Synthetic and Systems Biotechnology 9 (2024) 436–444

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.keaipublishing.com/en/journals/synthetic-and-systems-biotechnology

Original Research Article

Dynamics design of a non-natural transcription factor responding to androst-4-ene-3,17-dione

Ming Zhao [1], Mengkai Hu [1], Rumeng Han, Chao Ye, Xiangfei Li, Tianwen Wang, Yan Liu, Zhenglian Xue, Kun Liu [*]

Anhui Engineering Laboratory for Industrial Microbiology Molecular Breeding, College of Biology and Food Engineering, Anhui Polytechnic University, Wuhu 241000, China

ARTICLE INFO

Keywords:
Protein design
Artificial transcription factor
Molecular dynamics
Steriods
Site-mutations

ABSTRACT

The production of androst-4-ene-3,17-dione (AD) by the steroidal microbial cell factory requires transcription factors (TFs) to participate in metabolic regulation. However, microbial cell factory lacks effective TFs that can respond to AD in its metabolic pathway. Additionally, finding and obtaining natural TFs that specifically respond to AD is a complex and onerous task. In this study, we devised an artificial TF that responds to AD, termed AdT, based on structure-guided molecular dynamics (MD) simulation. According to MD analysis of the conformational changes of AdT after binding to AD, an LBD in which the *N*- and *C*-termini exhibited convergence tendencies was used as a microswitch to guide the assembly of a DNA-binding domain lexA, a linker $(GGGGS)_2$, and a transcription activation domain B42 into an artificial TF. As a proof of design, a AD biosensor was designed and constructed in yeast on the basis of the ligand-binding domain (LBD) of hormone receptor. In addition, the transcription factor activity of AdT was increased by 1.44-fold for its variant F320Y. Overall, we created non-natural TF elements for AD microbial cell factory, and expected that the design TF strategy will be applied to running in parallel to the signaling machinery of the host cell.

1. Introduction

Transcription factor (TF) can be used as an element of biosensors to participate in the regulation of intracellular substance metabolism flux [1–3]. However, obtaining natural TFs with desired functions from natural resources is very cumbersome. TFs comprise a small part of genetic diversity, and the identification of a natural TF for a given molecule is time- and labor-intensive. In other words, discovering protein elements with specific functions is extremely burdensome, requiring strategies such as the combination of genomic screens [4] and functional assays to identify and isolate TFs [5]. High-throughput detection and screening are common methods for discovering new functional or highly active elements; however, even if high-throughput screening methods are established, the naturally occurring elements must be identified, isolated, and obtained from the genome, and these steps also require extensive time and effort.

Although it is extremely difficult to discover natural TF elements, previous research work has shown that random or saturation mutations can lead to higher specificity of LuxR [6], as well as NahR, DmpR, XylR, TetR, LysG, and AraC in response to new inducers [7–9]. The binding of substitution natural inducers has a high probability of destroying the allosteric effect, so it is difficult to change the specificity of the natural ligand of the TF. Fortunately, computational design can provide mutated TFs, George Church et al. calculated and designed to change the specificity of the inducer of LacI to obtain a response to small molecules such as sucralose and trehalose without destroying the allosteric effect of LacI [8]. Another example, the molecule-binding sites of the proteins were redesigned based on shapes complementary, and this design was used to generate activators of transcription of the steroid digoxigenin [10]. Nevertheless, current methods for designing ligand-binding proteins rely on the semirational design and/or directed evolution of proteins, and a computational design has not been developed to the extent that it can solve completely different challenges. In essence, the difficulty of re-engineering natural TFs has limited the development of a number of

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding author.
E-mail address: liuk@ahpu.edu.cn (K. Liu).
[1] Ming Zhao and Mengkai Hu contributed equally to this work.

https://doi.org/10.1016/j.synbio.2024.04.001
Received 5 November 2023; Received in revised form 3 March 2024; Accepted 2 April 2024
Available online 6 April 2024

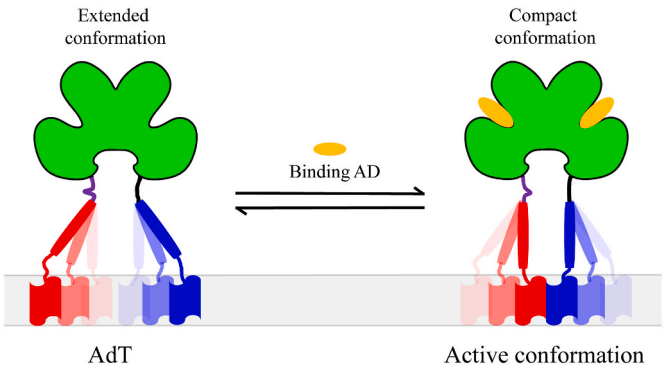

Fig. 1. Schematic of activation model for AdT. During the activation, AD was transported into cells to bind to extended conformation AdT, which led to a compact AdT conformation with activity. Red: lexA, Purple: linker, Green: LBD, Blue: B42, Yellow: AD.

required TFs. David and co-workers proposed the concept of *de novo* protein design and successfully designed a series of non-natural proteins, such as biosensors for the SARS-CoV-2 spike protein [11,12]. Although the researchers did not obtain desirable protein structural and functional characteristics through the contiguous stacking of amino acid residues from "Met", the use of an energy-minimum design has dramatically increased the possibility of *de novo* protein design.

A TF contains 3 characteristic domains: DNA-binding domain, binding domain of small molecule effectors and transcription activation domain. The design of non-natural TFs based on the modular assembly of protein functional domains may achieve the expected goals. Molecular dynamics (MD) is a commonly used method for computer simulation, and it has a wide range of applications in protein structure analysis, functional mechanism, and dynamic properties, such as protein folding, ligand–protein interactions, and dynamic change processes [13–15]. MD can supplement experimental phenomena to explain the mechanism of occurrence of phenomena. A higher level is to predict the occurrence or results of biology based on MD simulation to assist in experimental design and develop a feasible experimental plan. MD simulation has been developed into a mature technique that can be effectively used to understand the structure–function relationships of biomacromolecules. Such as, the effect of substrate–peptide interaction on the conformational kinetics of LptDE to clarify the mechanism of lipopolysaccharide translocation [16]. MD was performed to clarify the molecular mechanisms of puerarin and acetylcholinesterase and identify puerarin analogs as inhibitors of acetylcholinesterase [17]. TF was designed based on structure-guided MD simulation for progesterone biosensor [14]. In addition, MD simulations of *Caenorhabditis elegans* TF SKN-1 bound to its DNA site suggested that movements of two key Arg side chains between the major groove and the backbone of DNA generate distinct conformational sub-states [18]. These strategies of designing protein elements by MD might be simpler and less laborious.

