIDscore for 13 out of 24 TNFRSF members which have known interactions with TNFSF members. Out of the 33 correct pairing between TNFSF and TNFRSF members, 28 pairs could be found in the top five (including 25 pairs in the top three) seats in the AISIDscore ranking. Surprisingly, the specific interactions between TNFSF10 (TNF-related apoptosis-inducing ligand, TRAIL) and its decoy receptors DcR1 and DcR2 gave the highest AISIDscore in the list, while the structures of DcR1 and DcR2 are unknown. The data strongly suggests that AlphaFold-Multimer might be a useful computational screening tool to find novel specific protein bindings. This AISID method may have broad applications in protein biochemistry, extending the application of AlphaFold far beyond structure predictions.



Citation: Fu, Z.-Q.; Sha, H.L.; Sha, B. AI-Based Protein Interaction Screening and Identification (AISID). *Int. J. Mol. Sci.* **2022**, *23*, 11685. <https://doi.org/10.3390/ijms231911685>

Academic Editor: Ray Luo

Received: 8 September 2022

Accepted: 1 October 2022

Published: 2 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: proteins binding; computer-aided screening; AlphaFold; AISID

1. Introduction

Over the last several decades, various algorithms have been developed to model unknown structures based on the information deposited at the Protein Data Bank (PDB). The AlphaFold2 is an Artificial Intelligent (AI) program that utilizes deep learning to perform protein structure prediction [1], which demonstrated an outstanding ability to predict the single-chain protein structures during the 14th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP14) competition [2]. Other platforms have also been developed for protein structure modeling, either independently or through exploiting the existing version of AlphaFold2 [3–8]. Lately, the AlphaFold2 team have released AlphaFold-Multimer, which added the capability to model protein complexes of multiple chains [9]. More recently, it has been shown that AlphaFold-Multimer can be utilized to predict protein complex structure with accuracy for soluble protein complexes, transmembrane ABC transporter complex, and protein complexes containing intrinsic disordered proteins (IDP) [3,10–14].

While precise atomic structures of protein complexes could provide insight of the detailed static interactions, identifying novel specific protein interactions may have a broader interest in protein characterizations. To identify the possible binding partner for a bait protein from a long list of protein candidates, it would be very helpful that a computational screening method can quickly predict binding potentials and prioritize the candidates for further biochemical assays. Such a method could revolutionize the protein characterization process. In this study, we presented an AI-based protein interaction

screening and identification (AISID) method extending AlphaFold-Multimer's success in structure prediction towards identifying specific protein interactions.

2. Results and Discussions

To predict a heterodimer protein complex structure, the AlphaFold-Multimer generates 25 (by default) complex structure models with a model confidence value for assessing the quality of each model. We derive a metric score (denoted as AISIDscore hereafter) from the model confidence values, and the AISIDscore is optimized for assessing binding potentials between the protein pairs. We hypothesize that the AISIDscore can serve as a quality indicator to evaluate the binding potential between the protein pair. In the computational screening, a bait protein and a panel of protein ligand candidates can be input into the AlphaFold-Multimer for calculations. A higher AISIDscore might indicate the higher likelihood of specific binding between the bait and the protein ligand.

To test our method, 27 human TNFRSF members and 18 human TNFSF members were utilized. In all, 24 TNFRSF members out of the 27 can specifically bind at least one of the 18 TNFSF members as documented by the Uniprot database [15]. The specific binding between a TNFRSF member and its ligand TNFSF member(s) leads to distinct signaling pathways, including cell apoptosis, inflammation, and proliferations [16,17]. In each test, two protein sequences, one from a TNFRSF member and the other from a TNFSF member, were input into the AlphaFold-Multimer. The whole screening test generated a panel of AISIDscores for 27×18 pairs of TNFRSF and TNFSF members. For each TNFRSF member (as a bait protein), we ranked the AISIDscore among the 18 TNFSF members. The correct pairing was revealed as the highest AISIDscore for 13 out of 24 TNFRSF members, which have known specific interactions with TNFSF members. Out of the 33 correct pairings between TNFRSF and TNFSF members, 28 pairs could be found in the top five (including 25 pairs in the top three) seats of the AISIDscore rankings (Figure 1). The AISIDscore ranges from 0.02 to 0.85 with an average value of 0.46 for all pairs, and a much higher average value of 0.74 for the correct pairs. As a negative control, the AISIDscores between Hsp40 Hdj1 and TNFSF members were also calculated, resulted in a much lower average value of 0.15.

