FEBS Letters

FEBS PRESS science publishing by scientists

Why the c-Fos/c-Jun complex is extremely conserved: An *in vitro* evolution exploration by combining cDNA display and proximity ligation

Shou-Kai Kang, Xin-Yi Chu, Tian Tian, Peng-Fei Dong, Bai-Xue Chen and Hong-Yu Zhang

Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, China

Correspondence
H.-Y. Zhang, Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
Tel: +86-27-87285085
E-mail: zhy630@mail.hzau.edu.cn

(Received 8 February 2019, revised 31 March 2019, accepted 10 April 2019, available online 6 May 2019)

doi:10.1002/1873-3468.13388

Edited by Judith Zaugg

Transcriptional regulation involves a series of sophisticated protein–protein and protein–DNA interactions (PPI and PDI). Some transcriptional complexes, such as c-Fos/c-Jun and their binding DNA fragments, have been conserved over the past one billion years. Considering the thermodynamic principle for transcriptional complex formation, we hypothesized that the c-Fos/c-Jun complex may represent a thermodynamic summit in the evolutionary space. To test this, we invented a new method, termed One-Pot-seq, which combines cDNA display and proximity ligation to analyse PPI/PDI complexes simultaneously. We found that the wild-type c-Fos/c-Jun complex is indeed the most thermodynamically stable relative to various mutants of c-Fos/c-Jun and binding DNA fragments. Our method also provides a universal approach to detect transcriptional complexes and explore transcriptional regulation mechanisms.

Keywords: cDNA display; c-Fos/c-Jun; evolution; proximity ligation; transcription factor

All biological functions are dependent on transcriptional regulation, which establishes complex relationships between genotype and phenotype. Species produce different phenotypes to adapt to the environments during evolution. In this process, the transcriptional regulation machines are usually updated. However, some transcriptional regulation systems are extremely conserved. For instance basic leucine zipper (bZIP) proteins play several critical roles in the biological processes of all eukaryotes. Some transcriptional complexes composed of bZIP dimers, such as c-Fos/c-Jun, and their binding DNA fragments, are conserved in metazoan species that diverged approximately one billion years ago [1–3]. This evolutionary conservation was considered to be associated with the affinity between the bio-macromolecules that form the transcriptional complex. Aaron *et al.* investigated the dimerization among bZIP proteins of five metazoan species that represent different stages of metazoan evolution. Their study found that compared with bZIP dimers which only appear in parts of the five species, the dimers which are conserved in all species have stronger protein–protein interaction (PPI) strength [3]. In addition, an extensive investigation of transcription factors (TFs) and their binding DNA sequences in different species showed that a stronger protein–DNA interaction (PDI) strength leads to higher specificity and more conserved TF binding [4]. Together, a thermodynamic principle for transcriptional complex formation can be proposed: the more thermodynamically stable the complex, the more conserved it is during evolution. Thus, an interesting question arises: whether the c-Fos/c-Jun system has reached a thermodynamic summit in the evolutionary space?

Abbreviations
bZIP, basic leucine zipper; EMSA, electrophoretic mobility shift assay; PDI, protein–DNA interaction; PoC, proof-of-concept; PPI, protein–protein interaction; TFs, transcription factors.

To address this question, we need to perform an *in vitro* evolution in a large sequence space, during which the mutants of c-Fos, c-Jun and DNA fragments will be selected for forming stable interaction complexes. If the conventional c-Fos/c-Jun system reached the thermodynamic summit in the evolutionary space, it will have the highest frequency in the selected interaction complex population. Nevertheless, it remains a great challenge for determining PPI and PDI simultaneously in the population. Various methods, including high-throughput-sequencing-based techniques, have been developed in the past decades to determine PPI or PDI [5–8]. However, most of these methods measure the congeneric physical interactions between interaction pairs, which cannot be used to determine PPI and PDI synchronously. Although some mass-spectrometry-based methods have been invented to identify proteins binding to specific genomic loci [9–11], they are difficult to be performed in a high-throughput manner.

To synchronously obtain the PPI and PDI information in the interaction complex population, we developed a new method. It is well-known that DNA sequences can be determined more conveniently than protein sequences. Several methods, such as SNAP display [12], phage display [13], ribosome display [14], mRNA display [15,16] and cDNA display [17,18], have been developed to form a barcoded protein by linking a DNA fragment to a protein. Of these methods, cDNA display is of great interest, because it could produce high-content ($> 10^{12}$) stable cDNA–protein fusions, providing a robust display technology for detecting molecular interactions [17,18]. To read out the PPI/PDI information, we resorted to the proximity ligation technique, which has been successfully used to determine PPI or PDI [19,20]. This technique could ligate a cDNA barcode to another intracomplex protein's barcode, or to the intracomplex binding DNA fragment. In this way, the information of PPI and PDI contained in the interaction complex population could be decoded by DNA sequencing. Therefore, this method could determine PPI/PDI of transcriptional complexes in one pot, which is called the One-Pot-seq method.

In this study, firstly, the feasibility of the One-Pot-seq method was examined by a proof-of-concept (PoC) experiment. Then, based on this method, the *in vitro* evolution of c-Fos/c-Jun system was performed. Our work not only takes a brief glance into the evolution of the bZIP transcriptional system, but also provides a new method for determining internal interactions of DNA–protein complexes, which may help to elucidate the mechanisms of transcriptional regulation.