Androst-4-ene-3,17-dione (Androstenedione; AD) is a key precursor for the synthesis of glucocorticoids, saline corticosteroids, and oral contraceptives by the steroidal microbial cell factory [19,20]. However, the lack of transcription factors that respond to AD acts on the steroidal microbial cell factory, which seriously hinders the progress of the green transformation of steroidal manufacturing industry from chemical synthesis to biosynthesis [21–23]. Therefore, there is an urgent need to obtain a transcription factor that can respond to AD to regulate the metabolic flow to the target steroidal compound. To solve this challenge, we designed a yeast biosensor that responds to AD to characterize its non-natural transcription factor.

In this study, we designed a non-natural TF, termed AdT, which consisted of DNA-binding domain (DBD) lexA [24], a linker GGGGSGGGGS, the ligand-binding domain (LBD) from human hormone receptor [25], and the activation domain B42 [26]. This design concept stems from the functional cognition of the 3 characteristic domains of natural TFs. In detail, The LBD induced AdT conformational transition from extended conformation to compact conformation upon the binding of AD (Fig. 1). Specifically, if domains lexA and B42 are considered to be within sufficient proximity in space [27], AdT will exhibit TF activity. The activity mechanism of AdT was understood using MD to simulate the conformational change of AdT and ensure that the conformation changed to the desired states. As a proof of design, a yeast biosensor that responds to AD was constructed to characterize the availability of non-natural AdT. In addition, the AD-binding pocket of the AdT was changed by site-mutations to obtain higher TF sensitivity. We successfully designed AdT, and used MD simulations to justify the feasibility of design, indicating that we might create non-natural TFs for other molecules. We believe that the design will provide regulatory element for the microbial cell factory, and is applicable to a range of other TFs for proteins that have an LBD and/or receptor.

2. Materials and methods

2.1. Construction of a three-dimensional full-length atom model of AdT

The MD simulation of a non-natural protein started with obtaining a reliable three-dimensional (3D) spatial structure of the protein. Currently, the establishment of 3D protein models is mainly based on template-based modeling. However, the 3D protein model was assembled according to the templates provided by multiple different proteins, which led to low conformational reliability at the interface from different protein templates. In addition, this assembly method isolated the natural interaction force between the templates, resulting in low reliability of the whole 3D model, especially for models with extremely low homology. The iterative threading assembly refinement (I-TASSER) server was used to build the 3D protein model of AdT [28]. A hierarchical approach was applied to I-TASSER for AdT structure prediction and structure-based feature description. It first identified structural templates from the database PDB via the multiple threading method LOMETS, AdT 3D models with full-length atomic was constructed via iterative template-based fragment assembly simulations. Functional perceptiveness of AdT was then obtained by re-threading the 3D models through the protein function database BioLiP. At present, the I-TASSER server is open to the public (https://zhanggroup.org//I-TASSER/).

2.2. Molecular docking

A highly reliable model of the 3D spatial structure was selected as the model for molecular docking of the ligand AD. The LBD in the AdT model was from human progestin receptor; therefore, the binding pocket was in the LBD. In addition, it was reported that Gln 725 of receptor and oxygen at position C3 of progestin are the key sites forming hydrogen bonds [29]. These results were important for judging whether molecular docking was reasonable. The lowest energy state of AD and the AdT model was docked using the software package *AutoDock 4.2.6*. The best docking model of AD and AdT was obtained, and the advantages and disadvantages of the interaction were evaluated in real time based on the principles of geometric topology, chemical environment, and energy complementarity between the two molecules. *PyMOL* was then used to visualize the results of molecular docking.

2.3. MD simulations

MD simulation conforms to the principle of the classical Newtonian mechanical equation of motion. The motion of all molecules or particles of AdT and its complex met the classical Newtonian equation of motion, the hypothesis of statistical mechanics, and the superposition of the interactions between particles in the moving system. In short, the MD simulation of AdT and its complex was the solution of Newton's equation of motion in Cartesian coordinates. Each MD system was calculated

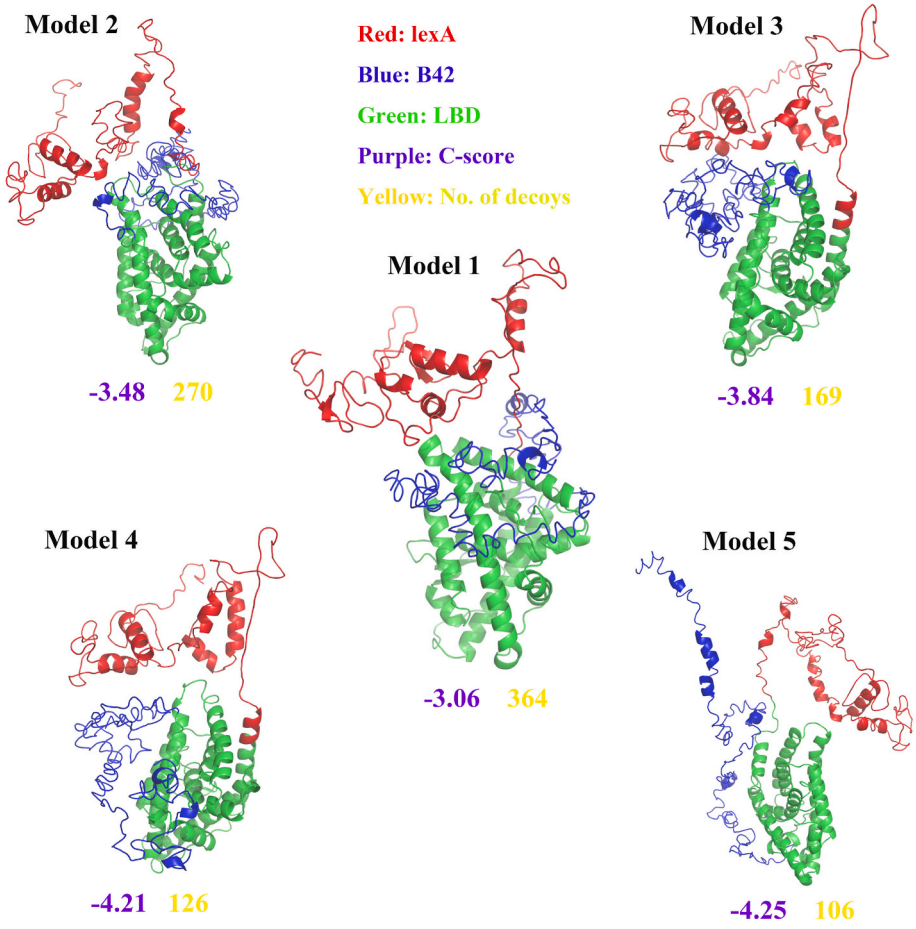

Fig. 2. The establishment of non-natural protein AdT models. Model 1–5 were supplied by I-TASSER with a high confidence, and chosen Model 1 was a more reliable structure based on obtained the high *C*-score and decoys.

