Prediction of TrkB Complex and Antidepressant Targets Leveraging Big Data

Xufu Xiang(✉), Chungen Qian, Xin Liu, and Fuzhen Xia

The Key Laboratory for Biomedical Photonics of MOE at Wuhan National Laboratory for Optoelectronics - Hubei Bioinformatics and Molecular Imaging Key Laboratory, Systems Biology Theme, Department of Biomedical Engineering, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
xiang_xufu@hust.edu.cn

Abstract. Since Major Depressive Disorder (MDD) represents a neurological pathology caused by inter-synaptic messaging errors, membrane receptors, the source of signal cascades, constitute appealing drugs targets. G protein coupled receptors (GPCRs) and ion channel receptors chelated antidepressants (ADs) high-resolution architectures were reported to realize receptors physical mechanism and design prototype compounds with minimal side effects. Tyrosine kinase receptor 2 (TrkB), a receptor that directly modulates synaptic plasticity, has a finite three-dimensional chart due to its high molecular mass and intrinsically disordered regions (IDRs). Leveraging breakthroughs in deep learning, the meticulous architecture of TrkB was projected employing Alphafold 2 (AF2). Furthermore, the Alphafold Multimer algorithm (AF-M) models the coupling of intra- and extra-membrane topologies to chaperones: mBDNF, SHP2, Etc. Conjugating firmly dimeric transmembrane helix with novel compounds like 2R,6R-hydroxynorketamine (2R,6R-HNK) expands scopes of drug screening to encompass all coding sequences throughout genomes. The operational implementation of TrkB kinase-SHP2, PLCγ1, and SHC1 ensembles has paved the path for machine learning in which it can forecast structural transitions in the self-assembly and self-dissociation of molecules during trillions of cellular mechanisms. In silicon, the cornerstone of the alteration will be big data and artificial intelligence (AI), empowering signal networks to operate at the atomic level and picosecond timescales.

Keywords: Major Depressive Disorder (MDD) · Antidepressants (ADs) · Tyrosine kinase receptor 2 (TrkB) · Alphafold Multimer algorithm (AF-M) · Big Data · Artificial Intelligence (AI)

1 Introduction

MDD, a systemic psychiatric disorder, may have a wide range of implicit etiologies: synaptic misconnection, metabolic abnormalities and immune inflammation. The lifetime prevalence of MDD surpasses 20 % in the worldwide population, and the unavailability of specific medications renders one-third of patients unresponsive to treatment

S. Zhang et al. (Eds.): BigData 2023, LNCS 14203, pp. 149–165, 2023.
https://doi.org/10.1007/978-3-031-44725-9_11

[1]. As the initial spot of the signal cascade, transmembrane receptors are indispensable in gaining and transferring signals across a thousand trillion interconnected synapses and as the objective of fifty percent of prescription medicine [2, 3]. Several ADs have a direct link to the 5-hydroxytryptamine receprots (5-HTRs) of GPCRs and the ion channel-type glutamate receptors NMDAR and AMPAR for the onset of action [4]. Subtypes of 5-HTRs binding maps with ADs have been constructed then using virtual pharmacological screenings of millions of compounds to catch novel non-hallucinogenic antidepressants [5–9]. The results of conjugating NMDAR with S-ketamine reveal that S-ketamine takes effect more rapidly than conventional ADs by inhibiting Ca^{2+} influx to intracellular [10]. Previously, the predominant hypothesis claimed that ADs indirectly positively regulate TrkB via NMDAR to trigger the synaptic plasticity mechanism [11]. However, in a latest report, robust R-ketamine release therapeutic advantages by directly binding to TrkB [12]. This casts doubt on the broadly held 5-HT and NMDAR hypotheses and propels TrkB to spring up as a momentous priority receptor in building neat antidepressants.

Determining protein structure is essential for pharmaceutical research, while there is a deficit of architectural insights into TrkB despite the profusion of foundational research results. TrkB modulates postsynaptic protein expression and synaptogenesis through the MAPK, PIK3/mTOR, and PLC pathways, intimately associated with a multitude of psychiatric conditions, including depression, Parkinson's disease, and schizophrenia. Like many other RTKs, due to their variable transmembrane topology and molecular mass outweighing the upper limit of conventional resolution approaches, only fifty percent of TrkB sequences have solved. Several RTKs, including EGFR, INSR, and ALK, have accurate measurements of the extracellular segment thanks to recent methodological enhancements in Cryo-electron microscopy (Cryo-EM). However, the membrane-spanning and cytoplasmic sections of all ligand-binding multimers were presented in low resolution due to IDRs. The dearth of architectural information has sparked debate on a variety of hot-button issues, along with the following: Why the extracellular TrkB segment reacts to distinct NGF-family members having diverse biological responses? Does TrkB's transmembrane helix bridging angle alter the on/off state and strength of its subsequent enzymatic activity, and is it a feasible curative landing point?How does TrkB's intracellular kinase element engage over 140 chaperones to assemble signal gatherings and automatically disentangle post-phosphorylation?Only a few layouts of RTKs bound to key partners were fixed since resolution entails a rigorous selection of docking sequences, kinase phosphorylation phase manipulation, and crystallographic strategies. AF2, a deep learning-based artificial intelligence algorithm, predicted and released the structures of all protein sequences, making them easily accessible before resolution. Previously, the GPCRs and ATP-binding cassette receptors were correctly identified as the experimental structures adopting AF2. RTKs predicting topologies are rarely utilized because of their trisecting topologies and characterization of dimer activation. As a result of AF2 advancements, biologists can construct biochemically active aggregates utilizing the protein complex prediction algorithm AF-M. Even if TrkB is unresolved, AF-M can build its signal assemblies, and clear out if the TrkB dimeric transmembrane helix's architecture does bind to ADs.