It is of great interest to examine the ability of AISID for specific interactions between proteins with unknown structures. The TNFRSF members TNFRSF10C and TNFRSF10D (TRAIL DcR1 and TRAIL DcR2) can interact specifically with their ligand TNFSF10 (TRAIL), while their structural information is missing. Surprisingly, the pairing between the DcR1 and DcR2 and their ligand TRAIL gave the highest AISIDscore in the list (Figures 1 and 2A,B). Biochemical studies have shown that both DcR1 and DcR2 could specifically bind TRAIL with the K_D of ~ 1 nM. DcR1 and DcR2 do not interact with other TNFSF members [18,19]. In another case of CD30/CD30L, both CD30 and its ligand CD30L are with unknown structures. The prediction showed that the CD30/CD30L pair provided the second highest AISIDscore in the list (Figure 2C). The studies using Surface Plasmon Resonance (SPR) have indicated that CD30 specifically interacted with CD30L with the K_D of ~ 5 nM [20]. Therefore, the prediction results using the AISIDscore are consistent with the previous biochemical studies. The data clearly indicated that the AISIDscore derived from the AlphaFold-Multimer might be a useful indicator to identify the specific interactions.

As a positive control, we took advantage of a nanobody that can specifically bind TNF α [21]. The protein complex structure of TNF α and this specific nanobody was deposited into PDB on 13 October 2016. We asked the AlphaFold-Multimer to perform the complex structure prediction between the nanobody and 18 TNFSF members. To prevent the AlphaFold-Multimer using the structure information in PDB, we also perform the calculation with the cutoff date of 1 January 2016. In both cases, the calculations gave the highest AISIDscore for the correct pairing of TNF α and the nanobody (Table 1). These data strongly suggest that AISID may be utilized as a valid method for identifying novel specific protein bindings.

Table 1. AISIDscores of TNF α nanobody against TNFSF superfamily.

TNFSF/AISIDscore	Default	Cutoff Date 1 January 2016
TNF α	0.871	0.869
TRAIL	0.713	0.705
4-1BBL	0.699	0.392
LIGHT	0.590	0.709
TNF β	0.563	0.515
TNFSF13	0.552	0.454
TNFSF4	0.484	0.118
TNFSF12	0.271	0.101
TNF γ	0.261	0.311
CD70	0.233	0.353
CD153	0.223	0.096
TNFSF11	0.215	0.265
TNFSF13B	0.204	0.253
TNFSF18	0.163	0.276
FASLG	0.161	0.155
EDA	0.150	0.158
TNFSF15	0.098	0.089
CD40LG	0.081	0.082