through the combined atom force field of *GROMOS96 54A7* of the GMX protocol in the *GROMACS 2019.6* package and accelerated with a graphics card GPU NVIDIA GeForce RTX 4090 [30,31]. First, the charges and coordinates of small molecules were calculated and transformed in Automated Topology Builder (http://atb.uq.edu.au/). The AdT, counterions, AD, and H_2O were added into a dodecahedral solvent box with the simple point-charge water model at physiological salt ion concentrations. Next, the integral of Newtonian motion was iteratively calculated through the algorithmthe leap-frog algorithm. The cutoff radius of van der Waals interactions and short-range electrostatics was re-written to 1.4 nm using the scheme of Verlet-cutoff. The long-range electrostatic forces was computed using a method that is the particle mesh Ewald [32]. Simulations of energy minimization was executed using the algorithm of steepest descent minimization, and both 600-ps NVT (30 °C) and NPT (1 bar) were executed, which these executions were aimed to obtain an initial equilibrium phase. Each MD simulation was calculated for a total of 50 ns, in which the dynamic trajectories of conformational change were written every 10 ps? Each MD simulation was performed at least 3 times.

2.4. Analysis of the MD trajectories

All results for motion trajectories, H-bond properties, and conformation changes were visualized using *VMD*, *LigPlot*$^{+}$, and *Qtgrace* (https://plasma-gate.weizmann.ac.il/Grace/) [33]. Furthermore, the *g_mmpbsa* protocol was utilized to calculate the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) binding free energy [34]. The selection of the center of geometry was based on the Algorithms and Formulas for Modeling of Molecular Systems section (http://www.gromos.net/).

2.5. Characterization of the TF activity of AdT

Hydrocortisone, AD, 17β-estradiol, pregnenolone, androst-1,4-dien-3,17-dione (ADD), dehydroepiandrosterone (DHEA), and testosterone were purchased from aladdin (Shanghai, China). HBC was supplied by Sigma–Aldrich (shanghai, USA), and the MilliQ (MQ)-H_2O was used as pure water. FastDigest restriction enzymes and high-fidelity DNA polymerase were obtained from Fermentas Co., Ltd. (Thermo Fisher, USA). The plasmid extraction kit and a one-step cloning kit were supplied by Solarbio Science & Technology Co., Ltd. (Beijing, China). The gel extraction kit was obtained from Yeasen Biotech Co., Ltd. (Shanghai, China). Other reagents and chemicals with a higher purity available were supplied by multiple companies.

To verify the TF activity of AdT, a *Saccharomyces cerevisiae* whole-cell biosensor was constructed in this study. *Escherichia coli* DH5α and *S. cerevisiae* BY4741 (MATa; *his3Δ1*; *leu2Δ0*; *met15Δ0*; *ura3Δ0*; *gal80Δ0*) were used as hosts for the construction of biosensor to verify AdT activity. The monomer AdT, which was sequentially connected by the amino acid sequences of lexA, $(GGGGS)_2$, LBD (residues 213–469), and B42, was synthesized its gene after codon optimization by Genscript (China). The gene *AdT* was ligated into the pESC-leu^{-} controlled by the ADH1 promoter by digestion with *Bam*HI and *Hin*d III restriction. In addition, *egfp* was ligated into the pESC-his^{-} controlled by a new promoter CYC1s by digestion with *Bam*HI and *Xho*I restriction. A synthetic promoter CYC1s composed of terminator DEG1 [35], 3 lexA sites [36] and promoter CYC1 controls the transcription of *egfp*. The AdT can bind to the DNA sequence of the promoter CYC1s because the domain lexA

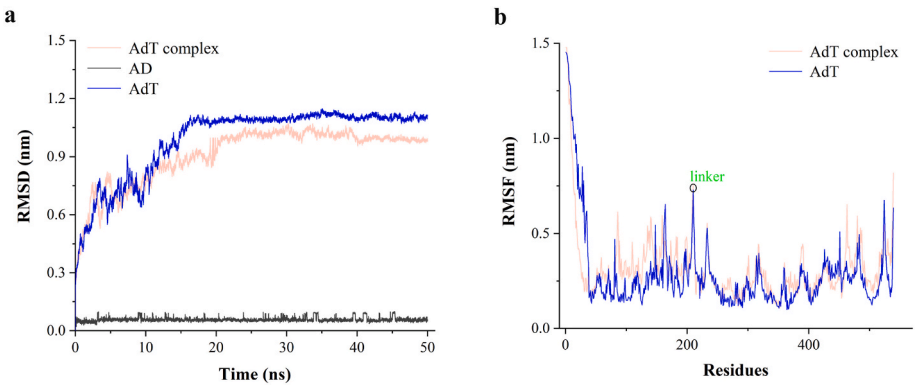

Fig. 3. (A) Evaluation of system balance of AdT complex, AdT, and AD. (b) The stability of amino acid residues. The linker provided greater regional flexibility, which might bring lexA closer to B42 to increase AdT's activity.

contained in AdT specifically recognizes and binds to lexA site. DH5α and *S. cerevisiae* were cultured in medium LB and YPD or nutrient-deficient synthetic medium, respectively. 100 μg/mL ampicillin was added into LB medium to hold plasmid as required for DH5α.

2.6. Improve the activity of AdT by site-mutations

The primers of site-mutations for AdT were designed using software SnapGene for primer design. In addition, the single point mutation of AdT was performed using a site-directed mutagenesis kit, which was supplied by Agilent Technologies Co., Ltd. (Agilent, USA). The gene fragments were connected together by overlapping PCR, which were the AdT mutant genes, and then they were ligated into the plasmid pESC-leu$^-$, respectively. The mutagenesis was sequenced by Sangon Biotech.

3. Results and discussion

To ensure that LBD and linker induced AdT conformational transition from extended conformation to compact conformation upon the binding of AD. Thus, AdT conformational trajectories were simulated via MD to determine whether the binding of AD to its LBD resulted in a shorter distance between the lexA and B42 (Fig. 1).

