Reprogramming cells with synthetic proteins

Xiaoxiao Yang*, Vikas Malik*, Ralf Jauch

Conversion of one cell type into another cell type by forcibly expressing specific cocktails of transcription factors (TFs) has demonstrated that cell fates are not fixed and that cellular differentiation can be a two-way street with many intersections. These experiments also illustrated the sweeping potential of TFs to “read” genetically hardwired regulatory information even in cells where they are not normally expressed and to access and open up tightly packed chromatin to execute gene expression programs. Cellular reprogramming enables the modeling of diseases in a dish, to test the efficacy and toxicity of drugs in patient-derived cells and ultimately, could enable cell-based therapies to cure degenerative diseases. Yet, producing terminally differentiated cells that fully resemble their in vivo counterparts in sufficient quantities is still an unmet clinical need. While efforts are being made to reprogram cells non-genetically by using drug-like molecules, defined TF cocktails still dominate reprogramming protocols. Therefore, the optimization of TFs by protein engineering has emerged as a strategy to enhance reprogramming to produce functional, stable and safe cells for regenerative biomedicine. Engineering approaches focused on Oct4, MyoD, Sox17, Nanog and Mef2c and range from chimeric TFs with added transactivation domains, designer transcription activator-like effectors to activate endogenous TFs to reprogramming TFs with rationally engineered DNA recognition principles. Possibly, applying the complete toolkit of protein design to cellular reprogramming can help to remove the hurdles that, thus far, impeded the clinical use of cells derived from reprogramming technologies.

Asian Journal of Andrology (2015) 17, 394–402; doi: 10.4103/1008-682X.145433; published online: 09 January 2015

Keywords: cellular reprogramming; protein design; protein engineering; synthetic transcription factors; transactivation

SWITCHING CELL FATES WITH TRANSCRIPTION FACTOR PROTEINS

The notion that organismic development does not stubbornly follow a predetermined path but is a rather plastic process has been made long before DNA was recognized as the carrier of inheritable information.1 Nearly 100 years later, cellular reprogramming was first demonstrated when live tadpoles arose after transferring the nuclei of frog intestine derived somatic cells into oocytes.2 This observation led to the realization that even fully differentiated cells must still contain the complete genetic blueprint required to build a whole organism. That transcription factors (TFs) are remarkably powerful to drive cellular reprogramming, that is to convert one cell type into another, was first demonstrated by turning a mouse fibroblast into a muscle cell with a cDNA encoding a single TF, MyoD (Myod1).3 Subsequently, a cocktail of four TFs, Krüppel-like factor 4 (Klf4), c-Myc, Sry (sex determining region y) box2 (Sox2) and octamer binding protein 4 (Oct4), was discovered to induce pluripotent stem cells (iPSCs) when forcibly expressed in fibroblasts of mouse4 and human.5,6 Since this seminal work, cellular reprogramming has become a mainstream research activity. Like embryonic stem cells derived from the blastocyst, iPSCs can be passaged in culture for indefinite periods and given appropriate growth conditions, can be differentiated into all cell types of the body.7 The latter can straightforwardly be demonstrated by transplanting iPSCs back into blastocysts, but it is a challenge to recapitulate this process in vitro.