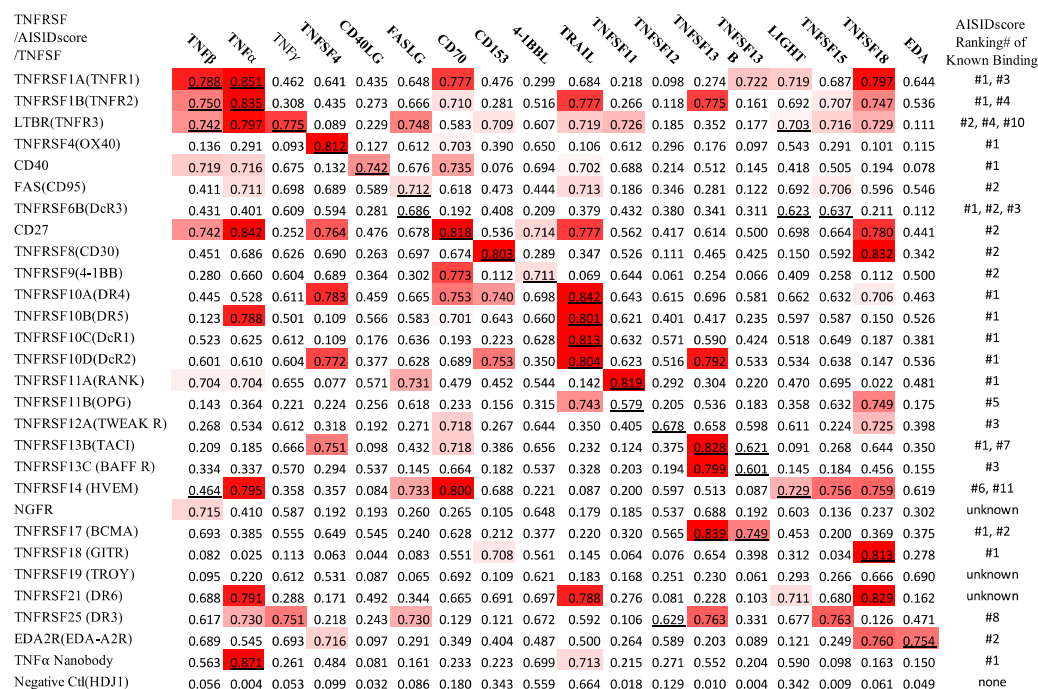
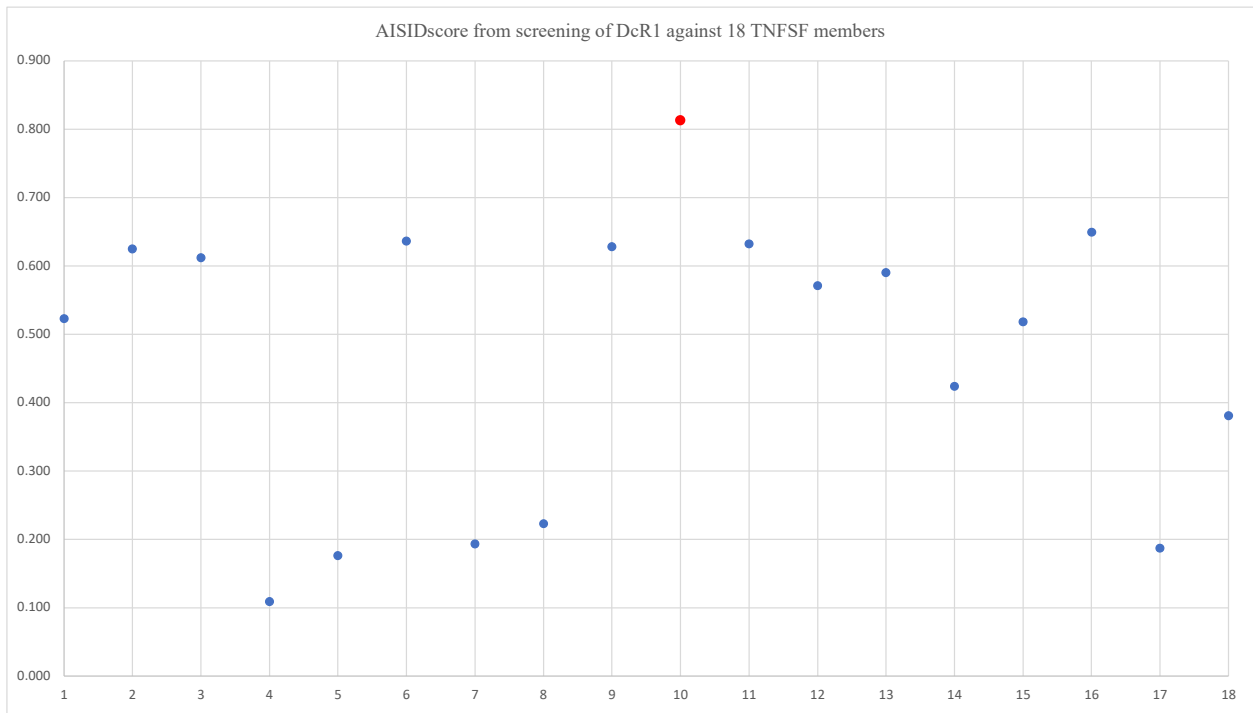