In the first step, utilizing AF2, the full-length monomer of TrkB was built with atomic-level precision. Due to IDR, the full-length dimeric structure of TrkB was not

topology-compliant. To tackle this problem, the TrkB sequence was divided into extracellular, transmembrane, and intracellular segments. TrkB extracellular portion relating to mBDNF has predicted, and it is consistent with the resolved structure possessing the similar intrinsic and specific ligand-binding regions. Creating the model of TrkC linking NTF3 indicates that AF-M rapidly pair between homologous receptors and ligands. The structural information help design microproteins with regulatory activity for a specific pair. In physiological membrane environment, the transmembrane helix dimers maintained stable, and binding pocket for the novel ADs: IHCH-7086 and (2R, 6R)-HNK was localized in the helix crossing point. This discovery expands the list of pharmacological targets from GPCRs and ion channel-type receptors to include RTKs. AI detected the essential posture of the assemblies during phosphorylation using related signal substrates like PLCγ1, SHP2, and SHC1. In conclusion, the TrkB's snapshots during the signal cascade have been effectively reproduced using AF-M. Networks of physiological or pathological signal pathways comprising spatiotemporal information will be built in computers with artificial intelligence. In the next generation, AI will provide crucial points of the unresolved protein architectures associated with pathologies, broadening the scope of structure-based drug discovery to all coding sequences.

2 Methods

2.1 AF2 Predicts the Full-Length Structure of TrkB

To predict the full-length structure of TrkB, the first one was obtained directly from the AF Protein Structure Database (Last updated in Alphafold database, version 2022–06−01, created with the AF Monomer v2.0 pipeline). The second one is based on AF Monomer V2.2, manually inputting the full-length sequence of TrkB obtained from uniport. The third one is based on the monomer_casp14 program, which improves the average GDT of Monomer by about 0.1. AF2 was downloaded from github and run locally as described (https://github.com/deepmind/AF), using default parameters and the database version used: pdb _mmcif and uniport are 2022–08−03, the rest of the database is the default database.

2.2 AF-Multimer Predicts the Structure of Protein Complexes

Human full-length sequences of TrkB, BDNF, SHP2, PLCγ1, and SHC were obtained from uniport, and sequences were selected for combination to construct multimers as needed, and the sequence combinations used are shown in the Supplementary Material. Run the AF-multimer program and set '--model_preset = multimer' to output the PDB file of the top 5 predicted complex structures sorted by PLDDT. All raw structures not shown are shown in the supplementary material.

2.3 Conformational Optimization of Drug Small Molecule Drawing

The (2R, 6R)-HNK, IHCH-7086 structures were drawn using ChemDraw, calling the Chem.AllChem module of RDKi (http://rdkit.org) using the Embed Molecule function using Experimental-Torsion Basic Knowledge Distance Geometry (ETKDG) algorithm

to generate 3D conformations based on the modified distance geometry algorithm, optimize and calculate the energy using MMFF94 force field, and finally select the lowest energy conformation as the initial conformation for docking.

2.4 Protein-Drug Molecule Docking

The highest confidence 427–459 double alpha helix structure obtained by AF-M was taken, and Smina was selected as the docking software for molecular docking with (2R, 6R)-HNK and IHCH-7086. The positions of the active centers were as follows: X-center $= -7.729$, Y-center $= 2.684$, and Z-center $= -6.868$, where the approximation (exhaustiveness) of docking was 80, the box size of docking was 40 Å, and 80 Å conformations were generated each time, and the optimal conformations were selected for molecular dynamics simulations.

2.5 Analysis of Protein-Ligand Interactions

Upload the pdb file of protein complex to PILP for interaction analysis, set the A chain as the main chain, obtain the salt bond, hydrophobic bond, $\pi\pi$ bond, etc. generated with the interacting ligands, and visualize the output pse file with pymol.

2.6 Sequences Alignment

Clustal Omega is used to perform multiple sequence alignment, input the sequence information obtained from uniport, and set the output form as ClustalW with character counts. Visualizing the comparison results using Jalview. The higher the similarity, the darker the color of the residues.

2.7 Molecular Dynamics Simulation

The conformation of ranked 1 transmembrane helical dimer coupled with (2R, 6R)-HNK predicted by AF-M was added with 20% CHOL + 80% DOPC as the membrane environment. Gromacs2019.6 was chosen as the kinetic simulation software and amber14sb as the protein force field. Small molecules were used to produce topology files based on GAFF2 (Generation Amber Force Field) force field. The TIP3P water model was used to add TIP3P water model to the complex system to create a water box and add sodium ions to equilibrate the system. Under elastic simulation by Verlet and cg algorithms respectively, PME deals with electrostatic interactions and energy minimization using steepest descent method for maximum number of steps (50,000 steps). The Coulomb force cutoff distance and van der Waals radius cutoff distance are both 1.4 nm, and finally the system is equilibrated using the regular system (NVT) and isothermal isobaric system (NPT), and then the MDS simulations are performed for 100 ns at room temperature and pressure. In the MDS simulations, the hydrogen bonds are constrained by the LINCS algorithm with an integration step of 2 fs. The Particle-mesh Ewald (PME) method is calculated with a cutoff value of 1.2 nm, and the non-bond interaction cutoff value is set to 10 Å. The simulation temperature is controlled by the V-rescale temperature coupling

method at 300 K, and the pressure is controlled by the Berendsen method at 1 Å. NVT and NPT equilibrium simulations were performed at 300 K for 30 ps, and finally, the finished MDS simulations were performed for 50 ns. The root mean square deviation (RMSD) was used to observe the local site variation during the simulation (the cut-off point was set to 0.2). The radius of gyration (Rg, radius of gyration) is used to evaluate the tightness of the structure of the system. The root mean square function (RMSF) is used to observe the local site metastability of the system during the simulation, and the solvent accessible surface area (SASA) is used to observe the size of the solvent accessible surface area of the complex during the simulation.

2.8 Binding Free Energy Calculation for Proteins and Small Molecules

The MDS trajectory is used to calculate the binding free energy by the following equation:

$$\begin{aligned}\Delta G_{bind} &= \Delta G_{complex} - \left(\Delta G_{receptor} + \Delta G_{ligand}\right) \\ &= \Delta E_{internal} + \Delta E_{VDW} + \Delta E_{elec} + \Delta G_{GB} + \Delta G_{SA}\end{aligned}$$

In the above equations, $\Delta E_{internal}$ internal represents internal energy, ΔE_{VDW} represents van der Waals interaction and ΔE_{elec} represents electrostatic interaction. The internal energy includes the bond energy (Ebond), angular energy (Eangle), and torsion energy (Etorsion); ΔG_{GB} and ΔG_{GA} are collectively referred to as the solvation free energy. Among them, GGB is the polar solvation free energy and GSA is the nonpolar solvation free energy. For ΔG_{GB}, the GB model is used for calculation (igb = 8). The nonpolar solvation free energy (ΔGSA) is calculated based on the product of surface tension (γ) and solvent accessible surface area (SA), GSA = 0.0072 × SASA. The entropy change is neglected in this study due to the high computational resources and low precision. This study is neglected. This algorithm is implemented by Gmx_MMPBSA.