The MyoD experiment already demonstrated that differentiated cells do not have to be pushed up all the way the “Waddington canal” to a completely undifferentiated cell type before subsequent re-specification to an alternative cell fate. Instead, the MyoD conversion of fibroblasts directly to muscle indicates that direct transdifferentiation can be accomplished without a pluripotent intermediate. Yet, analogous to the MyoD example, lineage conversion was initially only reported between cell types originating from a similar developmental trajectory such as the interconversion of two immune cell types, B cells (lymphoid cells) to macrophages (myeloid cells),5 exocrine to endocrine pancreatic cells,6 glial cells to neurons,7 brown fat to muscle cells8 and fibroblasts to cardiomyocytes.9 All of these transdifferentiation events are within the same germ layer (i.e. mesoderm to mesoderm or neuroectoderm to neuroectoderm). Eventually, conversions between cells originating from different germ layers could also be achieved as fibroblast as well as hepatocytes could be directly converted into functional neurons with defined TF cocktails.10,11 Apparently, even pronounced reprogramming barriers separating very distant cellular states can be crossed with small sets of lineage specifying TF proteins. The appreciation of this astonishing developmental plasticity sparked a plethora of studies attempting to re-direct cell fates. For example, blood cells,12 endothelial cells,13 hepatocytes,14–16 sertoli cells17 and thymic epithelial cells18 could be successfully generated with TF cocktails. Readers are referred to excellent reviews that discuss the progress of cellular reprogramming and lineage conversions in more detail.19–23 Remarkably, particularly potent reprogramming factors such as Oct4 were found to be able to
induce reprogramming alone in certain cell types and other cell types with small molecule supplementation.\textsuperscript{34–36} Moreover, a few studies have reported that TFs can be omitted altogether as induced pluripotent or neuronal cells could be produced solely with small molecules.\textsuperscript{37,38} However, efficiency of chemical reprogramming is considerably lower than TF based approaches and applications currently remain limited to mice. Hence, TF based cell lineage conversions continue to be the most effective and versatile approach. Therefore, we will focus our discussion on efforts to enhance TF based cell lineage conversions through protein engineering.

**ROADBLOCKS ON THE WAY TO FUNCTIONAL CELLS**

The excitement sparked by cellular reprogramming is catalyzed by its promise to lead to new clinical applications. One strategy is to conduct \textbf{“in vitro clinical trials”}\textsuperscript{26,39} That is, cells obtained from patients through biopsies, blood or urine samples are differentiated into disease-relevant cell types. Next, preselected drugs or drug libraries can be assessed for their toxicity and potential to exert curative effects on those cells. It is hoped that this approach will accelerate personalized therapies, facilitate drug discovery and avoid the prescription of drugs that are toxic or ineffective to certain patient populations. Moreover, reprogramming technologies can be used to model human diseases in a dish. Here, the behavior of cells derived from patients is compared to cells from healthy donors. If disease-causing mutations are known, the mutation can be engineered using genome editing technologies and genetically matched isogenic cell lines can be studied. This way, diseases can be understood at an unprecedented depth, cellular pathways can be mapped, biomarkers can be discovered and therapeutic strategies can be developed. Lastly, the holy grail of stem cell research is to produce functional cells that can be transplanted back into patients to remedy degenerative diseases.\textsuperscript{40} Encouragingly, diseases could be cured through cell therapies in animal models. For example, gene-corrected iPSC derived hematopoietic progenitors transplanted back into humanized sickle cell anemia mouse models could cure the animals.\textsuperscript{41} This has led to the hope that diseases caused by deficiencies in well-defined cell types such as type 1 diabetes,\textsuperscript{42} Parkinson’s disease\textsuperscript{43} and retinal degeneration\textsuperscript{44} are curable with cell-based therapies. Though, hematopoietic stem cells have been used in bone marrow transplants since the 1950’s, cell therapies in humans still pose major challenges, and daunting roadblocks remain. Most importantly, safety has to be rigorously assessed before transplanting the reprogrammed cells. iPSCs resemble cancer cells in many ways and are teratogenic when injected into mice. This poses a significant risk as incomplete differentiation, and remnant pluripotent cells could potentially lead to cancerous growth.\textsuperscript{45,46} Collectively, avoiding insertional mutagenesis, oncogenic TFs and pluripotent reprogramming intermediates could solve this problem. Furthermore, it is often problematic to terminally differentiate cells so that they fully replicate the function of the cells matured \textit{in vivo}. Cells have to be stable and need to be expandable so that they can be produced in sufficient quantities needed to support transplantation medicine. Ideally, reprogramming strategies should leave the genome unscathed, utilize cells that are genetically matched to the recipient with just the disease-causing loci corrected and produce an epigenetic state identical to the tissue embedded cell they are meant to replace. While optimized factor cocktails, novel culture conditions and small molecule compounds will likely further advance reprogrammed cells toward the clinic, we surmise that the engineering of the reprogramming TFs themselves provides a viable strategy to be further explored.