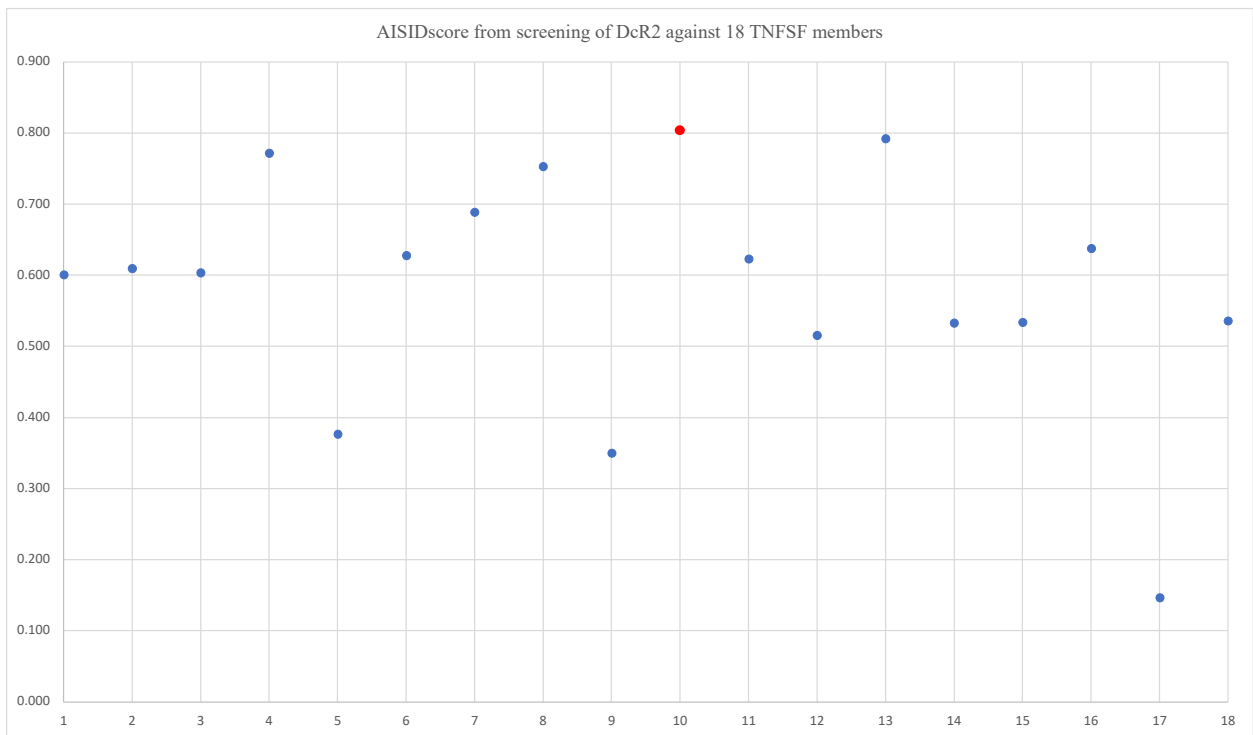