3.1. Establishment of a 3D model of AdT

A specific AdT was generated by fusing three domains, namely the prokaryotic domain lexA, a linker $(GGGGS)_2$, LBD, and the prokaryotic domain B42, to synthesize a non-natural TF (Fig. 2 model 1). We speculated that linker can regulate lexA's flexibility to control the spatial distance between domains, which may be able to improve AdT's activity. The amino acid sequence of this AdT was submitted to the I-TASSER server, and then 5 AdT models were obtained using I-TASSER based on pairwise structural similarity using the SPICKER program to cluster all decoys (Fig. 2). *C*-score is a confidence score for estimating the quality of predicted AdT models. It is calculated according to the significance of threading template alignments and the convergence parameters of the structure assembly simulations. *C*-score typically ranges −5 to 2, where a higher *C*-score signifies a model with higher confidence [37]. In addition, low-temperature replicas (decoys) generated during the simulation were clustered by SPICKER, and the top five cluster centroids were selected for generating full atomic models. High numbers of decoys suggested that the structure of AdT occurred more often in the simulation trajectory, indicative of a high quality model [37]. Hence, model 1 was selected as the MD simulation conformation because of its high confidence (*C*-score = −3.06) and high number of decoys (364).

3.2. Conformational stability of the AdT complex system

Two AdT structures were prepared, namely a AD-free AdT structure (as control) and a AD-bound AdT complex structure, and two independent MD simulations were performed using the *GROMACS 2019.6* package and the *GROMOS96 54A7* force field for 50 ns each. The conformation states of the AdTs were valuable only after the system in which it was located reached equilibrium. In addition, the dynamic balance of the MD system was generally evaluated by the total energy or root mean square deviation (RMSD). The RMSD of the AdT main chain atoms was calculated and used to evaluate the dynamic equilibrium of the calculation system and judge whether the system converges. As presented in Fig. 3a, the convergence of AdT docked with AD occurred later than that of AdT without AD. This result suggested that the conformation of AdT bound to AD reached dynamic equilibrium in more than 20 ns, and AD affected the convergence of the complex system and made AdT more stable (lower RMSD). Meanwhile, the conformation of the AdT complex was relatively stable for more than 20 ns, indicating that its conformation state had collective value. These results illustrated that 50 ns MD simulation to simulate the dynamic changes of AdT and its complex conformation was feasible and that the results were credible.

3.3. Amino acid residue stability of the AdT conformation systems

To prove that AD triggered lexA and B42 to be close proximity to each other. The proximity between the two domains was bound to cause changes in the conformation of amino acid residues in the lexA and B42. Therefore, the stability of amino acid residues in AdT must be judged from the overall structure, including AdT bound to AD and AdT without AD. Root mean square fluctuation (RMSF) is an analytical modality reflecting the fluctuation of each amino acid residue in a protein in the overall period. Namely, the flexibility or stability of amino acid residues inside the AdTs was evaluated using RMSF in the simulation systems. A large change of the ordinate indicated instability of the amino acid residues or greater flexibility of the regional amino acids. In short, protein conformational changes were mainly determined by amino acid residues with changes in RMSF. Fig. 3b illustrates that the RMSF of the *N*-terminal domain lexA (residues 1–202), linker (residues 203–212) and *C*-terminal domain B42 (residues 470–548) changed significantly regardless of whether AdT was bound by AD. This result suggested that both domains were flexible and variable, but the change in the distance between the lexA and B42 cannot be judged to be caused by binding between AD and AdT. The RMSF of amino acids 250–280 in the conformation of AdT complex was greater than that of the control, indicating that this region was flexible, and AD caused a change in the position of amino acid residues around this region. The MD simulation results revealed that AD induced the instability of amino acid residues in the lexA and B42 after binding to AdT. Compared to the findings for the

Table 1
Properties of H-bonds between AD and donors.

Donor–Acceptor	Occupancy (%)	Distance (Å)	Angle (°)
Q251–AD	13.6	2.9	15.1
T420–AD	6.9	3.3	17.0

control, it was uncertain that the flexibility of the domain was caused by the amino acid residues or secondary structure transformation around the LBD after binding to AD.

3.4. Non-bonding interactions

To further prove that the changes of the lexA and B42 were exactly caused by the binding of AD, non-bonding interactions, the binding free energy of the complex system, and the contribution of amino acid residues to the binding free energy were analyzed. Bound AD structured the core region of AdT, and the function of this region, LBD, is to specifically recognize hormones in a part of the hydrophobic pocket (Fig. 2 model 1) [25]. The specificity is determined by the H-bonds formed between the ketone group of AD and the amino acid residues. In this regard, hydrogen bonds between AD and the AdT complex were monitored over the simulation. The hydrogen bond analysis results are presented in Table 1, the 3-oxygen atoms in the A ring and the 19-oxygen atoms of AD bound to the NH_2 group of Q251 (a key residue) and side chain of T420, respectively, giving rise to H-bonds that together determined the specificity between AD and its LBD. This result was consistent with the reported conclusion [25].

Although binding free energy was calculated using several strategies, MM-PBSA has been widely used because of its accurate calculation results and fast calculation speed. The binding free energy was the sum of the energy generated by six interactions: van del Waals energy between AdT and AD, electrostatic energy between AdT and AD, non-polar solvation of the complex system (SASA energy), polar solvation energy of the complex system, WCA energy, and SAV energy. As presented in Fig. 4a, AD-bound AdT contributed to binding free energy, which further confirmed that AD caused the fluctuation of AdT amino acid residues. Van del Waals energy, SASA energy, and electrostatic energy contributed positively to the binding free energy; however, polar solvation energy had a negative effect on the binding free energy of the complex system. In addition, the van der Waals interaction between AdT and AD was the main contributor to the binding free energy of the complex system. In addition, van der Waals force is closely related to the hydrophobic force of small ligand molecules. This result suggested that the main interaction between AD and amino acids in the complex system might be derived from the hydrophobic force. Moreover, the contribution of amino acid residues to the free energy of the AdT complex indicated that the interactions between amino acid residues within the LBD and AD were major contributors to the free energy of binding, and non-polar amino acid residues, such as Leu, Met, Trp, and Tyr, played a positive role (Fig. 4b–c). Totally, Gln 251 and Thr420 specifically recognized the oxygen atoms of AD through H-bonds, resulting in fluctuation of the amino acid residues around bound-AD. The binding process was mainly caused by the binding free energy generated by the interaction between AD and non-polar amino acid residues of the LBD. These results indicated that AD-bound to AdT was the fundamental cause of the conformational change of the LBD, resulting in changes of the lexA and B42.