Materials and methods

Preparation of templates for cDNA display

Construction of the DNA templates with cDNA display formation for *c-Fos*/*c-Jun* and glutathione S-transferase (*gst*) were prepared by PCR using the primers listed in Table S1. The DNA sequence used in this experiment was identical as previously used [21]. The mouse brain cDNA library (CWBio, Beijing, China) and pGEX-4T-1 plasmid served as templates for the first round PCR. MFFos/SRFos, MFJun/SRJun and MFGST/SRGST were used as primers. The PCR products were purified with an EZNA Cycle Pure Kit (OMEGA, Norcross, GA, USA). Using the purified products as a template, the second round PCR was performed with MT7 and SR, and the third round PCR was performed with MT7 and LHR-59. After the third PCR, the full-length DNA sequences were obtained, containing a T7 RNA polymerase promoter, an ε-enhancer, a Shine-Dalgarno sequence for ribosome binding at the 5′ end, a His-tag sequence, a spacer sequence and the complementary sequence of the puromycin linker DNA (LHR) at the 3′ end. The purified PCR products were then ligated with the pMD-18T vector, transformed into competent *E. coli* DH5α cells and sequenced by Wuhan TsingKe Biological Technology Co., Ltd. (Wuhan, China).

Preparation of barcoded proteins by cDNA display

Barcoded proteins were prepared according to the procedures of previous studies [18,21,22], with some modifications. The templates were *in vitro* transcribed with the T7 Ribo-MAX Large Scale RNA Production System (Promega, Madison, WI, USA) and purified using a MicroElute RNA Clean-up kit (OMEGA), detected by 1.5% agarose gel electrophoresis and quantified using EON (Biotek, Winooski, VT, USA). The purified mRNA and a puromycin linker [18] (at a proportion of 1 : 1.5) were ligated using T4 RNA ligase (Takara, Dalian, Liaoning, China). The mRNA-puromycin conjugate solution was then concentrated to 20 μL using a 30 kDa MWCO Amicon Ultra-0.5 mL centrifugal filter unit (Millipore, Burlington, MA, USA) at 8000 ***g*** for 20 min. An *E. coli* S30 Extract System for Linear Templates (Promega) was used for *in vitro* translation.

The cDNA-displayed proteins were reverse-transcribed with SuperScriptIII reverse transcriptase (Invitrogen, Carlsbad, CA, USA) at 42 °C for 30 min. In order to perform double-stranded blunt end ligation, we synthesized the second strand of cDNA using the Klenow fragment (Takara) with the primer T7P, which was phosphorylated at the 5′ end and biotin-modified at the fourth base. 400 μL washed Streptavidin C1 beads (ThermoFisher, Waltham, MA, USA) were added to the mixture containing mRNA–protein fusions and incubated for 30 min at room temperature,

and then washed three times with 1 × NEB buffer 4. Then the barcoded proteins were released with Endo V (NEB). The fusions were prepurified using 200 μL of Metal Affinity Resin (TALON, Dalian, Liaoning, China) to remove the frameshift mutation sequence.

Preparation of the bait DNA pool

In the proof-of-concept experiment, the bait and mutant DNA sequences were the same as previously used [21]. The bait DNA sequence was composed of three cognate c-Fos/c-Jun-binding DNA fragments-TPA responsive element (TRE), and the mutant TRE DNA fragment was a combination of six tandem mutant motifs without c-Fos/c-Jun-binding ability (Table S1). The bait and mutant DNA fragments were prepared by two-step PCR. The variable regions of the two DNA sequences were generated by the first round overlap extension PCR using MDF/MDR and CDF/CDR for cognate and mutant fragments respectively. The different 5′ and 3′ constant sequences were joined via a second PCR with the primer OPDF/KMR and COPD/KMR. The PCR products were analysed by 2% agarose gel electrophoresis and purified using a High Pure PCR Product Purification kit (Roche, Basel, Switzerland). KMF-4/KMR was used to amplify the DNA fragments in the subsequent solution-phase selection. Desthiobiotin-modified primers DIB and KMR were used to amplify the products during the solid-phase selection and *in situ* proximity ligation. For these steps, Pfu Taq DNA polymerase (Thermo) was used to reduce the errors introduced by the PCR process.

For the *in vitro* evolution platform, which was used to evaluate the feasibility of One-Pot-seq in a complex system, we redesigned the bait DNA pool. The new DNA pool consisted of three components: the bait DNA (130 bp) and two mutant pools of the bait DNA. The full length of the mutants' DNA was 134 bp, which contained the ggat barcode. The mutant I comprised six tandem nucleotide fragments (TGANTCA), which included a modification of the fourth base of the TRE motif to N (N = A, G, C or T), whereas the mutant II was composed of six NNNNNNN fragments. The samples were prepared with KMF/KMR as described above. The final DNA pool was mixed at a ratio of 1: 1: 10, and the diversity of 3 μg of this DNA pool was around 2.0×10^{13}.

Preparation of c-Fos/c-Jun random *trans* mutant library

The *trans* mutant library was prepared using transfer mutagenic PCR according to a previously reported method [23]. The mixture was dispensed into 16 PCR tubes, with 96.8 and 96.5 μL in the *c-fos* (F1) and *c-jun* (J1) tubes, respectively, 88 μL in the F2-F8/J2-J8 tubes. 1.2 μL of *c-fos* (80.5 ng·μL^{-1}) and 1.5 μL of *c-jun* (59 ng·μL^{-1}) were added into tubes F1 and J1 respectively. Following a denaturation step (94 °C, 1 min), the reaction temperature reached 62 °C, 1 μL of 50 mM $MnCl_2$ (Sigma, St. Louis, MO, USA) and 1 μL of rTaq DNA polymerase (Takara) were added into tubes J1/F1. After four cycles of PCR amplification (94 °C, 1 min; 62 °C, 1 min; 72 °C, 3 min), 10 μL of PCR products in tubes J1/F1 were transferred to tubes J2/F2 as the DNA template. The PCR program was started. Once the PCR program reached the annealing step, the PCR program was paused. 1 μL of 50 mM $MnCl_2$ (Sigma) and 1 μL of rTaq DNA polymerase (Takara) were added into tubes J2/F2. The second-step PCR was performed using the same program as the first-step PCR. The same procedure used for tubes J2/F2 was repeated for tubes J3-J8/F3-F8. The results of the error-prone PCR were examined by agarose gel electrophoresis. PCR products were purified using a Cycle Pure kit (OMEGA). The library was then sequenced to evaluate the mutation efficiency and calculate the diversity of the random library.