Figure 1. Heatmap of AISIDscores from 29 bait proteins (27 TNFRSF members, TNF α nanobody for positive control, Hsp40Hdj1 for negative control) against 18 TNFSF members. The AISIDscore of correct pairings between bait protein and TNFSF members are underlined. For each TNFRSF member, AISIDscores were ranked among 18 TNFSF members, with the ranking positions of the correct pairs are listed at the rightmost column. Three TNF receptor (NGFR, TROY, and DR6) do not have known ligands in TNFSF members and are labeled as “unknown”. The human Hsp40Hdj1 is labeled as “none”.



(A)



(B)

Figure 2. Cont.

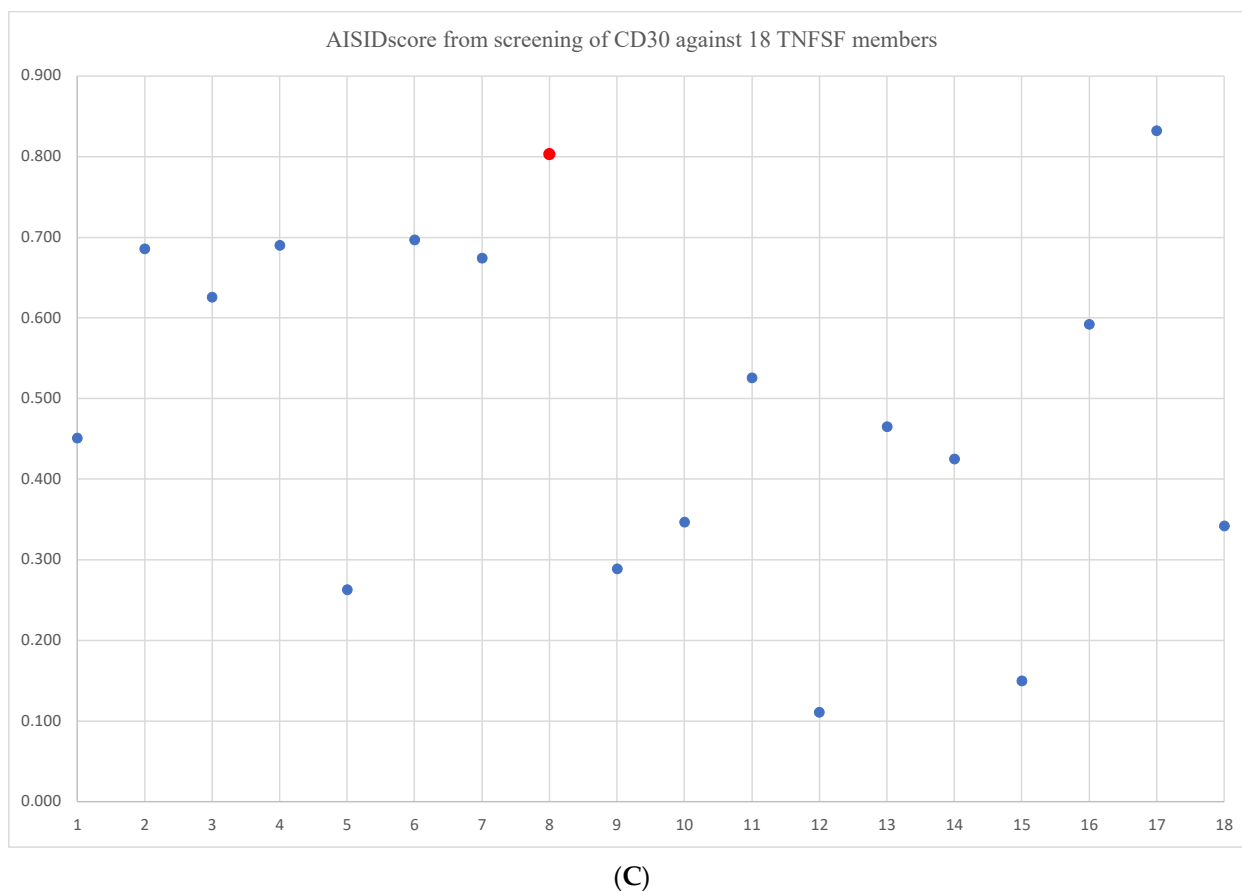


Figure 2. Plot of AISIDscores from screening the TNFRSF members, which do not have solved complex structures. (A–C), represent DcR1, DcR2, and CD30, respectively, with the 18 TNFSF members in same order as Figure 1. The red circle indicates the correct pairing.

It is worth noting that a relatively high sequence homology exists in TNFRSF and TNFSF family members. For example, the TNF α and TRAIL share ~25% sequence identity and ~60% sequence similarity. Our data indicate that the AISID method can reveal the specific interactions even among the highly homologous TNFSF family members. It is highly likely that the AISID may greatly speed up the screening for unknown specific protein bindings by prioritizing a long list of potential binding partners. The AISID-guided biochemical assay may follow to confirm the bindings.

Many protein interactions are mediated by intrinsic disordered proteins (IDP). Some IDPs have been showed to undergo a disorder-to-order transition upon recognizing their physiological partners [22]. Trying to postulate the binding partners for IDP by examining its primary sequence would be extremely challenging. The AISID can put the IDP in the context with its potential partners to predict the binding, which may mimic the fold-and-bind scenario for IDP.

The detailed atomic structure of predicted complex model is not the focus of AISID screening. It is structure-blind, and no other prior biological information, except only a pair of protein sequences, are needed for calculating the AISIDscore. Users will not need to examine the accuracy of the model complex structures, which makes the method user-friendly for people working in wet lab.

The presented AISID method may have significant biological impact in general biochemistry and cell biology. To characterize a protein with unknown functions, it is important to identify its binding partners. Using the traditional methods such as a yeast two hybrid system has been shown to be problematic and time-consuming. The AISID method may provide a fast alternative to identify the possible binding partner for a bait

protein from a long list of protein candidates. Our method offers an AISIDscore which quantitatively measures the likelihood for a particular protein-protein interaction. The data showed that the AISIDscore was accurate in revealing specific protein interactions even among highly homologous proteins. Such a method may revolutionize how protein interactions are characterized.

3. Methods

In this presented AISID method for specific protein bindings screening, the AlphaFold-Multimer [9] was used to calculate the ranking metric. The AlphaFold-Multimer is a publicly available software package (<https://github.com/deepmind/alphafold>, accessed on 18 March 2022) recently released by DeepMind which extends the AlphaFold2 to multiple chains during both training and inference with native support for multiple-chain featurization and symmetry handling, surpassing those of inference-only modifications to the AlphaFold2.