3.5. Effect of AD on the secondary structure of AdT

To more intuitively analyze the effect of bound AD on the conformational change of AdT, the secondary structures of proteins were analyzed. These conformational trajectories were analyzed to construct the probability of states for the binding of AD to its LBD, revealing the preferred locations of lexA and B42. The results of the MD simulations indicated two highly different probable conformational states (Fig. 5a and b). After AD entered the pocket and formed hydrogen bonds, the bound AD directly contacted helices, leading to a β-turn change in the *C*-terminal extension (residues 448–459) within the LBD. This result

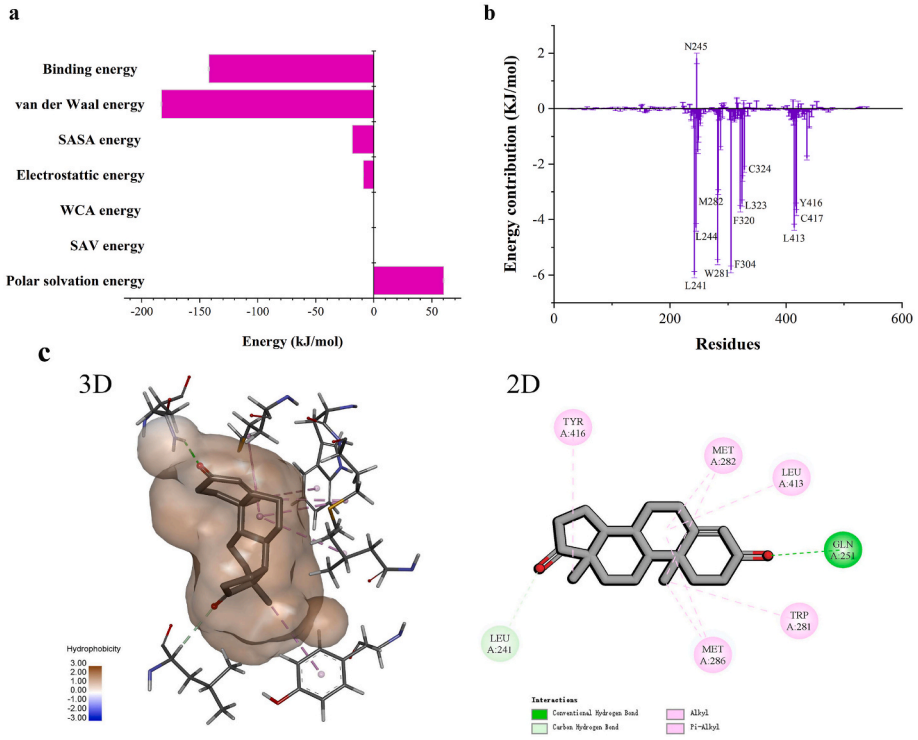

Fig. 4. (A) Computational result of binding free energy of bound-AD system. (b) Contribution of binding free energy of amino acid residues in the bound-AD system. (c) The main residues within the hydrophobic cavity of the pocket. Non-polar amino acids play a positive role in the binding of AD around the pocket.

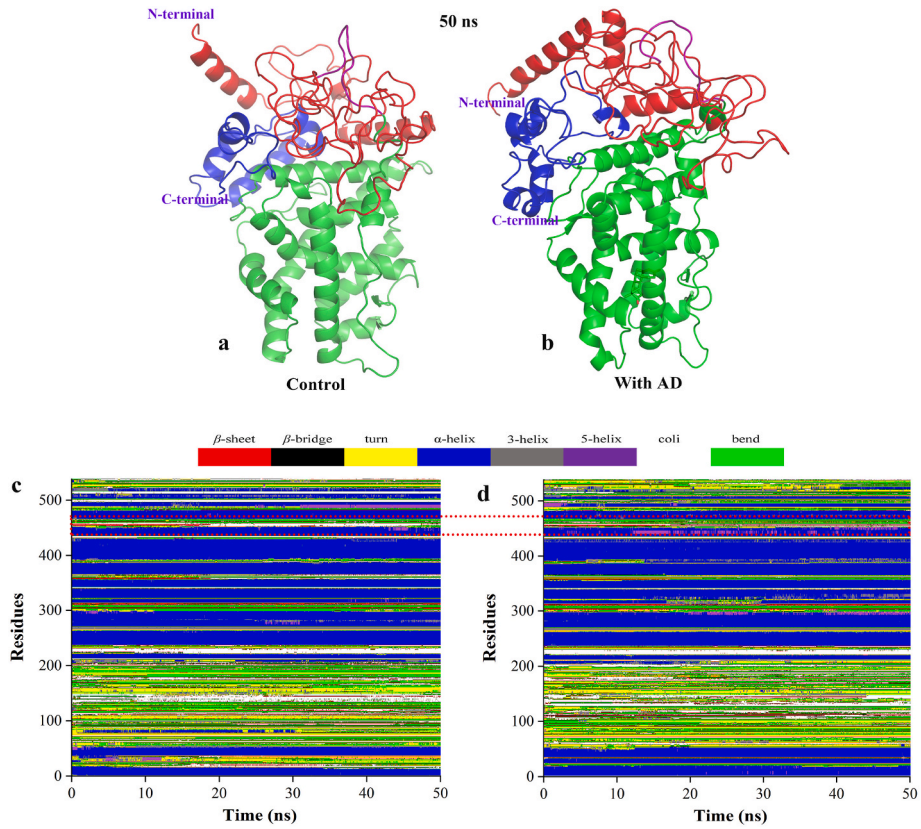

Fig. 5. Analysis of the conformational transition of AdTs by MD simulation. (a) AdT conformation at 50 ns. **(b)** AdT complex conformation at 50 ns. **(*c*–d)** Compared changes of protein secondary structures within AdTs. The red dashed box shows the area with a large change in LBD conformation.

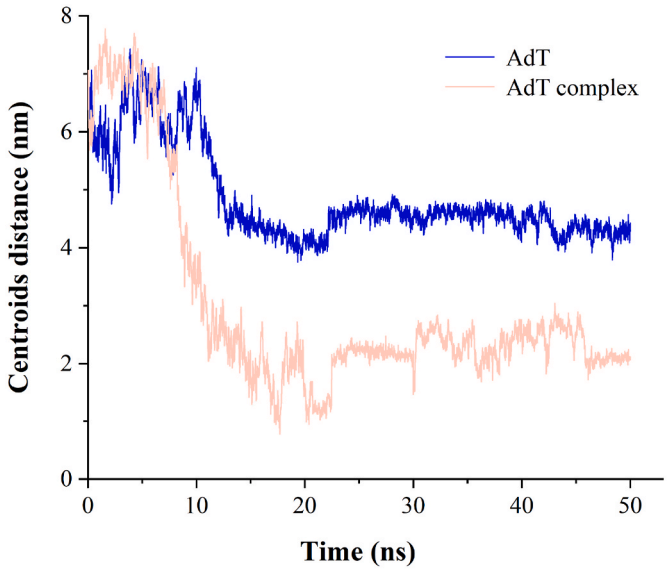

Fig. 6. Evaluation of distance of center of geometry between lexA and B42 in AdT and AdT complex independently.

indicated that the conformational changes in helices caused a change in another regional conformation, namely the 12-residue *C*-terminal extension region. Therefore, we can speculate that the β-turn change of the *C*-terminus occurred through an antiparallel β-sheet interaction between amino acid residues 355–357 and 451–453 (Fig. 5c and d). The perturbations in helixs caused the β-turn angle in the *C*-terminus to change, which drove B42 to tilt to the upper left. This in turn positioned the *N*-terminus closer to B42 (Fig. 5b). Conversely, the *C*-terminal β-turn angle of the AD-free model became stabilized, and it drove B42 toward the left and inside, which positioned the helix of the *N*-terminus above and across the B42 and lexA strands (Fig. 5a). Compared to the results in Fig. 5b, this result demonstrated that the *N*-terminus was not sufficiently close to B42 (Fig. 5a).