Selection of the interaction complexes in solid phase

In this process, we performed two-step solid-phase purification to select the interaction complexes. First, we used the desthiobiotin tag in the bait DNA to eliminate nonspecific binding to DNA. 3 μg of bait DNA pool was immobilized using 200 μL Dynabeads MyOne C1 (Invitrogen). Then the beads were blocked with free biotin in the binding buffer [5 mM Tris-HCl (pH 7.5), 1 M NaCl, 0.5 mM EDTA] and equilibrated with selection buffer. The 100 μL mixture contained barcoded proteins were incubated with the bait DNA on the beads at 4 °C with rotation for 2 h. The beads were washed with selection buffer twice to eliminate nonspecific binding protein. Then the interaction complexes were eluted with selection buffer containing 5 mM biotin. The supernatant was dialysed with a Thermodialysis tube (Thermo) to remove the excess biotin. Next, we performed the second-step purification with the His-tag in barcoded proteins to remove the nonspecific binding DNA. The supernatant was incubated with 200 μL Dynabeads His-Tag beads (Thermo) at room temperature for 30 min. The beads were washed with 400 μL of Wash Buffer to remove the unbound DNA. Bound barcoded proteins and DNA were then amplified by PCR and used as templates in the next round of selection.

Selection of the interaction complexes in solution phase

Based on the different electrophoretic mobility shift, the interaction complexes can be isolated. The 100 μL barcoded protein mixture and 3 μg of bait DNA pool were incubated in selection buffer [50 mM Tris-HCl (pH 8.0), 40 mM NaCl, 5 mM $MgCl_2$, 3 mM KCl and 2% glycerol (vol/vol)], at 4 °C with rotation for 2 h. The mixtures were then run on a 6% native polyacrylamide gel at 20 mA. The

positions of the barcoded proteins were defined by FITC label in the puromycin linker, and the barcoded protein–DNA complexes showed the smallest shift in the gel. The band larger than the barcoded proteins was cut and collected. The gel containing the interaction complexes was purified using a QIAquick Gel kit (QIAGEN, Hilden, Germany). The eluted fractions were amplified by PCR and used as templates in the next round of selection.

Capture of interaction complexes by proximity ligation

During this process, we also conducted two-step consecutive affinity-purification to increase the selection pressure and decrease the number of false positives. The 200 μL Dynabeads MyOne C1 beads, coated with desthiobiotin tag bait DNA was used to eliminate the nonspecific barcoded proteins, the same with the first step of solid-phase selection. After incubation, the beads were washed twice with selection buffer and then equilibrated with T4 DNA polymerase buffer. The end of the fragment was filled by adding T4 DNA polymerase (Takara) and dNTP; the reaction was held at 11 °C for 20 min and then terminated by the addition of 5 μL of 0.5 M EDTA. The supernatant was discarded, and the beads were washed three times with T4 DNA ligase buffer. The 4 μL T4 DNA ligase was then added to perform *in situ* proximity-based ligation at 16 °C for 6 h with rotation in 200 μL ligation buffer.

After the proximity ligation, 5 μL of protease K (Thermo) was added, and the mixture was incubated at 65 °C for 12 h. Subsequently, an additional 2 μL of protease K was added. After protein degradation, the interaction complexes were eluted with selection buffer containing 5 mM biotin. The supernatant was dialysed with a Thermodialysis tube (Thermo) to remove the excess biotin. Next, the second-step purification was also performed with the desthiobiotin modification in the second strand cDNA of the barcoded proteins to remove the unligated DNA fragments. T4 DNA polymerase (Takara) was added to remove the desthiobiotin modification in the unligated DNA fragments, and then Dynabeads MyOne C1 beads were used once again to pull down the ligated DNA fragments.

Interaction complex amplification by *in situ* nested PCR

The beads mentioned in the above step were resuspended with 100 μL 1 × EB buffer [10 mM Tris-Cl (pH 8.5)] to be used as the template for nested PCR. Different primers were used to detect the PDI or PPI. The primers were listed in Table S1. The PDI ligation products were amplified using KMF-4/SR for the first round of PCR and NMD/NSR for the second round of PCR. In the case of PPI detection, SR was used in the first round and NSR was used in second round of PCR. The final PCR product was purified using an EZNA Cycle Pure Kit (OMEGA). The purified PCR product was ligated to the vector, transformed into competent cells and sequenced by Wuhan TsingKe Biological Technology Co., Ltd. (Wuhan, China).

Analysis of high-throughput sequencing data

Consensus contigs of pairwise reads were generated using FLASH [24]. To eliminate aberrant sequences arising from mutations during PCR and sequencing, only contigs with 130 or 134 bp length were regarded as reliable contigs, then the primer sequences in contigs were removed. Finally, the filtered contigs were analysed with BLAST program [25] and the frequencies and rates were calculated with R package.