When used in complex structure prediction, the AlphaFold-Multimer generates a merit index, the model confidence, defined as a weighted combination of ipTM and pTM:

$$\text{Model Confidence} = 0.8 \times \text{ipTM} + 0.2 \times \text{pTM} \quad (1)$$

which is used to assess the overall accuracy of a predicted complex structure model. The two weighted factors ipTM and pTM are for evaluating inter-chain and intra-chain quality, respectively. In cases when the predicted model with either one or both composite proteins does not have high accuracy, the 'Model Confidence' thus defined could skew the evaluation of specific bindings, especially when not all domains of the proteins folded meaningfully. For identifying the potential specific binding between a bait protein and a ligand protein, a metric using only the inter-chain part would be more appropriate. In heterodimer complex structure predictions, the AlphaFold-Multimer generates 25 models by default, each with a model confidence. Due to the reasons described above, we derive a new metric (denoted as AISIDscore for simplicity) to assess binding potential between the proteins pair:

$$\text{AISIDscore} = \max\{1.25 \times (\text{Model Confidence}) - 0.25 \times \text{avp}\} \quad (2)$$

Here, 'avp' is the averaged percentage pLDDT [1] score of all amino acid residues in the predicted complex structure model, and 'max{}' runs through the 25 models to calculate the maximum. The AISIDscore thus defined is intended to highlight binding potential assessment by weighting down the impact of possible folding errors of individual chains.

The sequence of the bait protein to be investigated, and each sequence from the list of candidate proteins to be screened, are used to create a composite sequence file as input to the AlphaFold-Multimer for heterodimer complex structure predictions. As computing time of a prediction job heavily depends on the size of sequence, the list of composite sequence files is sorted by the number of amino acid residues from small to large. The sorting helps optimize the project production cycle (not the machine time of a prediction job), especially when carrying out a large screening project on a server with limited computing nodes and/or the candidate proteins are quite different in sizes. Batch jobs with inputs from the sorted list of composite sequence files are submitted through a load-balanced scheduler to the computer server running the AlphaFold-Multimer. Each of the composite sequence file will instruct the AlphaFold-Multimer to generate 25 (by default) predicted heterodimer complex structure models. The AISIDscore is then derived for the pair from harvesting and analyzing the outputs. Upon the finish of the screening project, a list of AISIDscores will be generated, each for one bait-ligand pair.

For proof of concept, we picked 29 bait proteins including 27 human TNFRSF (tumor necrosis factor receptor superfamily) member proteins, a TNF α nanobody as positive control, and protein Hsp40 Hdj1 as negative control. The 18 human TNFSF (tumor necrosis factor superfamily) member proteins are used as the ligand proteins to be screened. Human

TNFRSF and TNFSF are two of the well-studied protein superfamilies with many known specific bindings, which is good for methods validation and evaluation. Figure 1 shows the results from the 29 screening projects.

The AISID screening focuses on finding the potential specific binding proteins, not the detailed atomic structure of the predicted complex models. The sequences of bait and ligand proteins are the only inputs to the AlphaFold-Multimer for calculating the AISIDscores. No other prior biological information is needed. For routine uses, one simply needs to examine the AISIDscore in search for the potential specific binding protein(s). The workflow diagram of a typical screening project is summarized in Figure 3. A group of in-house helper scripts have been created to facilitate executing the steps. From sequences to the ranked list of final AISIDscores, a screening project could be carried out by wet-lab scientists without special training in modeling, nor in interpreting details of an atomic structure.