3 Result

3.1 AF2 Predict the Full-Length and Dimeric Structures of TrkB

- Predicting the full-length structure of TrkB monomers by three methods using AF2

The fixed structures predicted by the three modalities were similar, TrkB consists of five major structures, Domain 1: two CR clusters sandwiching three LRRs, Domain 2: Ig C1, Domain 3: Ig C2, Domain 4: transmembrane α-helix, and Domain 5: kinase domain (Fig. 1A). However, the main difference between the predicted results could be attributed to the disordered sequences on both sides of the transmembrane helix and at the N and C-terminals. The PLDDT values of these sequences are shown to be low and as orange noodles by AF2. The disordered sequences are freely distributed, resulting in the predicted extracellular, transmembrane, and intracellular relative positions of TrkB not conforming to their spatial distribution and separated from each other at the plasma membrane.

- AF-M predicts the full-length dimer structure of TrkB without ligand/binding ligand and predicts the spatial distribution of the structure irrationally

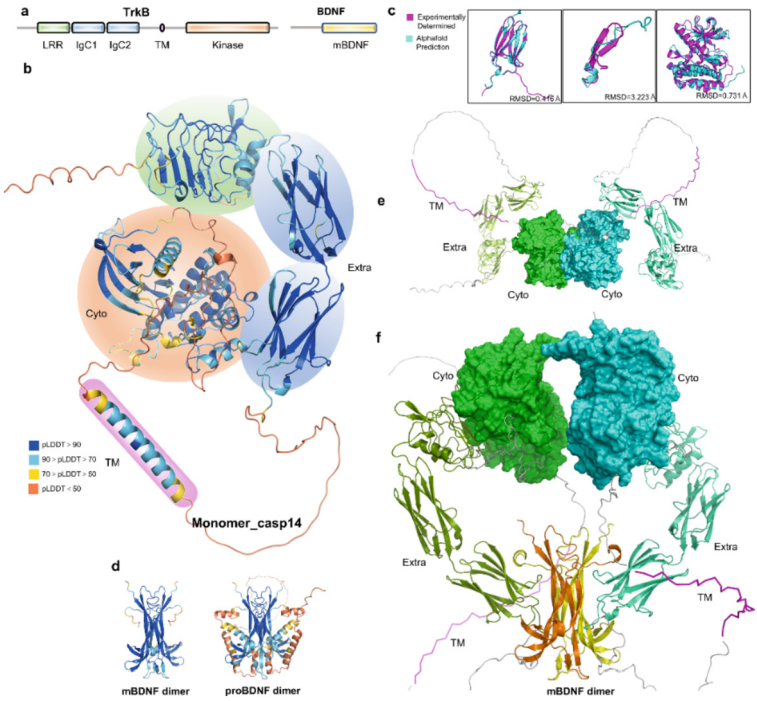

Fig. 1. TrkB monomer, dimer structure predicted by AF2. (A) Schematic diagram of the structural domains of TrkB and BDNF. (B) The full-length structure of TrkB predicted by casp14, with different interval confidence levels displayed in the structure in the corresponding colors. Fixed structures are circled and colored in line with (A). (C) Structural alignment between the AF predicted structure and the experimentally resolved structure, in order: IgC2 [PDB:1WWB], juxtamembrane IDR [PDB:2MFQ], kinase domain [PDB:4ASZ], the RMSD between the predicted structure and the experimental structure is marked in the lower right corner, respectively. (D) mBDNF and proBDNF homodimeric structures predicted by AF-M. (E) The ligand-free dimeric structure of TrkB predicted by AF-M does not possess a reasonable membrane topology. (F) AF-M predicted structure of the activation dimer of TrkB-binding mBDNF. (Color figure online)

AF-Multimer, used for protein complex prediction, has been released recently and has demonstrated excellent performance in constructing dimers and multimers. It has been used for dimer construction of TrkB to understand the usability of AF2 for RTKs, which have dynamic conformations and a wide range of binding partners and complex binding forms. The structure of TrkB has not yet been resolved at high resolution under Cryo-EM, and AF2 does not allow direct access to experimental dimer structures. Nevertheless, the extracellular dimer and transmembrane dimer of the TrkB homologous protein TrkA have been resolved to use it as a reference for the predicted structure of TrkB to evaluate the availability of AF-M predictions.

The homodimers of the two forms of BDNF were constructed independently (Fig. 1D). mBDNF dimer is highly similar in structure to the already resolved mBDNF/NTF-4 heterodimer 55. The main body of mBDNF consists of three pairs

of β-strands and four linked β-hairpin loops, and the dimer is centrosymmetric with the long axes of two pairs of long β-strands contact (Fig. 1D). In contrast, the structure of the proBDNF dimer has not yet been resolved, and the predicted structure has four incomplete helices at the N-terminal of each monomer. The presence of these helices caused the lower end of the β-strands to be pulled to both sides, which disrupted the intrinsic binding interface of the NGF-β family at the IgC2 of the Trk family. This alteration revealed the reason for the low affinity of proBDNF to TrkB and high affinity to p75, providing a structural explanation for the opposite synaptic effect of proBDNF and mBDNF. Consequently, AF predicted mBDNF homodimers, demonstrating that it is highly usable in the prediction of dimeric ligands and can rapidly predict all the putative intrafamily paired dimer conformations when homologous family dimeric structures are available.

Based on the successful construction of mBDNF dimer, we further predicted the conformation of TrkB full-length dimeric activation state, TrkB-mBDNF-mBDNF-TrkB. The major dimerization interface of all predicted results was the extracellular segment, and the kinase dimerization structure in the ligand-free dimerization structure was separated, indicating centrosymmetric but no contact. In ranked 1 as an example: mBDNF bridges the bipartite structure of the extracellular segment, while the kinase segment is incorrectly placed at the top of the N-terminal of the extracellular segment and does not contact each other (Fig. 1F). The transmembrane helix is recognized as a disordered structure, indicating that the addition of mBDNF does not rescue the irrational spatial distribution of TrkB dimeric conformation predicted by AF-M.