To avoid the imprecise conclusion caused by intuition, we selected the centers of geometry of the lexA and B42 as the basis for computation using the "gmx distance" during the whole trajectory to measure the distance of the center of geometry. The computational results suggested that the distance of the center of geometry between the AD-bound and AD-free structures differed by 1.8–2.7 nm after the two AdT systems reached equilibrium at approximately 23 ns, indicating that AD brought the domain lexA closer to B42 (Fig. 6). In fact, the first design of AdT was without linker. MD simulation results showed that there was almost no difference in the centroid distance between the two domains (not shown). MD guided us to insert a linker between the flexible regions of the domain to obtain an AdT, in which lexA and B42 were close enough (Fig. 3b). Therefore, we believe that the prerequisite for positive TF activity is close proximity of lexA and B42. In other words, these results suggested that conformational transitions in the LBD indeed acted as a microswitch to control AdT activity.

3.6. *Availability of AdT as a TF*

We theoretically designed a TF element AdT by MD simulation to confirm that the aforementioned non-natural AdT binds AD to achieve high-sensitivity TF function by changing its own conformation to control the transcription of the downstream reporter gene enhanced green fluorescent protein (e*gfp*, Fig. 7a). AdT as a non-natural AD-responsive TF was applied to construct *S. cerevisiae* biosensor. A strong promoter ADH1 was used in an AdT gene expression cassette. In addition, eGFP

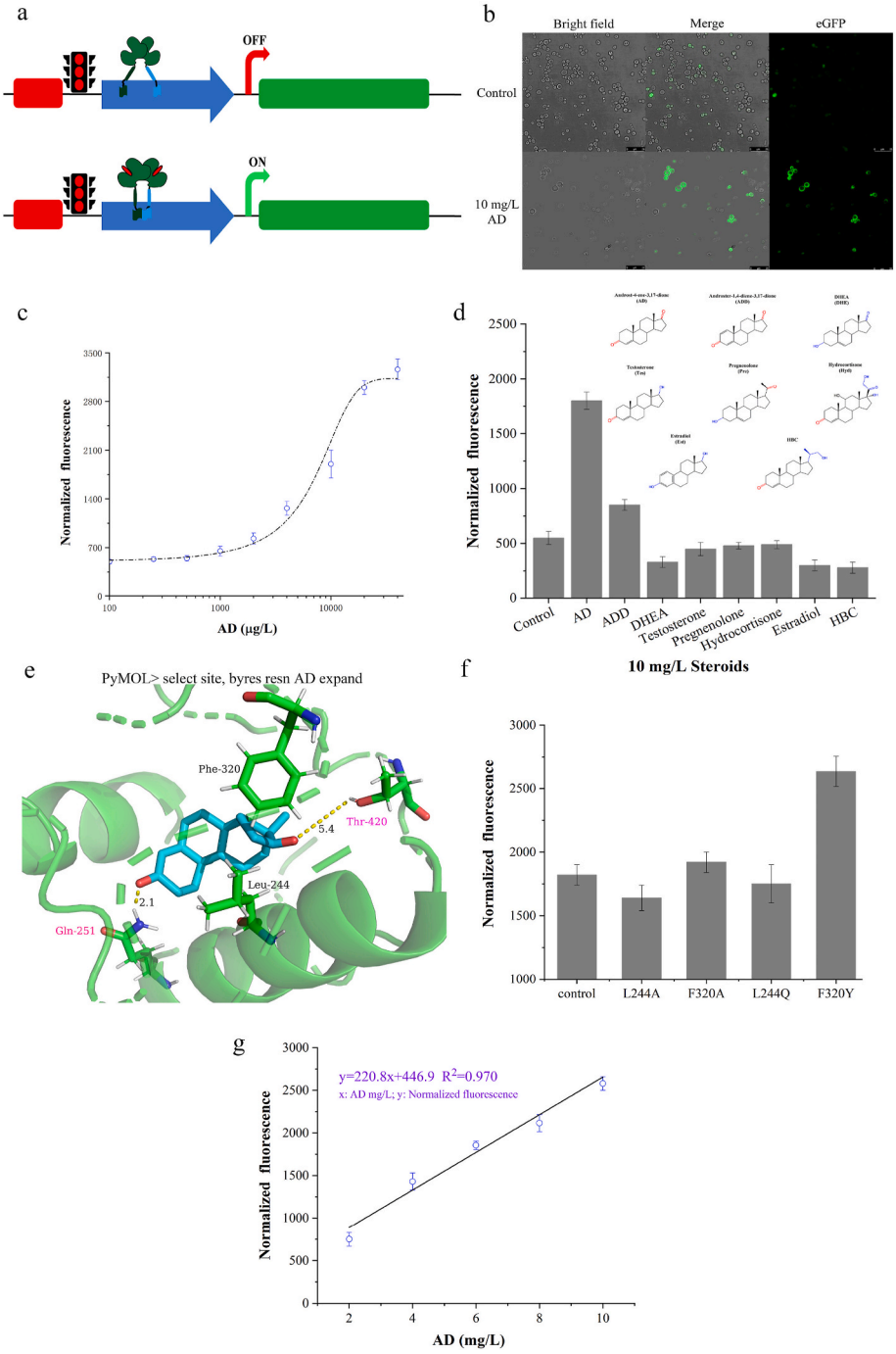

Fig. 7. Validation and optimization of transcription factor AdT in yeast. (a) Schematic of gene switch for strict control of *egfp* transcription by AdT. **(b)** Analysis the eGFP fluorescence in yeast biosensor. **(c)** Testing of sensitivity. **(d)** Testing of specificity of the AD biosensor. **(e)** Analysis of steric resistance of AD based on AdT docking with AD. **(f)** Enhancing AdT activity through site-mutations. **(g)** Testing of linearity.

controlled by the promoter CYC1s on a plasmid pESC-his$^-$ was used as a visible signal. The promoter CYC1s was derived from CYC1, which contained 3 prokaryotic DNA-binding sites recognized by the AdT to avoid non-specific transcription in the eukaryotic system. The terminator DEG1 like traffic light prevented undesired transcriptional events [35]. Hence, the transcription events of *egfp* were strictly controlled by AD (Fig. 7b).