Results and Discussion

Pipeline for One-Pot-seq

The pipeline for One-Pot-seq method is as follows. First, the target proteins are covalently linked with their corresponding double-stranded DNA fragments using cDNA display, forming barcoded proteins. Barcoded proteins are then incubated with the bait DNA and enriched using consecutive affinity-purification (solid-phase selection) or iterative subtractive differential electrophoretic mobility shift assay (EMSA) (solution-phase selection) [26]. The solid-phase selection is appropriate for screening the complexes with different masses. Whereas, the solution-phase selection is suitable to screen the interaction complexes with similar molecular masses, and is helpful for maintaining the conformation of the complexes. The screened protein–protein and protein–DNA complexes are joined via *in situ* blunt end proximity ligation. The captured ligation products are amplified by *in situ* nested PCR and analysed by DNA sequencing to decode the interacting PPI/PDI complexes (Fig. 1).

Proof-of-concept experiment

To examine the feasibility of the One-Pot-seq method, a proof-of-concept experiment was performed to determine the cognate binding DNA fragments of the c-Fos/c-Jun complex (Fig. 2). The barcoded protein library was prepared by cDNA display, during which DNA templates for *c-fos, c-jun* and glutathione S-transferase (*gst*) were pooled at a ratio of 1 : 1 : 10, and *gst* was used as a negative control. The DNA library consists of bait DNA containing three TRE sequences, and mutant DNA without the ability to bind c-Fos/c-Jun with a proportion of 1 : 10 (Table S1). We performed the solid-phase and solution-phase selection to enrich the

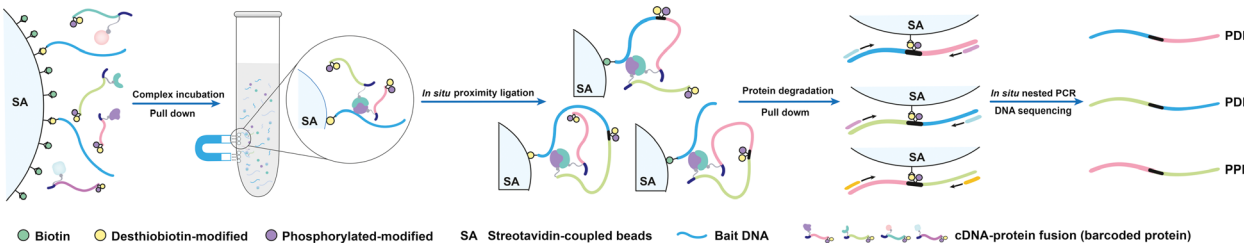

Fig. 1. One-Pot-seq for PDI/PPI complexes. Following the solution-phase screening, the barcoded proteins prepared by cDNA display and bait DNA are further enriched by two-step solid-phase purification. The first step uses the desthiobiotin tag in the bait DNA to eliminate unbound proteins, during which one tenth of the surface of beads is used and the rest surface is blocked with free biotin; meanwhile, the proximal blunt end ligation was performed *in situ*. The desthiobiotin tag on the barcoded proteins was utilized in the second-step purification to remove unbound DNA. Subsequently, *in situ* nested PCR amplification with different universal primer pairs is performed and DNA sequencing is used to decode the PPI/PDI involved in the transcriptional complex.

specific interaction complex. In the solid-phase selection, DNA library was immobilized using Dynabeads MyOne C1 and then incubated with the barcoded protein library. The specific interaction complexes were separated through two-step solid-phase selection. In the solution-phase selection, the barcoded protein library was incubated with the DNA library. The interaction complex, which has lower electrophoretic mobility than