In conclusion, we found that AF-M does not provide a reasonable full-length dimeric structure of TrkB with or without binding ligands. This phenomenon is attributed to the fact that the resolution results of RTKs Cryo-EM could not provide a sufficient density for the transmembrane and intracellular segments due to the lack of rigid connections between the extracellular segment and transmembrane helix, and only the dimeric extracellular segment was resolved at high resolution and uploaded to the database. Despite the inability to give a full-length dimeric structure, AF-M provides a dimeric conformation of the kinase segment and the extracellular segment of the bound ligand, derived from the mimicry of the local structures of RTKs obtained by NMR and X-ray over the past 30 years. This indicates that AF-M has the ability to predict the fixed structure of RTKs undergoing dimerization. Hence, we split the sequence of TrkB into extracellular/transmembrane helix/intracellular segments for dimerization prediction based on previous studies of EGFR and predicted their complex conformation upon binding ligands, signal components, or drugs.

3.2 AF-M Accurately Predicts the Ligand-Binding Structure of the Extracellular Segment of the TrkB Dimer

- Prediction of potential TrkB extracellular ligand-free binding dimer structure with inter-monomeric binding via β-turn

The predicted TrkB extracellular ligand-free dimer was obtained by AF-M. The docking of ranked 1–ranked 3 occurs at Domain 3, while the difference between the

structures is that the two monomers Domain 3 are increasingly spaced apart with decreasing confidence until no contact is made; the docking of ranked 4–5 occurs at Domain 1 and IDR. These findings differed from the previously experimentally obtained dimeric structure of Trk family Domain 3, wherein the crystal structure is based on the two monomers with overlapping N-terminus and C-terminus, which results in the loss of βA at the N-terminus of Domain 3. Such a structure is considered erroneous because Domain 2 exists on Domain 3 N-terminal, and the N-terminal sequence serves as a bridge between Domain 2 with 3. Despite the incorrectly constructed Domain 3 dimer in the database, AF-M can determine and correct its implausibility. In ranked 1, βA was successfully identified and linked to Domain 2, and the monomers contacted each other through the loop between βA and βB (ABL), i.e., the interaction force of S297-W301 (Fig. 2A and 2B). A total of five hydrogen bonds and two hydrophobic bonds were formed between the residues, and a π-π stacking was formed between H299 and W301. These residues formed a negatively charged finger-like protrusion and a concavity, and the two monomers formed a chimeric structure with each other, which was consistent with the subsequent prediction of a specific ligand-binding region (Fig. 2C). Furthermore, M379 and G380 near the membrane were also identified as bind residues in the disordered region. AF-M offers the possibility of an unresolved TrkB pre-dimer, i.e., ABLs that contact Domain 3 specifically with ligands forming chimeras with each other, and the ligand-binding surface is buried until the ligand is inserted and opened to stimulate the downstream signals.

- Predicting the ligand-bound dimeric structure of TrkB extracellularly with specific binding region and intrinsic binding region between Domain 3 and mBDNF

Next, AF-M was used to predict the conformation of the extracellular segment of TrkB bound to mBDNF. The five results showed a high degree of agreement, with differences arising from the free distribution of the disordered region. The complex structure of mBDNF retains its centrosymmetric homodimeric conformation and binds to the Domain3 of TrkB. Conversely, the extracellular segment of TrkB has a crabpincer shape with two monomers each monomer is attached to a single chain, alternating between the front and back of mBDNF. This finding is similar to the results of the TrkA extracellular segment. The most different secondary structure is Domain 1, wherein β1–β3 starting from the N terminus is shorter in the crystalline structure than in the predicted structure, and most of the articulated sequences have a loop-like morphology (Fig S5B). These results predicted a TrkB Domain 1 superhelical topology compared to the resolved TrkA.

In conclusion, in the structure prediction of extracellular segment-binding ligands, AF-M is unable to predict the effect of PTMs on the structure, resulting in a relative angle between domains deviating from the true structure. However, the predicted structures are highly accurate for the ligand-bound receptor structural domains, constructed with reference to the experimental structures between ligand-same family receptor members of the same family in the database. Previous studies have shown that several RTKs can form heterodimeric pairings with homologous receptors. Also, ligand family members can heterodimerize, with dozens of possible pairings between the Trk family and the NGF-β family alone. AF-M facilitates the assembly of all the ligand-receptor binding structures using only the sequences, thereby circumventing the limitation of the resolved structures. Thus, we can retrieve highly matched design structure microproteins and

small-molecule drugs using the three-dimensional structural information of proteins, thereby minimizing the side effects of the traditional antibody binding to the same family of receptors and regulating the activation and inactivation of receptor kinases at the atomic level.