**DESIGNING BETTER PROTEINS**

Bioengineering proteins to either enhance their activity or to install completely novel functions has been successfully accomplished in numerous instances. Day-to-day laboratory operations utilize a range of artificially enhanced proteins. Those include DNA polymerases with thoroughly optimized fidelity,\textsuperscript{47–49} proteases with engineered activity and substrate specificity\textsuperscript{50} and fluorescent proteins with increased brightness.\textsuperscript{51,52} Likewise, protein therapeutics are often rationally improved. In particular coagulation factors to treat bleeding disorders were bioengineered in a variety of ways.\textsuperscript{53} For example, factor IX was engineered to have prolonged activity by fusing it to a Fc fragment\textsuperscript{54} and the coagulation factor VIIIa was rationally mutagenized for inactivation resistance and optimized secretion profiles.\textsuperscript{55,56} More ambitious goals include the engineering of whole pathways leading to the biotechnological synthesis of new products.\textsuperscript{57,58}

What are the methods protein designers use to achieve their engineering goals? A rather simple way is to concatenate functional protein domains or even whole proteins. Examples include fusions of green fluorescent protein with antibody fragments that increase their brightness\textsuperscript{59} or attaching effector domains such as nuclease to artificial TFs with customized DNA sequence preferences.\textsuperscript{60} In addition, functional regions such as phosphorylation sites, protease cleavage sites, and signaling sequences can be rationally modified to install desired properties.\textsuperscript{55,58} Most commonly, rational and randomization strategies are combined to achieve the desired results. Using the knowledge of the protein’s sequence, structure, conservation and functional insights gained from site-directed mutagenesis experiments can lead to the selection of functionally important structural elements. Such elements could be individual or a small set of amino acids, secondary structure elements or subdomains. Frequentely, design efforts target catalytic centers, substrate binding pockets or macromolecular contact interfaces. Those elements can then be modified taking biophysical parameters such as charge, size, and hydrophobicity, as well as functional data and sequence information of homologs into account. Yet, rationally predicting how a specific structural modification affects protein activity is a daunting task as our understanding about the structural basis for protein function remains limited. Therefore, protein designers often subject, structural elements earmarked for protein optimization to directed evolution.\textsuperscript{61} This strategy requires a carefully designed randomization strategy, which can include error-prone polymerase chain reaction,\textsuperscript{62} site-directed mutagenesis with randomized oligos\textsuperscript{63} and “chimeragenesis,” that is the recombination of protein fragment libraries.\textsuperscript{64} Libraries of modified proteins now undergo a screening and selection procedure to identify variants with improved functionality. Selection systems include binding assays such as phage display,\textsuperscript{65} ribosome display,\textsuperscript{66} enzymatic assays,\textsuperscript{67} tests for protein stability,\textsuperscript{68} genetic complementation combined with phenotypic read-outs\textsuperscript{69} and \textit{in vitro} compartmentalization.\textsuperscript{70} Obviously, selection system design is critical as desired protein variants would escape detection if the screen cannot rigorously discriminate between enhanced and unwanted variants of the designed protein.\textsuperscript{71}

Remarkably, efforts are being made to design proteins entirely from scratch using fragment libraries of nonnatural peptide sequences with minimal architectural constraints. Given the mindboggling number of theoretically possible protein sequences this seems like a herculean feat. Nevertheless, \textit{de novo} design has led to the creation of some functional sequences.\textsuperscript{71,72} Thus far, examples for the engineering of TF proteins are still rather rare. Here we ask whether the toolkit of protein engineering could be employed to design reprogramming TFs to more
Effectively engineer cell lineage conversions and to bring progress to regenerative biomedicine.

**ENGINEERING SYNTHETIC REPROGRAMMING FACTORS**

**Enhancing reprogramming efficiency with potent transactivation domains**

The optimization of reprogramming strategies has been a priority for many laboratories as the original protocol was rather inefficient. Efficiency enhancements could be achieved by supplementing the media, altering the factor cocktails, changing the sequence of factor addition, adding small molecules or removing reprogramming roadblocks. In addition, some studies resorted to protein engineering to improve reprogramming (Table 1). Based on the assumption that reprogramming TFs mainly act by inducing mRNA synthesis of their target genes, several engineering efforts were made to increase the transactivation potential of TFs by fusing them to potent transactivation domains (TADs) (TAD-TFs, Figure 1).