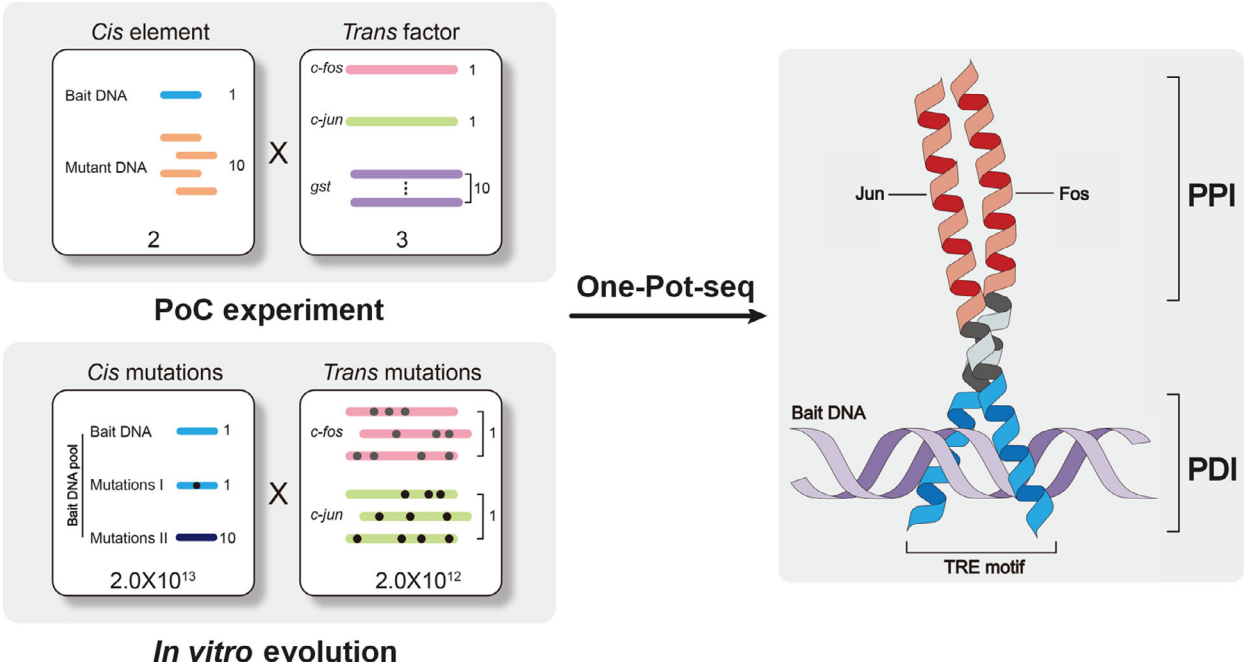

Fig. 2. A highly parallel assay for exploring the affinity and specificity of the c-Fos/c-Jun transcriptional complexes with One-Pot-seq. In PoC experiment, we conducted a 2 × 3 library-by-library screen, the *cis* element was composed of bait DNA, consisting of three TPA responsive element (TRE) sequences and mutant DNA without the ability to bind c-Fos/c-Jun, at a proportion of 1 : 10, the *trans* factors (barcoded proteins) were prepared by cDNA display, during which DNA templates for *c-fos, c-jun* and glutathione S-transferase (*gst*) were pooled a ratio of 1 : 1 : 10, and *gst* was used as a negative control. In *in vitro* evolution, the assay was performed in a high-order complex mixture, 2.0×10^{12} *cis* mutations × 2.0×10^{13} *trans* mutations. The *cis* mutations consisted of three components: the bait DNA, mutants I and II, the mutant I comprised six tandem nucleotide fragments (TGANTCA), which included a modification of the fourth base of the TRE motif to N (N = A, G, C or T), whereas the mutant II was composed of six NNNNNNN fragments (See Materials and methods for details). The final DNA pool was mixed with a ratio of 1 : 1 : 10, and the diversity of 3 μg of this DNA pool was 2.0×10^{13}. The *trans* mutations were prepared with a 30 pmol randomly mutated c-fos/c-jun library, of which the diversity was approximately 2.0×10^{12}. After the iterate rounds One-Pot-seq selection, the native c-Fos/c-Jun transcriptional complexes showed the best performance in terms of selectivity and affinity in both the PoC experiment and the *in vitro* evolution.

free DNA and barcoded protein, were separated by EMSA. After one round of selection, the interaction complex was analysed by pilot PCR which use the barcodes of the proteins as templates. This analysis showed that in both kinds of selection, c-fos and c-jun were enriched, whereas gst was largely filtered out (Fig. 3a), indicating solution-phase and solid-phase selection were effective.

Next, we performed the *in situ* proximity ligation in the solid-phase selection system. Two-step consecutive affinity purification was conducted to improve the specificity of the approach (Fig. 1). In order to reduce ligation probability of the intercomplexes, bait DNA was immobilized at a low surface density on streptavidin-coated beads, which occupied approximately one tenth of the bead surface area. This occupancy ration (0.1) is much lower than that used in the tethered conformation capture method (0.4) [27]. After the proximity ligation, we subsequently purified the ligation products to remove the unbound DNA with streptavidin-coated beads through the desthiobiotin tag in the second strand of cDNA. Given that the 5′ of protein barcodes and 3′ of bait DNA are linkable, the ligation patterns of PPI and PDI are unique, as shown in Fig. 4. In the meantime, the determinative ligation form also facilitates the follow-up *in situ* nested PCR, because the primers are only designed along the ligation direction.

The captured PPI and PDI ligation products were amplified by *in situ* nested PCR with different primer pairs. The PCR products were cloned and randomly chosen for sequencing. The results were used to define the PPI/PDI involved in the transcriptional complexes. Based on the experimental design, the expected size of PDI and PPI lengths would be around 420 bp and 720 bp respectively (Fig. S1). The results showed that all of the ligated sequences corresponded to c-Fos-c-Jun, c-Fos-DNA and c-Jun-DNA and lacked GST sequences, bait DNA mutants or c-Jun-c-Jun ligation (Fig. 3b,c). The absence of c-Jun-c-Jun ligation products is consistent with a previous report that the c-Fos-c-Jun heterodimer, rather than the c-Jun-c-Jun homodimer, binds DNA [28,29]. Therefore, the combined use of cDNA display and proximity ligation provides a platform to determine transcriptional complexes in one pot.

In vitro evolution by One-Pot-seq

After One-Pot-seq showed its feasibility, the *in vitro* evolution of c-Fos/c-Jun complex was performed based on this method to investigate whether the conventional interaction complex reached the thermodynamic summit. The content of the initial gene pool (*i.e.* c-Fos, c-Jun and binding DNA fragment library) used in evolution is important. A higher content gene pool is more possible to contain proteins and DNA fragments capable of forming stable complexes. The barcoded protein used in One-Pot-seq was constructed by cDNA display, which can produce high-content libraries ($> 10^{12}$). The randomly mutated *c-fos*/*c-jun* DNA templates which were used for cDNA display were generated using eight consecutive rounds of serial transfer mutagenic PCR [24]. The *c-fos*/*c-jun* mutants were cloned and randomly chosen for sequencing. The results showed that the average extent of mutagenesis was approximately 8% at the amino acid level, and the mutated sites were evenly distributed. Given that each DNA mutant was unique, and the complexity of the library was determined by the total number of molecules generated in the mutagenic PCR, we inferred that the diversity of a 30 pmol *c-fos*/*c-jun*