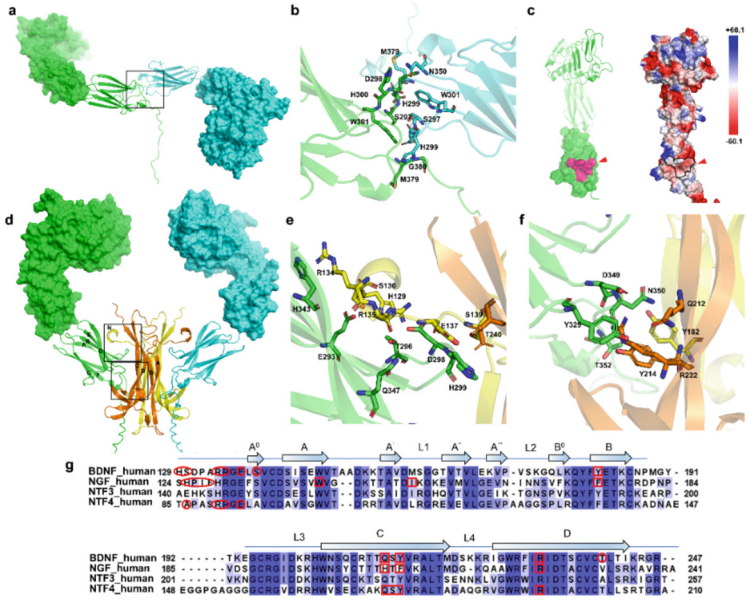

Fig. 2. Structure prediction of pre-activated and activated dimer of TrkB extracellular segment. (A) AF-M prediction of the potential TrkB extracellular segment pre-activated dimer structureb. (B) Detailed graphical representation of the contact residues of the pre-activated TrkB pre-dimer. (C) TrkB pre-dimer contact residues form finger-like protrusion and concavity, marked in purple (left). Circled in black stroke on the surface potential energy map (right). (D) AF-M prediction of the extracellular activation state dimer of TrkB that binds mBDNF. The two interact regions are framed in black. (E) Residue details of the TrkB-binding mBDNF-specific binding region are shown. (F) Detail map of residues in the intrinsic binding region of TrkB-binding mBDNF is shown. (G) Sequence alignment of NGF-β family, where residues in the specific binding region are circled and residues in the intrinsic binding region are boxed. Information of BDNF is from the predicted structure of TrkB-mBDNF. Information of NGF NTF4 is from the resolved structure of TrkA-NGF and TrkB-NTF4.

3.3 AF-M Predicts the Antidepressant Binding Pocket at the Crossover of TrkB Transmembranesegment Dimer

- The predicted TrkB transmembrane segment dimer structure is significantly different from the resolved structure of TrkA.

The transmembrane helix dimer of TrkB was successfully constructed by AF-M (Fig S6A). The transmembrane helix dimerization structure of TrkA has been resolved

previously. The predicted transmembrane helix crossover pattern of TrkB is significantly different from that of TrkA, with TrkA monomers further apart from each other and the overall structure similar to X-type. On the other hand, the structure of TrkB tends to be parallel, and the helix of TrkB is longer than that of TrkA, with the crossover site ASVVG located at the center of the helix, while the crossover site of TrkA is near the N-terminal. Sequence comparison revealed that the crossover site SXAVG of TrkA and TrkC was shifted up by 7 residues compared to TrkB, indicating that the crossover site of TrkB is highly specific in the Trk family (Fig. 3D). The sequence comparison of TrkB between different species revealed that the transmembrane helix is conserved in the family. While the transmembrane helix is a common drug binding site for GPCRs and ion channel-type receptors, the complex formation structures of various antidepressants with the 5HT family and the glutamate receptor family have been resolved. Previous studies on EGFR have shown that the transmembrane helix transmits signals from the extracellular segment to the intracellular segment by changing the rotation angle and docking site. This alteration implies that the transmembrane helix of TrkB has the potential to modulate the activation strength of the intracellular kinases and becomes a target structure for antidepressants.

- Molecular dynamics simulations reveal a potential antidepressant binding pocket at TrkB docking.

The transmembrane helix structures predicted by AF-M were derived from the imitation of similar structures in the database. The AF-predicted transmembrane helix structures have been demonstrated to be highly accurate based on the comparison between the GPCR and ABC proteins with the resolved structures. The five TrkB transmembrane helix dimerization structures predicted by AF-M differed only in relative angles, with consistent crossover sites, indicating that TrkB dimers are formed in cholesterol-rich lipid rafts. To demonstrate the stability of the predicted structures in the membrane environment, the predicted ranked 1 dimer helix was placed in a 20% CHOL + 80% DOPC environmental box, and molecular dynamic simulations were performed for 100 ns, and the dimer structures were found to be cross-stable until the end of simulation.

In order to obtain the interaction of the drug with the transmembrane helix in the physiological membrane environment, the (2R, 6R)-HNK docked dimeric transmembrane helix structure was subjected to MDS in 20% CHOL + 80% DOPC environment (Fig. 3C). We found that the transmembrane helix showed a tendency to converge and reach a steady state in the second half of the simulation. The RMSD values fluctuate from 0–50 ns, which is caused by the initial instability of the docked acquired conformation inside the box, which declines rapidly at 50 ns and enters the steady state at 60 ns. The fluctuation of Rg corresponds to the ripple correspondence in RMSD, wherein the fluctuation rises in the first period, has a small peak at 10 ns, starts to reach a second peak at 50 ns and finally enters the stable zone at 75–100 ns, indicating that the system is shifting from the unstable to the stable state at this moment. The SASA of the protein gradually decreases in 0–100 ns, indicating a favorable binding and gradual protein tightening. HBNUM showed 0–2 connections of hydrogen bonding during the simulation.

Postural fingerprinting of the stable conformation revealed that the interaction of (2R, 6R)-HNK with the dimeric transmembrane helix originates from hydrogen bonding and the surrounding hydrophobic amino acids, whereby (2R, 6R)-HNK produced hydrogen

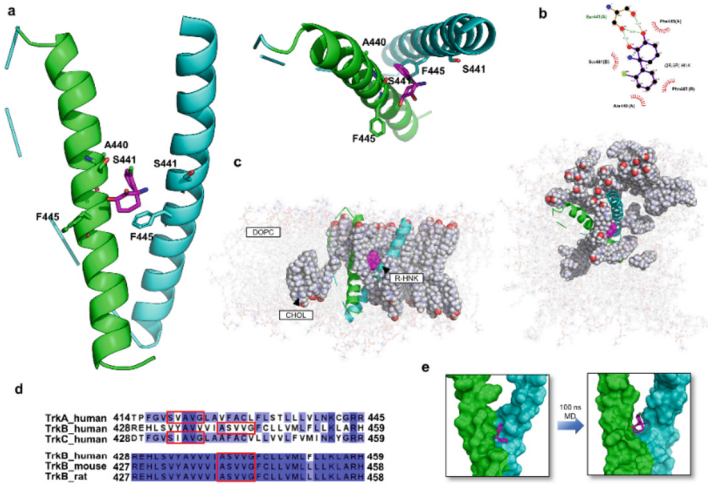

Fig. 3. AF-M predicted TrkB transmembrane helix dimer after molecular dynamics simulation to localize the drug binding pocket (A) TrkB transmembrane helical dimer with (2R, 6R)-HNK coupled structure after 100 ns MDS. Contact residues are shown in stick form. The top view is shown on the right. (B) Pose fingerprint of the (2R, 6R)-HNK contact with the surrounding TrkB transmembrane helix residues. The most stable contact occurs on Ser441, with a total of three hydrogen bonding contacts. (C) The final conformation of the TrkB transmembrane helix dimer with the (2R, 6R)-HNK-coupled structure in the lipid box, with cholesterol being displayed in a spherical shape, after 100 ns MDS simulation, the top view(right). (D) Sequence comparison of the human Trk family (top), and sequence comparison of TrkB across species (bottom), reveals that the crossover sequence is highly specific in the same family and highly conserved between species, and the helical crossover residue sequence is boxed in red. (E) After 100 ns MDS simulation, (2R, 6R)-HNK is transferred from the protein surface to the crossover gap and a drug binding pocket is generated. (Color figure online)