**Viral protein 16-transactivation domain**

The viral protein 16 (VP16) is a 490 amino acids TF protein of the herpes simplex virus with strong transactivating activity (Figure 1a). Its potent TAD was mapped and molecularly dissected more than 25 years ago and found to consist of an acidic C-terminal region spanning approximately 80 amino acids. Immediately after its discovery the VP16-TAD has been utilized to engineer chimeric TFs with enhanced activity. More recently, VP16 has also been utilized to enable cellular reprogramming. When the VP16-TAD was fused to the Xenopus ortholog of the pancreatic and duodenal homeobox1 (Pdx1), a chimeric protein could induce the conversion of liver cells to pancreatic cell in transgenic tadpoles. A similar Pdx1-VP16 fusion induced insulin biosynthesis and ameliorated glucose tolerance in mouse diabetic models. In an effort to enhance iPSC formation the VP16-TAD was fused to pluripotency reprogramming factors. In this study, a core fragment of the VP16-TAD (residues 446-490) was attached to Oct4, Sox2, Klf4 and Nanog TFs separated by a glycine-rich linker (Figure 1a). With the exception of Klf4, the engineered TAD-TFs substantially outperformed the wild-type proteins with regards to both the efficacy and the kinetics of iPSC generation in mouse and human cells. Moreover, Oct4-VP16 alone could efficiently reprogram mouse embryonic fibroblasts into germline-competent iPSCs. An Oct4 construct containing three C-terminal VP16 copies arranged in tandem exhibited the highest efficiency (Figure 1b).

A separate study also reported that fusions of VP16 to mouse Oct4, human Oct4 and Xenopus Xlpou91 could support reprogramming as well as rescue Oct4 null ESCs. However, the authors did not observe a substantial enhancement in the reprogramming efficiency by the engineered proteins. The differences between the two studies could be caused by variations in the reprogramming conditions and construct.

**Table 1: Engineered reprogramming factors**

| Protein modified | Type of modification | Reprogramming factor modulation with potent TADs | Effects on reprogramming and differentiation | References |
|------------------|----------------------|-----------------------------------------------|--------------------------------------------|------------|
| Oct4, Sox2, Nanog | Fusing the MyoD TAD to Oct4 | VP16-TAD linked to Oct4, Sox2 and Nanog | Enhanced reprogramming of mouse and human fibroblasts; single factor reprogramming with Oct4-VP16 | 93 |
| Oct4 and Xlpou91 | VP16 linked to Oct4 and its Xenopus orthologue | Sox17-TAD linked to Sox2, Sox18EK and Sox4EK | iPSC induction similar to WT cocktail | 95 |
| Sox2, Sox18, Sox4 | YAP-TAD linked to Oct4, Sox2 and Nanog | VP16-TAD fused to Xlpou91 Pdx1 | Conversion of liver to pancreas in Xenopus tadpoles | 90 |
| Pdx1 | VP16-TAD fused to mouse Pdx1 | Induces insulin synthesis and ameliorates glucose tolerance in diabetic mouse models | 91,92 |
| MyoD | DNA-binding domain of MyoD fused to the TAD of myocardin | Improves directed differentiation of human mesenchymal stem cells into skeletal myocytes | 149 |
| MeI2c | MyoD TAD linked to MeI2c | Targeted activation of endogenous reprogramming factors | 102 |
| TALEs | Designer TALEs with VP64 targeting the distal Oct4 enhancer | Substitutes for exogenous Oct4 to generate iPSCs with comparable efficiency | 120 |
| MyoD and E12 | Swap of three amino acids from MyoD into E12 | Mutant E12 can convert fibroblasts into myocytes | 122 |
| Sox17 and Sox2 | Point mutation in the DNA binding domain of Sox17 | Conversion of Sox17 into an iPSC inducer with about 4-fold greater efficiency than Sox2 | 142 |