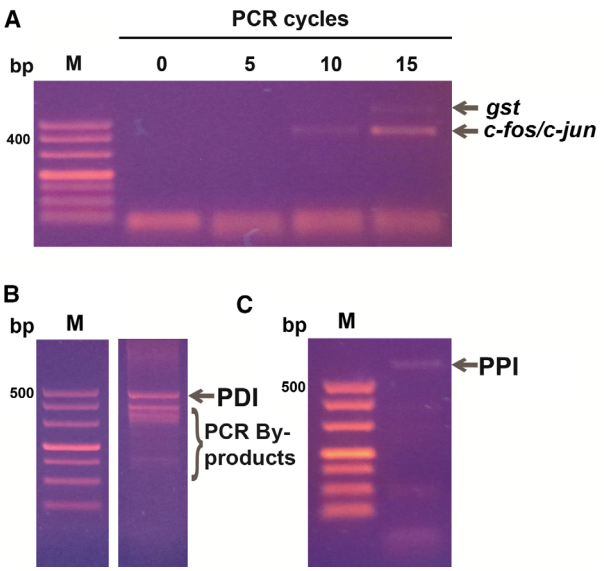

Fig. 3. Affinity enrichment of DNA-binding protein complexes. PoC experiment results of c-Fos/c-Jun transcriptional complex by One-Pot-seq. (A) Enrichment of c-Fos and c-Jun transcriptional factors. Pilot PCR with universal primer MT7/LHR-59 was used to detect *c-fos/c-jun* and *gst*. Lanes 1/2 are the 10th round PCR products and lanes 3/4 are the 20th PCR products. Lanes 1/3 were amplified with the solid-phase selection template and lanes 2/4 were amplified with the solution-phase selection template. This result shows that both *c-fos* and *c-jun* were enriched, whereas *gst* was largely filtered out. (B) Final PCR products of PDI ligation. (C) Final PCR products of PPI ligation. Based on the design, the expected sizes of the different amplification products are known. The lengths of PDI and PPI were around 420 bp and 720 bp respectively. These results indicate that proximity ligation was successful. Nested PCR with different primer pairs were used to amplify the PDI and PPI ligation products.

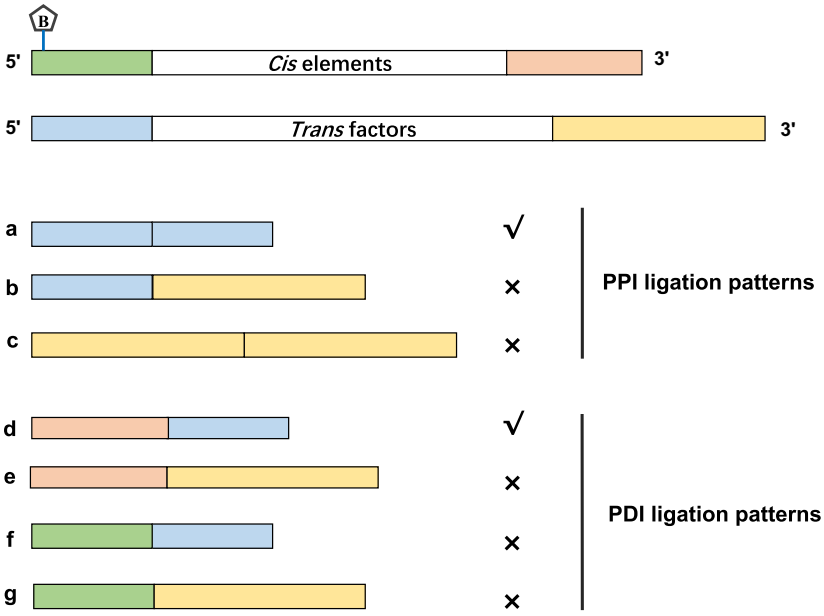

Fig. 4. Different ligation patterns for protein–protein and protein–DNA interactions (PPI and PDI). For PPI, 5′ of *trans* factors are ligated. For PDI, 3′ of *cis* element ligated with the 5′ of *trans* factor.

library (*trans* mutant library) was approximately 2×10^{12} [30]. The DNA fragment library (*cis* mutant library)consisted of cognate c-Fos/c-Jun-binding DNA fragment and their mutants produced by random synthesis. The complexity of a 3 μg DNA library is 2×10^{13}.

The *cis* and *trans* mutant libraries of c-Fos/c-Jun transcriptional system were incubated and subjected to five successive rounds in PoC experiment by EMSA (Fig. 2). To evaluate the enrichment efficiency, the fifth round of selected DNA was amplified by PCR and sequenced on a Hi-seq instrument to obtain paired-end 100 + 100 nucleotide reads, which can cover the full length of bait DNA. A total of 1.59 Gb were sequenced, and basic information regarding the sequencing results is shown in Table S2. The cognate c-Fos/c-Jun-binding DNA fragment was efficiently enriched, with the occupancy increased from 8.3% to 18.8% in the total sequences. In the random sequences, the most enriched sequences also consisted of the cognate c-Fos/c-Jun-binding motif (Fig. S2), accounting for 1.4% of the total random reads. The enriched protein and DNA were further purified by solid-phase purification, during which proximity ligation was performed *in situ*. The finally captured ligation products were amplified and cloned. Of the ligated sequences, the wild-type PPI/PDI interactions are predominant, with 23% (7/30) corresponded to c-Fos-c-Jun, 31% (14/45) corresponded to c-Fos-DNA and 33% (9/27) corresponded to c-Jun-DNA, indicating that the wild-type c-Fos/c-Jun-DNA complex is most stable (Fig. 5), which is consistent with previous