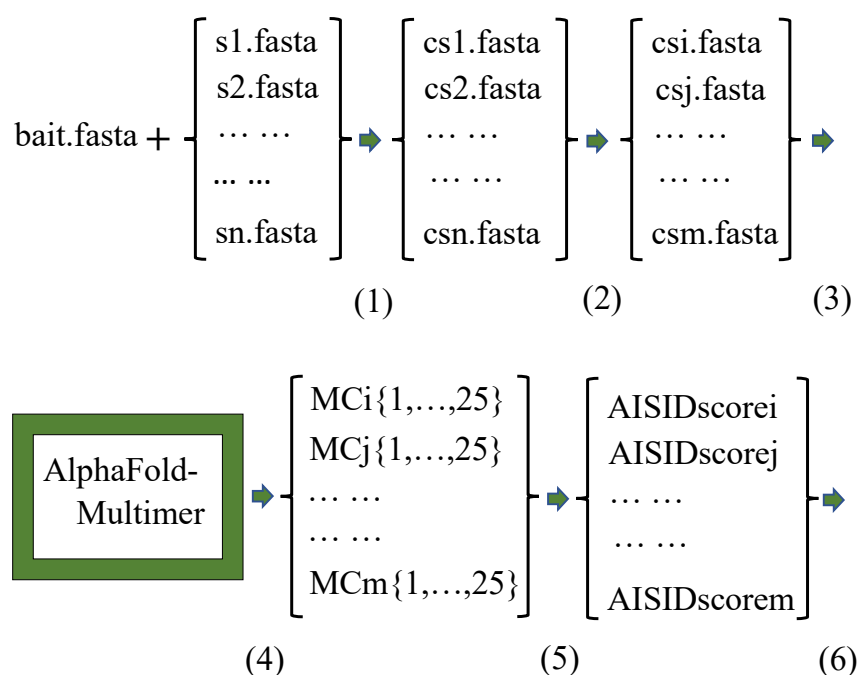


Figure 3. Workflow diagram of specific protein binding screening: (1). Create a list of composite sequence files of FASTA-format. Each file contains two separate sequences, one is that of the bait protein, the other is from one of the ligand proteins against which to be screened. (2). Sort the list by the number of amino acid residues in the composite sequence files from small to large. (3). Through a load-balanced scheduler, submit the batch jobs onto the server running the AlphaFold-multimer. (4). Once the project finishes, each job generates 25 predicted heterodimer complex structure models. (5). Harvest outputs and calculate the AISIDscore for each bait-ligand pair, resulting in a list of AISIDscores for the screening project. (6). Sort the list by AISIDscore from large to small for evaluation.

4. Conclusions

In our test sets, 27 human TNFRSF members and 18 human TNFSF members were utilized. The specific interactions between TNFRSF members and TNFSF members have been extensively studied. Relatively high sequence homologues exist among TNFRSF members and TNFSF member, while each TNFRSF member can specifically bind TNFSF members to initiate distinct signaling pathways. The interaction network constituted by TNFRSF and TNFSF members provide us with an excellent database to investigate the AISID method for revealing specific protein bindings. We reason that this small but very well-informed database may be ideal for testing our method. Our results showed that out

of the 33 correct pairings between TNFRSF and TNFSF members, 25 pairs can be found in the top three seats of the AISIDscore rankings (Figure 1).

The data in this study strongly suggest that the AISID method has the capability to identify novel protein interactions. It may greatly speed up the screening for unknown specific protein bindings by prioritizing a long list of potential binding partners. The AISID method may have broad applications for protein characterizations in protein biochemistry, extending the application of the AlphaFold far beyond structure predictions.