The sensitivity and specificity of AD biosensor was evaluated. The response time of 6 h at 30 °C was chosen for subsequent studies (OD_{600} ≈ 5). The yeast sensor was able to respond AD with high sensitivity, with a half-maximum effective concentration (EC_{50}) of 10 mg/L (Fig. 7c). To further test the specificity of biosensor, we chose 8 steroids with similar structures and found that the biosensor was highly specific for AD (Fig. 7d). However, the fluorescence intensity only increased by 3.3 times compared to the control, which suggested that the activity of TF was not as expected. In order to increase the sensitivity of AdT, the AD pocket was changed via site-mutations. We speculated that the reduction of AD steric resistance can improve the stability of the H-bond between AD and Q251 and T420 (Fig. 7e). Hence, L244 and F320 that hinder AD from approaching T420 were mutated to alanine. But, the fluorescence intensity has not increased significantly. Then, the amino acid residues at these two sites were saturated with mutations,

respectively. Importantly, variant F320Y showed a sharp improvement in fluorescence intensity, which suggested AdT activity was improved 1.44-fold (Fig. 7f). We found that the electron cloud density maps of phenylalanine and tyrosine are very similar, which affirmed the "QTY" code [38]. The repulsion between the hydrophilic tyrosine and the hydrophobic AD might change the AD position in the pocket, which contributed to the formation of H-bonds. Lastly, the solutions with a known concentration of AD were prepared, and the results revealed a good linear relationship between fluorescence intensity and AD content (Fig. 7g). We successfully used modular assembly of protein functional domains based on structure-guided MD to obtain a non-natural element TF. In addition, we expect that the design TF strategy used in this work will be applied to running in parallel to the signaling machinery of the host cell, such as synthetic biology and high-throughput drug screening.

4. Conclusions

A non-natural protein element AdT was designed, and binding of AD to AdT caused a conformational transformation of the lexA and B42 domains by MD simulation, resulting in positive TF activity for AdT. Three models, namely the ligand AD alone, AdT, and AdT docked with AD, were independently performed using *GROMACS* software package for 50 ns, respectively. Then, these conformational trajectories were analyzed, and AD was revealed to cause the distinct conformational states of AdTs to be significantly different. In addition, the distance of the center of geometry was calculated, and the value suggested that AD caused the lexA closer to the B42. A yeast sensor was successfully constructed to prove AdT with TF activity, and the mutant F320Y plays a key role in improving 1.44-fold for AdT activity. We used MD strategies to design element TF, minimize the requirements for complex engineering, and highlight the power of designing *de novo* elements based on structure-guided MD.

CRediT authorship contribution statement

Ming Zhao: carried out the experiments. **Mengkai Hu:** carried out the experiments. **Rumeng Han:** carried out the experiments. **Chao Ye:** analyzed the data. **Xiangfei Li:** gudied the experiments. **Tianwen Wang:** gudied the experiments. **Yan Liu:** offered useful suggestions. **Zhenglian Xue:** offered useful suggestions. **Kun Liu:** supervised the study and wrote the manuscript with feedback from all authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was financially supported by the Scientific Research Foundation of the Higher Education Institutions of Anhui Province, China (Grant No. 2022AH050971), Key R&D and Achievement Transformation Projects (2023yf096), the Scientific Research Start-up Fund for Talents of Anhui Polytechnic University (2022YQQ064).