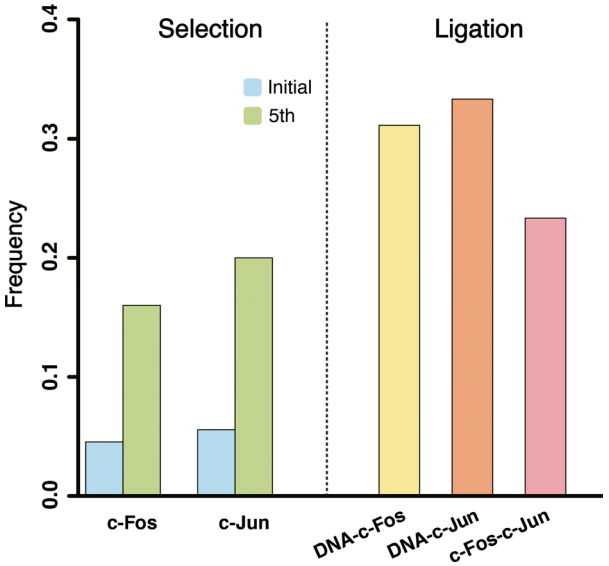

Fig. 5. *In vitro* evolution results of c-Fos/c-Jun transcriptional complex by One-Pot-seq. The frequency of wild-type c-Fos/c-Jun and their PDI/PPI complexes indicates that wild-type c-Fos and c-Jun were enriched after 5-round selection (left). After proximity ligation, the wild-type c-Fos/c-Jun was further enriched, the frequency of PDI/PPI complexes reaches 31% (c-Fos-DNA, yellow), 33% (c-Jun-DNA, orange), and 23% (c-Fos-c-Jun, pink) respectively (right).

reports [31–33]. This result strongly suggests that in the current evolutionary space (2.0×10^{12} *trans* mutations $\times$ 2.0×10^{13} *cis* mutations), the wild-type c-Fos/c-Jun system represents a thermodynamic summit, which well explains its evolutionary conservation.

The evolutionary conservation is usually the result of functional constraints. The extreme conservation observed in the c-Fos/c-Jun complex shows its important role in metazoans. Indeed, c-Fos/c-Jun regulated transcription participates in several most basic biologic process, including cell proliferation, differentiation, and apoptosis [34,35]. c-Fos/c-Jun was also involved in the pathogenesis of some diseases, such as cancer and immune diseases [36]. Since the conformation and composition of transcription complex are dependent on DNA binding [31,37,38], a precise elucidation of transcriptional regulation mechanism needs the determination of PPI/PDI involved in transcriptional complexes. One-Pot-Seq provides a solution to synchronously determine PPI and PDI, which is useful to clarify the mechanisms of transcription, and may be conducive to the studies of related biologic process and diseases. Moreover, the present methodology has a high content and supports high-throughput sequencing (a detailed description is presented in Fig. S3). Thus, this method could be expected to find applications in evolutionary biology, synthetic biology and drug discovery (peptide drug screening in particular).

Summary

Thermodynamic stability of biomacromolecule complexes were found positively correlated with their evolutionary conservation. In this study, using a combination of cDNA display and proximity ligation, a method to determine the PPI/PDI of transcriptional complexes in one pot was established, which we have called the One-Pot-seq method. On the basis of this method, we performed the *cis* and *trans in vitro* coevolution of c-Jun/c-Fos transcriptional complexes in a large evolutionary space (2.0×10^{12} *trans* mutations $\times$ 2.0×10^{13} *cis* mutations), and found that the conventional complex represents a thermodynamic summit, which is consistent with the extreme evolutionary conservation of the c-Fos/c-Jun system. In addition, this study proved the feasibility of the One-Pot-seq method. The high content, compatibility with high-throughput sequencing, and the ability to synchronously determine PPI and PDI confers this method with great potential in future applications.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant 31670779).

Author contributions

HYZ conceived the project. SKK and HYZ designed experiments. SKK, XYC, TT and BXC performed experiments. SKK, PFD and HYZ analysed data. HYZ, XYC, SKK and TT wrote the paper.

References

1 Deppmann CD, Alvania RS and Taparowsky EJ (2006) Cross-species annotation of basic leucine zipper factor interactions: Insight into the evolution of closed interaction networks. *Mol Biol Evol* **23**, 1480–1492.
2 Amoutzias GD, Veron AS, Weiner J 3rd, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG and Robertson DL (2007) One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity. *Mol Biol Evol* **24**, 827–835.
3 Reinke AW, Baek J, Ashenberg O and Keating AE (2013) Networks of bZIP protein-protein interactions diversified over a billion years of evolution. *Science* **340**, 730–734.
4 Kurafeiski JD, Pinto P and Bornberg-Bauer E (2018) Evolutionary potential of cis-regulatory mutations to cause rapid changes in transcription factor binding. *Genome Biol Evol* **11**, 406–414.
5 Piehler J (2005) New methodologies for measuring protein interactions *in vivo* and *in vitro*. *Curr Opin Struct Biol* **15**, 4–14.
6 Zhu J, Larman HB, Gao G, Somwar R, Zhang Z, Laserson U, Ciccia A, Pavlova N, Church G, Zhang W *et al.* (2013) Protein interaction discovery using parallel analysis of translated ORFs (PLATO). *Nat Biotechnol* **31**, 331–334.
7 Gu L, Li C, Aach J, Hill DE, Vidal M and Church GM (2014) Multiplex single-molecule interaction profiling of DNA-barcoded proteins. *Nature* **515**, 554–557.
8 Slattery M, Zhou T, Yang L, Dantas Machado AC, Gordân R and Rohs R (2014) Absence of a simple code: how transcription factors read the genome. *Trends Biochem Sci* **39**, 381–399.
9 Byrum SD, Raman A, Taverna SD and Tackett AJ (2012) ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. *Cell Rep* **2**, 198–205.
10 Byrum SD, Taverna SD and Tackett AJ (2013) Purification of a specific native genomic locus for proteomic analysis. *Nucleic Acids Res* **41**, e195.
11 Kennedy-Darling J, Guillen-Ahlers H, Shortreed MR, Scalf M, Frey BL, Kendziorski C, Olivier M, Gasch AP and Smith LM (2014) Discovery of chromatin-associated proteins via sequence-specific capture and mass spectrometric protein identification in *Saccharomyces cerevisiae*. *J Proteome Res* **13**, 3810–3825.
12 Kaltenbach M and Hollfelder F (2012) SNAP display: *in vitro* protein evolution in microdroplets. *Methods Mol Biol* **805**, 101–111.