bonding with S441 and hydrophobic forces with V437, V438, A440, V442, and F445 (Fig. 3A). Further analysis of drug-residue interactions revealed three hydrogen bonds created between (2R, 6R)-HNK and S441 at a distance of about 3 Å (Fig. 3B). The comparing of (2R, 6R)-HNK before and after the simulation showed that the molecule moved from the protein surface to the center of the crossover, where it formed a TrkB-specific hydrophobic pocket that was exactly on the two dimer crossover sequences (Fig. 3E). This phenomenon suggested that (2R, 6R)-HNK effectuates the extracellular segment signal on the kinase segment by anchoring the transmembrane helical dimer conformation, thereby exerting a synaptic plasticity and enhancing the antidepressant effect.

3.4 AF-M Predicts Intracellular Signal Assemblies of TrkB with Multiple Binding Modes Between Kinase and Signal Proteins

- AF predicts TrkB kinase homodimers with centrosymmetry between monomers centered on the activation loop.

The kinase segments of the RTK family are structurally similar, with the main structure consisting of an N-lobe composed of α-helixes and a C-lobe composed of β-folds, as well as a hinge bridging the two globes, while a loop sequence, known as the activation loop, exists on the C-lobe and plays a critical role in the phosphorylation cascade. For TrkB, there are five majors phosphorylated tyrosines (pTyr), including Y516 in the proximal membrane region recognized by AF as an incomplete helix, Y702, Y706, Y707 in the activation loop, and Y817 in the C-terminal IDR tail. The predicted TrkB kinase segment shows a consistent conformation with the resolved kinase-inactive structure, wherein the αC helix opens outward and the expanded binding pocket appears as C-helix out; the benzene ring of Phe in the DFG at β8 appears as DFG-out towards the hinge domain, indicating that the protein is inactivated and is in a closed state. Despite the presence of the DFG-in structure of CPD5N [PDB:4AT3] with the antagonist in the database, AF2 tends to select the inactive state of the kinase for output. Among all subsequent structures of the complex, the kinase presents an inactive conformation consistent with the monomer (Fig. 4B).

Furthermore, three models of self-inhibition have been identified in studies of RTKs, including cis-inhibition of the active site by the proximal membrane region and C-terminal tail, tightening of the activation loop, and blocking of the substrate binding site. In the predicted structure of the TrkB monomer, the juxtamembrane region is identified as a disordered scattered structure, the C-terminal tail has only 10 residues and does not block the ATP-binding site, and the activation loop shows an outwardly expanded relaxed structure. Therefore, AF did not provide a monomeric TrkB self-suppressive structure.

- Predicted TrkB kinase segment complex structure with PLCγ1 and SH2 structural domain bound to the kinase C-terminus

Tyrosine kinases transmit signals downstream by recruiting and activating substrate proteins. pTyr binds to the SH2 and PTB structural domains of the signal proteins to generate cascade phosphorylation. The SH2 and PTB structural domains are present in a diverse set of proteins containing a range of catalytic type and interacting domains, which provide a degree of specificity by recognizing pTyr residues and the surrounding residues. Three extensively studied signal proteins were selected for the construction of their complex structures with TrkB kinase segments to examine the efficacy of AF-M in predicting the intracellular molecular assemblies of tyrosine kinases. Notably, the AF was previously shown to be effective in the prediction of intracellular giant complexes, with quiescent structures. However, the tyrosine kinase signal is transmitted by a liquid-liquid phase separation mechanism, and the protein structure in the droplet exhibits a high degree of dynamics.

The PLCγ signal pathway is downstream of TrkB, which plays a major role in promoting calcium inward flow. The activation of the pathway leads stimulates CaMKII and subsequent synaptic plasticity, where PLCγ1 serves as the first substrate with two tandem SH2 domains. The previous structural analysis provided two reference binding modes; one from the binding of NSH2-CSH2 to FGFR1, in which NSH2 forms a binding pocket on pTyr at the C-terminus of the kinase 31, and the other is from the complex formed by FGFR2 and CSH2. Based on this structure, another activation model was proposed as follows: CSH2 contacts both kinases simultaneously, and in addition to

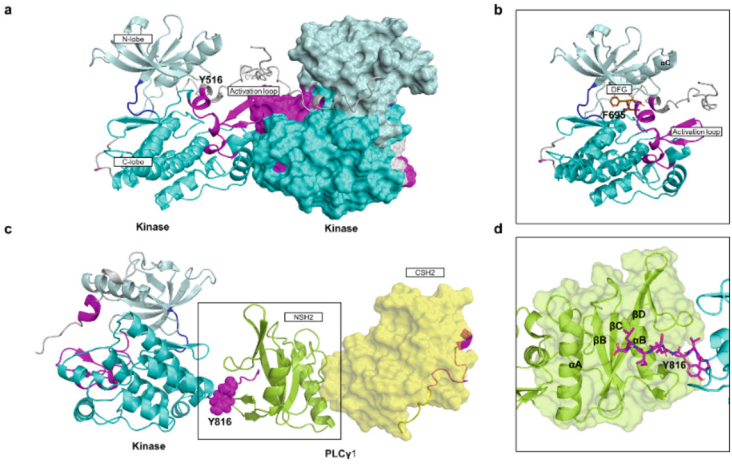

Fig. 4. AF-M predicted TrkB intracellular kinase segment homodimer, heterodimer bound to PLCγ1 NSH2-CSH2 (A) AF-M predicted homodimer of TrkB kinase segment. The two monomers are centrosymmetric with the active ring as the central point, and the active ring is close but not in contact. The proximal membrane region crosses the interface where the active ring is located. The hinge region is dark blue, and the proximal membrane region containing the autophosphorylated tyrosine, the active loop, and the C-terminal tail are purple. Same below. (B) AF-M predicted sequences related to kinase activity in TrkB kinase segment homodimers are shown. αC is C-helix out, DFG is DFG-out, and the kinase is in the off state. (C) AF-M predicted heterodimer of TrkB kinase segment with PLCγ1 NSH2-CSH2 structural domain, Y816 site in contact with NSH2, shown as spherical. (D) Details of the contact of the TrkB kinase segment with the PLCγ1 NSH2 structural domain are shown the C-terminal tail is embedded in the surface groove of NSH2 and is localized near αB. (Color figure online)