References

[1] Yu WW, Xu XH, Jin K, Liu YF, Li JH, Du GH, Lv XQ, Liu L. Genetically encoded biosensors for microbial synthetic biology: from conceptual frameworks to practical applications. Biotechnol Adv 2023;62:108077.
[2] He HH, Yang MF, Li S, Zhang GY, Ding ZY, Zhang L, Shi GY, Li YR. Mechanisms and biotechnological applications of transcription factors. Synth Syst Biotechnol 2023;8:565–77.
[3] Zhang HP, Zhang L, Xu YP, Chen SY, Ma ZY, Yao MD, Li FY, Li B, Yuan YJ. Simulating androgen receptor selection in designer yeast. Synth Syst Biotechnol 2022;7:1108–16.
[4] Ishihama A, Shimada T, Yamazaki Y. Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. Nucleic Acids Res 2016;44:2058–74.
[5] Grazon C, Baer RC, Kuzmanović U, Nguyen T, Chen M, Zamani M, Chern M, Aquino P, Zhang X, Lecommandoux S, Fan A, Cabodi M, Klapperich C, Grinstaff MW, Dennis AM, Galagan JE. A progesterone biosensor derived from microbial screening. Nat Commun 2020;11:1276.
[6] Collins CH, Arnold FH, Leadbetter JR. Directed evolution of *Vibrio fischeri* LuxR for increased sensitivity to a broad spectrum of acyl-homoserine lactones. Mol Microbiol 2005;55:712–23.
[7] Tang S, Fazelinia H, Cirino PC. AraC regulatory protein mutants with altered effector specificity. J Am Chem Soc 2008;130:5267–71.
[8] Taylor ND, Garruss AS, Moretti R, Chan S, Arbing MA, Cascio D, Rogers JK, Isaacs FJ, Kosuri S, Baker D, Fields S, Church GM, Raman S. Engineering an allosteric transcription factor to respond to new ligands. Nat Methods 2016;13:177–83.
[9] Della CD, van Beek HL, Syberg F, Schallmey M, Tobola F, Cormann KU, Schlicker C, Baumann PT, Krumbach K, Sokolowsky S, Morris CJ, Grünberger A, Hofmann E, Schröder GF, Marienhagen J. Engineering and application of a biosensor with focused ligand specificity. Nat Commun 2020;11:4851.
[10] Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL, Baker D. Computational design of ligand-binding proteins with high affinity and selectivity. Nature 2013;501:212–6.
[11] Quijano-Rubio A, Yeh H, Park J, Lee H, Langan RA, Boyken SE, Lajoie MJ, Cao L, Chow CM, Miranda MC, Wi J, Hong HJ, Stewart L, Oh B, Baker D. *De novo* design of modular and tunable protein biosensors. Nature 2021;591:482–7.
[12] Cao LX, Coventry B, Goreshnik I, Huang B, Sheffler W, Park JS, Jude KM, Marković I, Kadam RU, Verschueren KHG, Verstraete K, Walsh STR, Bennett N, Phal A, Yang A, Kozodoy L, DeWitt M, Picton L, Miller L, Strauch E, DeBouver ND, Pires A, Bera AK, Halabiya S, Hammerson B, Yang W, Bernard S, Stewart L, Wilson IA, Ruohola-Baker H, Schlessinger J, Lee S, Savvides SN, Garcia KC, Baker D. Design of protein-binding proteins from the target structure alone. Nature 2022;605:551–60.
[13] Cui Y, Chen Y, Liu X, Dong S, Tian YE, Qiao Y, Mitra R, Han J, Li C, Han X, Liu W, Chen Q, Wei W, Wang X, Du W, Tang S, Xiang H, Liu H, Liang Y, Houk KN, Wu B. Computational redesign of a PETase for plastic biodegradation under ambient condition by the GRAPE strategy. ACS Catal 2021;11:1340–50.
[14] Liu K, Zhang YS, Liu K, Zhao YQ, Gao B, Tao XY, Zhao M, Wang FQ, Wei DZ. *De novo* design of a transcription factor for a progesterone biosensor. Biosens Bioelectron 2022;203:113897.
[15] Li FW, Ma L, Zhang XW, Chen JF, Qi FF, Huang YY, Qu ZP, Yao LS, Zhang W, Kim E, Li SY. Structure-guided manipulation of the regioselectivity of the cyclosporine a hydroxylase CYP-sb21 from *Sebekia benihana*. Synth Syst Biotechnol 2020;5:236–43.
[16] Fiorentino F, Sauer JB, Qiu X, Corey RA, Cassidy CK, Mynors-Wallis B, Mehmood S, Bolla JR, Stansfeld PJ, Robinson CV. Dynamics of an LPS translocon induced by substrate and an antimicrobial peptide. Nat Chem Biol 2021;17:187–95.
[17] Liu S, Cao XL, Liu GQ, Zhou T, Yang XL, Ma BX. The in silico and in vivo evaluation of puerarin against Alzheimer's disease. Food Funct 2019;10:799–813.
[18] Etheve L, Martin J, Lavery R. Dynamics and recognition within a protein-DNA complex: a molecular dynamics study of the SKN-1/DNA interaction. Nucleic Acids Res 2016;44:1440–8.
[19] Xu LQ, Liu Y, Yao K, Liu HH, Tao XY, Wang FQ, Wei DZ. Unraveling and engineering the production of 23,24-bisnorcholenic steroids in sterol metabolism. Sci Rep 2016;6:21928.
[20] Zhou XL, Zhang Y, Shen YB, Zhang X, Zan ZH, Xia ML, Luo JM, Wang M. Efficient repeated batch production of androstenedione using untreated cane molasses by Mycobacterium neoaurum driven by ATP futile cycle. Bioresour Technol 2020;309:123307.
[21] Yao K, Wang FQ, Zhang HC, Wei DZ. Identification and engineering of cholesterol oxidases involved in the initial step of sterols catabolism in *Mycobacterium neoaurum*. Metab Eng 2013;15:75–87.
[22] Yao K, Xu LQ, Wang FQ, Wei DZ. Characterization and engineering of 3-ketosteroid-△1-dehydrogenase and 3-ketosteroid-9α-hydroxylase in *Mycobacterium neoaurum* ATCC 25795 to produce 9α-hydroxy-4-androstene-3,17-dione through the catabolism of sterols. Metab Eng 2014;24:181–91.
[23] Liu K, Wang FQ, Liu K, Zhao YQ, Gao B, Tao XY, Wei DZ. Light-driven progesterone production by InP–(*M. neoaurum*) biohybrid system. Bioresour Bioprocess 2022;9:93.
[24] Li Y, Wood K, Ptashne M, Ruden DM, Ma J. Generating yeast transcriptional activators containing no yeast protein sequences. Nature 1991;350:250–2.
[25] Tanenbaum DM, Wang Y, Williams SP, Sigler PB. Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. Proc Natl Acad Sci USA 1998;95:5998–6003.
[26] Ma J, Ptashne M. A new class of yeast transcriptional activators. Cell 1987;51:113–9.
[27] Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U. Yeast two-hybrid, a powerful tool for systems biology. Int J Mol Sci 2009;10:2763–88.
[28] Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods 2015;12:7–8.
[29] Williams SP, Sigler PB. Atomic structure of progesterone complexed with its receptor. Nature 1998;393:392–6.
[30] Fuller PJ, Yao Y, Jin R, He S, Martín-Fernández B, Young MJ, Smith BJ. Molecular evolution of the switch for progesterone and spironolactone from mineralocorticoid receptor agonist to antagonist. Proc Natl Acad Sci USA 2019;116:18578–83.

[31] Hua T, Li XT, Wu LJ, Iliopoulos-Tsoutsouvas C, Wang YX, Wu M, Shen L, Brust CA, Nikas SP, Song F, Song XY, Yuan SG, Sun QQ, Wu YR, Jiang S, Grim TW, Benchama O, Stahl EL, Zvonok N, Zhao SW, Bohn LM, Makriyannis A, Liu ZJ. Activation and signaling mechanism revealed by cannabinoid receptor-gi complex structures. Cell 2020;180:655–65.

[32] Wan JY, Xie J, Kong X, Liu Z, Liu K, Shi FF, Pei A, Chen H, Chen W, Chen J, Zhang XK, Zong LQ, Wang JY, Chen LQ, Qin J, Cui Y. Ultrathin, flexible, solid polymer composite electrolyte enabled with aligned nanoporous host for lithium batteries. Nat Nanotechnol 2019;14:705–11.

[33] Alam A, Kowal J, Broude E, Roninson I, Locher KP. Structural insight into substrate and inhibitor discrimination by human P-glycoprotein. Science 2019;363:753–6.

[34] Wilmes S, Hafer M, Vuorio J, Tucker JA, Winkelmann H, Lochte S, Stanly TA, Pulgar PK, Poojari C, Sharma V, Richter CP, Kurre R, Hubbard SR, Garcia KC, Moraga I, Vattulainen I, Hitchcock IS, Piehler J. Mechanism of homodimeric cytokine receptor activation and dysregulation by oncogenic mutations. Science 2020;367:643–52.

[35] Brambilla A, Mainieri D, Carbone MLA. A simple signal element mediates transcription termination and mRNA 3'end formation in the DEG1 gene of *Saccharomyces cerevisiae*. Mol Gen Genet 1997;254:681–8.

[36] Brent R, Ptashne M. Mechanism of action of the lexA gene product. Proc Natl Acad Sci USA 1981;78:4204–8.

[37] Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 2010;5:725–38.

[38] Zhang SG, Tao F, Qing R, Tang HZ, Skuhersky M, Corin K, Tegler L, Wassie A, Wassie B, Kwon Y, Suter B, Entzian C, Schubert T, Yang G, Labahn J, Kubicek J, Maertens B. QTY code enables design of detergent-free chemokine receptors that retain ligand-binding activities. Proc Natl Acad Sci USA 2018;115:8652–9.