13 Clackson T, Hoogenboom HR, Griffiths AD and Winter G (1991) Making antibody fragments using phage display libraries. *Nature* **352**, 624–628.
14 Hanes J and Pluckthun A (1997) *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci USA* **94**, 4937–4942.
15 Roberts RW and Szostak JW (1997) RNA–peptide fusions for the *in vitro* selection of peptides and proteins. *Proc Natl Acad Sci USA* **94**, 12297–12302.
16 Nemoto N, Miyamoto-Sato E, Husimi Y and Yanagawa H (1997) *In vitro* virus: bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome *in vitro*. *FEBS Lett* **414**, 405–408.
17 Yamaguchi J, Naimuddin M, Biyani M, Sasaki T, Machida M, Kubo T, Funatsu T, Husimi Y and Nemoto N (2009) cDNA display: a novel screening method for functional disulfide-rich peptides by solid-phase synthesis and stabilization of mRNA-protein fusions. *Nucleic Acids Res* **37**, e108.
18 Mochizuki Y, Biyani M, Tsuji-Ueno S, Suzuki M, Nishigaki K, Husimi Y and Nemoto N (2011) One-pot preparation of mRNA/cDNA display by a novel and versatile puromycin-linker DNA. *ACS Comb Sci* **13**, 478–485.
19 Lundberg M, Eriksson A, Tran B, Assarsson E and Fredriksson S (2011) Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res* **39**, e102.
20 Gustafsdottir SM, Schlingemann J, Rada-Iglesias A, Schallmeiner E, Kamali-Moghaddam M, Wadelius C and Landegren U (2007) In vitro analysis of DNA-protein interactions by proximity ligation. *Proc Natl Acad Sci USA* **104**, 3067–3072.
21 Tateyama S, Horisawa K, Takashima H, Miyamoto-Sato E, Doi N and Yanagawa H (2006) Affinity selection of DNA-binding protein complexes using mRNA display. *Nucleic Acids Res* **14**, e27.
22 Kang SK, Chen BX, Tian T, Jia XS, Chu XY, Liu R, Dong PF, Yang QY and Zhang HY (2015) ATP selection in a random peptide library consisting of prebiotic amino acids. *Biochem Biophys Res Commun* **466**, 400–405.
23 Wilson DS and Keefe AD (2000) Random mutagenesis by PCR. *Curr Protoc Mol Biol* **51**, 8.3.1–8.3.9.
24 Magoč T and Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963.
25 Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
26 Garner MM and Revzin A (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res* **9**, 3047–3060.
27 Kalhor R, Tjong H, Jayathilaka N, Alber F and Chen L (2011) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat Biotechnol* **30**, 90–98.
28 John M, Leppik R, Busch SJ, Granger-Schnarr M and Schnarr N (1996) DNA binding of Jun and Fos bZIP domains: homodimers and heterodimers induce a DNA confirmational change in solution. *Nucleic Acids Res* **24**, 4487–4494.
29 Ramidez-Carrrozzi V and Kerppola T (2003) Asymmetric recognition of nonconsensus AP-1 sites by Fos-Jun and Jun-Jun influences transcriptional cooperatively with NFAT1. *Mol Cell Biol* **23**, 1737–1749.
30 Olson CA, Nie J, Diep J, Al-Shyoukh I, Takahashi TT, Al-Mawsawi LQ, Bolin JM, Elwell AL, Swanson S, Stewart R *et al*. (2012) Single-round, multiplexed antibody mimetic design through mRNA display. *Angew Chem Int Ed* **51**, 2957–2963.
31 Seldeen KL, Deegan BJ, Bhat V, Mikles DC, McDonald CB and Farooq A (2011) Energetic coupling along an allosteric communication channel drives the binding of Jun-Fos heterodimeric transcription factor to DNA. *FEBS J* **278**, 2090–2104.
32 Seldeen KL, McDonald CB, Deegan BJ and Farooq A (2009) Single nucleotide variants of the TGACTCA motif modulate energetics and orientation of binding of the Jun-Fos heterodimeric transcription factor. *Biochemistry* **48**, 1975–1983.
33 Seldeen KL, McDonald CB, Deegan BJ, Bhat V and Farooq A (2009) DNA plasticity is a key determinant of the energetics of binding of Jun-Fos heterodimeric transcription factor to genetic variants of TGACGTCA motif. *Biochemistry* **48**, 12213–12222.
34 Angel P and Karin M (1991) The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. *Biochem Biophys Acta* **1072**, 129–157.
35 Ameyar M, Wisniewska M and Weitzman JB (2003) A role for AP-1 in apoptosis: the case for and against. *Biochimie* **85**, 747–752.
36 Wagner EF and Eferl R (2005) Fos/AP-1 proteins in bone and the immune system. *Immunol Rev* **208**, 126–140.
37 Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, Taipale M, Vaquerizas JM, Yan J, Sillanpää MJ *et al*. (2010) Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res* **20**, 861–873.
38 Li X, Wang W, Wang J, Malovannaya A, Xi Y, Li W, Guerra R, Hawke DH, Qin J and Chen J (2015) Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. *Mol Syst Biol* **11**, 775–792.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. The expected sequences of the proximity ligation products.

Fig. S2. The most prevalent sequence of bait DNA and protein.

Fig. S3. Preparation of One-Pot-seq samples for high-throughput sequencing.

Table S1. Oligonucleotides used for One-Pot-seq.

Table S2. Basic information of NGS dataset.