producing a pocket binding mode similar to that described above, the pTyr-containing tail at the C-terminus of CSH2 is inserted into the active groove of the other kinase. These two models provide insights into two types of kinase-substrate binding: 1. Substrate recruitment and phosphorylation are cis at the same tyrosine kinase monomer. 2. Substrate recruitment and phosphorylation are dependent on the dimerization of the kinase segment. Consequently, we selected NSH2-CSH2 of PLCγ1 and retained the pTyr-containing sequence downstream of CSH2 to construct the complexes of NSH2-CSH2 with TrkB kinase segment in a 1:1 ratio and CSH2 with kinase segment in a 1:2 ratio.

This finding indicated that AF-M predicted the signal chaperone contact domain and located the phosphorylation site in the case where the database contains the structure of the kinase complex with the signal companion. However, some differences were noted in the spatial distribution of specific residues from the resolved structure owing to sequence differences among the kinases and the additional possibilities offered by the database containing the pTyr peptide in a complex with the SH2 structural domain. Moreover, in the 1:2 complex, AF did not construct the theoretical model of transdimeric activation. CSH2 in all predicted structures was free from the dimerization of the kinase segment,

indicating that despite the existence of the theoretical model, AF-M could not construct an approximate linkage due to the lack of reference structures in the database.

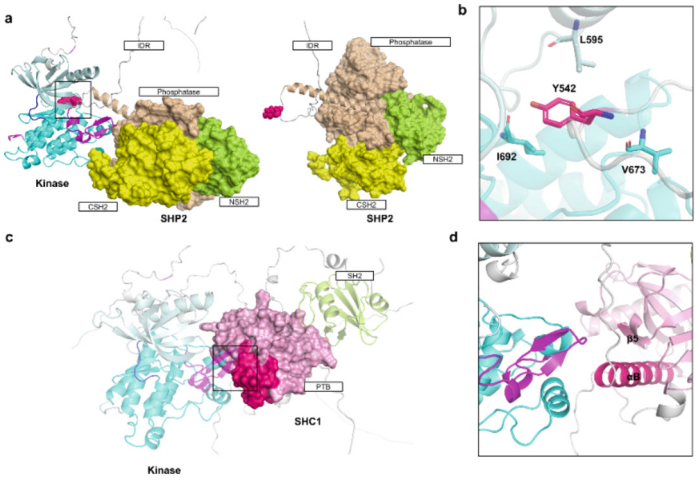

Fig. 5. AF-M predicted heterodimer of TrkB intracellular kinase segment bound to SHP2 and SHC1 (A) AF-M predicted heterodimer of TrkB intracellular kinase segment SHP2, with the structure of the disordered region of SHP2 extending into the active groove between the kinase globes. (B) Detail of AF-M predicted heterodimer of TrkB intracellular kinase segment SHP2 showing that autophosphorylated Y542 on the disordered region of SHP2 makes contact with I692, V673 and L595 of αC of C-lobe. (C) AF-M predicted heterodimerization of the TrkB intracellular kinase segment SHC1 with contact occurring in the disordered region while PTB is closer to the C-lobe compared to SH2. No contact occurs between PTB and the kinase, but the binding pocket of PTB (red) is toward the C-lobe. (D) Detail of AF-M predicted heterodimerization of the TrkB intracellular kinase segment SHC1, with β5 and αB of PTB producing binding pockets toward the C-lobe and close to the activation loop. (Color figure online)

- Prediction of the complex structure of TrkB kinase segment with SHP2 and SHC1 and identification of the different binding modes of kinase binding to SHP2 disordered region and SHC1 PTB

To further understand the effect of AF on the formation of kinase-signal protein complexes, two key signal proteins were selected to construct the dimer formed with kinase, including SHP2 and SHC1. SHP2 is an allosteric enzyme that consists of two tandem SH2 structural domains, phosphatase structural domain and C-terminal disordered tail, while NSH2 contacts with PTP and self-inhibits the phosphatase activity in the absence of biochemical reactions. In addition, SHC1 is a signal scaffold protein, and its PTB binds to the kinase structural domain, while SH2 further binds other signal proteins and large segments of the sequence show a disordered structure. Different from PLCγ1, neither SHP2 nor SHC1 obtained a resolved structure of its structural domains in a complex with the kinase segment, and the inferred contact site was derived from the contact conformation of the pTyr-containing polypeptide with SH2 and PTB.

The predicted structure of the TrkB kinase segment with SHP2 does not provide a conformation in which the PTP is activated by conformational change; however, it still adopts a self-inhibitory structure, and NSH2-CSH2 does not interact with the kinase (Fig. 5A). A phosphorylated peptide study showed that the spaced pTyr site sequentially crosses the surface of SH2 and causes NSH2 to lose the inhibition of the phosphatase; then, the kinase contacts and reacts with the main body of the phosphatase as the conformation during the allosteric activation is not available due to the absence of PTM information from AF-M. The contact in the predicted structure occurs in the C-terminal α-helix of the PTP structural domain of SHP2 and downstream to IDR, as it is embedded in the active groove of the kinase forming a multivalent and stable contact, where Y547 of SHP2, the phosphorylated site creates four hydrophobic bonds and one hydrogen bond between L595, I692, and V673 of the kinase segment (Fig. 5B). This interaction exhibits a transient local contact where SHP2 is recruited to the C-lobe as well as near the activation loop through the disordered region in the case of autoinhibition. Of the five possible conformations, four of the resulting contacts occur in the C-terminal disordered tail of SHP2. This interaction is consistent with recent studies on the liquid-liquid phase separation of SHP2, which undergoes LLPS in the autoinhibited state and is enhanced by the transformation of the phosphatase to an active conformation after mutation. Before this phenomenon, AF2-based phase separation prediction programs have been developed by scoring disordered regions, while AF-M has the potential to complete the screening and localization of phase separation targets.