Author Contributions: Conceptualization, Z.-Q.F. and B.S.; Methodology, Z.-Q.F. and B.S.; Software, Z.-Q.F. and H.L.S.; Formal analysis, Z.-Q.F., H.L.S. and B.S. Data curation, Z.-Q.F., H.L.S. and B.S. Writing, Z.-Q.F., H.L.S. and B.S. Investigation, Z.-Q.F. and B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study is in part supported by NIH R01HD102015.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Part of the calculations were carried out on the newly installed AlphaFold2.2.0 server at SER-CAT. Z.-Q.F. thanks SER-CAT at APS of Argonne National Laboratory and Bi-Cheng Wang for the support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)] [[PubMed](#)]
2. Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J. Critical assessment of methods of protein structure prediction (CASP)-Round XIV. *Proteins* **2021**, *89*, 1607–1617. [[CrossRef](#)]
3. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876. [[CrossRef](#)] [[PubMed](#)]
4. Mirdita, M.; Schütze, K.; Moriawaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making protein folding accessible to all. *Nat. Methods* **2022**, *19*, 679–682. [[CrossRef](#)] [[PubMed](#)]
5. Bryant, P.; Pozzati, G.; Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **2022**, *13*, 1265. [[CrossRef](#)]
6. Humphreys, I.R.; Pei, J.; Baek, M.; Krishnakumar, A.; Anishchenko, I.; Ovchinnikov, S.; Zhang, J.; Ness, T.J.; Banjade, S.; Bagde, S.R.; et al. Computed structures of core eukaryotic protein complexes. *Science* **2021**, *374*, eabm4805. [[CrossRef](#)]
7. Le, N.Q.K.; Ho, Q.T.; Ou, Y.Y. Using two-dimensional convolutional neural networks for identifying GTP binding sites in Rab proteins. *J. Bioinform. Comput. Biol.* **2019**, *17*, 1950005. [[CrossRef](#)]
8. Le, N.Q.K.; Sandag, G.A.; Ou, Y.Y. Incorporating post translational modification information for enhancing the predictive performance of membrane transport proteins. *Comput. Biol. Chem.* **2018**, *77*, 251–260. [[CrossRef](#)]
9. Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.W.; Green, T.; Židek, A.; Bates, R.; Blackwell, S.; Yim, J.; et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* **2021**. [[CrossRef](#)]
10. Gao, M.; Nakajima An, D.; Parks, J.M.; Skolnick, J. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nat. Commun.* **2022**, *13*, 1744. [[CrossRef](#)]
11. Tordai, H.; Suhajda, E.; Sillitoe, I.; Nair, S.; Varadi, M.; Hegedus, T. Comprehensive Collection and Prediction of ABC Transmembrane Protein Structures in the AI Era of Structural Biology. *Int. J. Mol. Sci.* **2022**, *23*, 8877. [[CrossRef](#)] [[PubMed](#)]
12. Yin, R.; Feng, B.Y.; Varshney, A.; Pierce, B.G. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci.* **2022**, *31*, e4379. [[CrossRef](#)] [[PubMed](#)]
13. Verburgt, J.; Zhang, Z.; Kihara, D. Multi-level analysis of intrinsically disordered protein docking methods. *Methods* **2022**, *204*, 55–63. [[CrossRef](#)]
14. Ivanov, Y.D.; Taldaev, A.; Lisitsa, A.V.; Ponomarenko, E.A.; Archakov, A.I. Prediction of Monomeric and Dimeric Structures of CYP102A1 Using AlphaFold2 and AlphaFold Multimer and Assessment of Point Mutation Effect on the Efficiency of Intra- and Interprotein Electron Transfer. *Molecules* **2022**, *27*, 1386. [[CrossRef](#)]
15. UniProt, C. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [[CrossRef](#)]
16. Dostert, C.; Grusdat, M.; Letellier, E.; Brenner, D. The TNF Family of Ligands and Receptors: Communication Modules in the Immune System and Beyond. *Physiol. Rev.* **2019**, *99*, 115–160. [[CrossRef](#)]

17. Kucka, K.; Wajant, H. Receptor Oligomerization and Its Relevance for Signaling by Receptors of the Tumor Necrosis Factor Receptor Superfamily. *Front. Cell Dev. Biol.* **2020**, *8*, 615141. [[CrossRef](#)]
18. Degli-Esposti, M.A.; Dougall, W.C.; Smolak, P.J.; Waugh, J.Y.; Smith, C.A.; Goodwin, R.G. The novel receptor TRAIL-R4 induces NF-kappaB and protects against TRAIL-mediated apoptosis, yet retains an incomplete death domain. *Immunity* **1997**, *7*, 813–820. [[CrossRef](#)]
19. Emery, J.G.; McDonnell, P.; Burke, M.B.; Deen, K.C.; Lyn, S.; Silverman, C.; Dul, E.; Appelbaum, E.R.; Eichman, C.; DiPrinzio, R.; et al. Osteoprotegerin is a receptor for the cytotoxic ligand TRAIL. *J. Biol. Chem.* **1998**, *273*, 14363–14367. [[CrossRef](#)]
20. Hargreaves, P.G.; Al-Shamkhani, A. Soluble CD30 binds to CD153 with high affinity and blocks transmembrane signaling by CD30. *Eur. J. Immunol.* **2002**, *32*, 163–173. [[CrossRef](#)]
21. Beirnaert, E.; Desmyter, A.; Spinelli, S.; Lauwereys, M.; Aarden, L.; Dreier, T.; Loris, R.; Silence, K.; Pollet, C.; Cambillau, C.; et al. Bivalent Llama Single-Domain Antibody Fragments against Tumor Necrosis Factor Have Picomolar Potencies due to Intramolecular Interactions. *Front. Immunol.* **2017**, *8*, 867. [[CrossRef](#)] [[PubMed](#)]
22. Toto, A.; Malagrino, F.; Visconti, L.; Troilo, F.; Pagano, L.; Brunori, M.; Jemth, P.; Gianni, S. Templated folding of intrinsically disordered proteins. *J. Biol. Chem.* **2020**, *295*, 6586–6593. [[CrossRef](#)] [[PubMed](#)]