In conclusion, we explored the potential conformation of AF-M in predicting the intracellular kinase segment dimerization and kinase segment binding to various key signal proteins. In the case of resolved structures with similar patterns, AF-M could accurately localize to the binding residues, such as NSH2 of PLCγ1. However, the monomer in the unresolved complex exhibits a self-inhibited structure with the blocked binding interface, and the predicted contacts occur at the IDR because the PTM information is not considered. In order to construct a complete, signal pathway network containing three-dimensional information is necessary to deduce additional conformations of kinases and companions at different reaction stages and to correlate the PTM information of the residues with these structures. On the other hand, it is essential to concatenate the previous experimental evidence and construct a three-dimensional model for artificial intelligence learning.

4 Discussion

MDD is a pressure stress-mediated physiological disorder that often leads to reduced synaptic plasticity, in which the TrkB receptor, the starting point of the synaptic plasticity pathway, is an RTK distinct from 5-HT and glutamate receptors 11. Its large extracellular receptor and intracellular kinase structures, as well as the highly variable transmembrane helices and disordered sequences between the two, have led to the resolution of its activation state being mostly limited to the extracellular segment. Following AF2 catch all protein in one draft, AF-M evolved to be able to predict complex structures, allowing us to construct TrkB dimerization activation structures as well as complex structures with intracellular signal molecules. Adopting a divide and conquer algorithm, we first

predicted the activation structure of the whole extracellular segment of TrkB binding to mBDNF and were able to rapidly perform homologous ligand-receptor pairing, providing three-dimensional information that can be used to design targeted agonists or antagonists that are highly specific to a given combination. For example, the design of mBDNF-mimetic microproteins targeting TrkB reduces the aberrant activation of TrkA and TrkC.

Transmembrane helices, due to their complex environment and their role as relay stations for signal cascades, often require careful design of membrane mimics for resolution. For unresolved transmembrane helix dimers, manual screening is required among numerous docking results. AF-M provides dimeric structures that are stable in the lipid raft environment and create pockets at the crossover that bind to novel antidepressants. This will strongly facilitate the design of drugs against RTKs, breaking the inherent impression that transmembrane helical drug targets exist only for GPCRs and ion channels with multiple transmembrane helices. It will also enhance the resolution of highly flexible dimeric transmembrane helices of RTKs under Cryo-EM and refine the transmembrane dimeric activation pattern of RTKs members in combination with MDS.

As a kinase segment that functions as a biochemical reaction, its potential chaperones exceed 300 species, and the complex conformation is currently difficult to capture due to the rapidly proceeding phosphorylation reaction. Attempts were made to construct a complex pattern of TrkB kinase segment with three typical signal chaperones, and the binding site and pattern of PLCγ1 were highly similar to that of the resolved FGFR2. Unfortunately, however, the binding site of SHP2 presenting a self-repressed structure was designated as being in the disordered region due to the absence of PTM information. Rather, SHC1 successfully distinguished a binding mode with the PTB structural domain binding pTyr. Further iterations of AF-M are needed in the prediction of kinase-signal chaperones to predict the metastable binding process of proteins in biochemical reactions, such as researchers adding PTM information when uploading structures, marking key sites regulating metastable conformations, and uploading multiple dynamic conformations of the binding process to the database for artificial intelligence learning.

AF-M extends the prediction to multimers, giving us access to the underlying structures of molecular machine-protein complexes that dominate physiological functions. Such breakthroughs allow me to stand at the beginning of the next era of structural science, using deep learning-based artificial intelligence programs represented by AF2, combined with MDS, to build a dynamic signal pathway network containing structural information. In this network, the self-assembly of small molecules in all physiological and pathological processes can be demonstrated, and the operation of the entire pathway can be controlled by modifying and controlling a few key targets, which will enable drug design to shift from single-target to global. In the foreseeable future, AI, a powerful assistant, will play an important role in breaking down difficult-to-treat diseases involving multiple mechanisms, such as MDD.

References

1. Malhi, G.S., Mann, J.J.: Depression. Lancet (London, England) **392**, 2299–2312 (2018)
2. Heine, M., Holcman, D.: Asymmetry between pre- and postsynaptic transient nanodomains shapes neuronal communication. Trends Neurosci. **43**, 182–196 (2020)
3. Santos, R., Ursu, O., Gaulton, A., et al.: A comprehensive map of molecular drug targets. Nat Rev. Drug Discov. **16**, 19–34 (2017)
4. Okaty, B.W., Commons, K.G., Dymecki, S.M.: Embracing diversity in the 5-HT neuronal system. Nat Rev. Neurosci. **20**, 397–424 (2019)
5. García-Nafría, J., Nehmé, R., Edwards, P.C., Tate, C.G.: Cryo-EM structure of the serotonin 5-HT(1B) receptor coupled to heterotrimeric G(o). Nature **558**, 620–623 (2018)
6. Kim, K., Che, T., Panova, O., et al.: Structure of a hallucinogen-activated Gq-coupled 5-HT(2A) serotonin receptor. Cell **182**, 1574-1588.e1519 (2020)
7. Cao, C., Barros-Alvarez, X., Zhang, S., et al.: Signaling snapshots of a serotonin receptor activated by the prototypical psychedelic LSD. Neuron **110**, 3154–3167 e3157 (2022)
8. Kaplan, A.L., Confair, D.N., Kim, K., et al.: Bespoke library docking for 5-HT(2A) receptor agonists with antidepressant activity. Nature **610**, 582–591 (2022)
9. Cao, D., Yu, J., Wang, H., et al.: Structure-based discovery of nonhallucinogenic psychedelic analogs. Science (New York, N.Y.) **375**, 403–411 (2022)
10. Zhang, Y., Ye, F., Zhang, T., et al.: Structural basis of ketamine action on human NMDA receptors. Nature **596**, 301–305 (2021)
11. Wang, C.S., Kavalali, E.T., Monteggia, L.M.: BDNF signaling in context: from synaptic regulation to psychiatric disorders. Cell **185**, 62–76 (2022)
12. Casarotto, P.C., Girych, M., Fred, S.M., et al.: Antidepressant drugs act by directly binding to TRKB neurotrophin receptors. Cell **184**, 1299–1313 e1219 (2021)