TAD: transactivation domain; Oct4: octamer binding protein 4; iPSC: induce pluripotent stem cells; MEFS: mouse embryonic fibroblasts; TALEs: transcription-activator-like effectors; YAP: yes-associated protein; Pdx1: pancreatic and duodenal homeobox 1; OSKM: Oct4, Sox2, Klf4, and c-Myc; VP16: viral protein 16; Klf4: krüppel-like factor 4; WT: wild-type.
Sox2-YAP fusion was reported to be most critical for the acceleration of iPSC generation. The kinetics of iPSC formation was markedly accelerated with iPSC colonies appearing on day 10, and the efficiency rose from <1% to ~40%. Furthermore, the reprogramming efficiency of iPSCs was improved when using the A-dTF the overall iPSC colony yield was increased to 20% of the Mef2c levels. Mef2c-VP16 fusions also showed some increase in iCM formation, and the position of the TADs at either the N- or C-termini of the TFs, as well as the inclusion of linker sequences.

Inducing endogenous reprogramming factors with TAL effectors

Artificial proteins based on C2-H2 zinc-finger proteins (ZFPs), transcription activator-like effectors (TALEs) and RNA-guided nuclease effector domains that enable genome editing at single base-pair resolution (reviewed in 94,97,98). While off-target effects have been a lurking concern, whole genome sequencing studies demonstrated that unwanted modifications are very rare. Recently, designer TALEs (dTALEs) and ZFPs have also been used to engineer transcriptional activators. TALEs consist of 33-34 amino acid repeat domains with hypervariable residues at position 12 and 13. The identity of the dipeptide at positions 12/13 determines a recognition code (HD = C, NG = T, NI = A, NS = A, C, G or T, NN = A or G), which allows to rationally create dTALEs that recognize DNA sequences of choice. As up to 33 TALE repeat domains can be arranged in tandem, genomic loci can be targeted with high precision. Although, dTALE design is somewhat more straightforward than ZFP design, initial efforts were undertaken using ZFP-VP16 fusions constructed to target a sequence proximal to the transcriptional start site of Oct4. In another attempt, the fusion of a KRAB domain fused to a designer ZFP could activate endogenous Oct4 protein in a series of cell lines. This was a surprising observation because the KRAB conventionally acts as transcriptional repressor. dTALE-VP16 fusions designed to bind proximal promoter sequences of SOX2, Klf4, c-MYC and Oct4 could activate reporter constructs, but only dTALEs targeting Klf4 and SOX2 could also activate the endogenous genes in 293FT cells. Moreover, dTALEs targeting the proximal promoter of Oct4 could activate the gene in NSCs where it is otherwise silenced. However, this strategy required the addition of histone deacetylase or DNA methyltransferase inhibitors suggesting that some chromatin loosening is needed for the dTALE-TF to access its target site.

Gao et al. asked whether dTALE-TFs can replace conventionally used reprogramming TFs by activating their endogenous counterparts. The authors used VP64-TADs (four tandem repeats of VP16) to engineer designer transcription activators (A-dTF) that target the p300 or the mediator complex has not yet been studied in a systematic manner. Hence, whether those TADs mediate a general, or a TAD specific mechanism of transcriptional activation awaits further exploration. Collectively, there appear to be no obvious rules of how TAD-TF chimeras should be designed to engineer lineage converting TFs. Rather, optimal constructs had to be empirically produced for each TAD-TF combination. For example, increasing TAD copy numbers can either boost or impede TAD-TF activity. Further parameters to be optimized include the length of the TAD fragment used, the position of the TADs at either the N- or C-termini of the TFs, and the inclusion of linker sequences.
was higher when Oct4 was used directly. The authors went on to show that A-dTf, targeting a distal Nanog enhancer could convert epiblast stem cells into ESCs. This study provides an elegant proof-of-concept that TALE-based TFs can replace native reprogramming factors. However, as the sole function of dTALE-TFs is to activate endogenous reprogramming factors that would eventually have to finish the job, it remains to be demonstrated whether this method can enhance cell lineage conversions.

**Engineering chromatin association of reprogramming factors**

**Turning E12 into a myogenic transcription factor**

The TAD-TFs and TFs endogenously activated by A-dTfs will likely engage the genome in the same manner as the native TFs as the DNA recognition domain is not modified. So far, only a few engineering efforts focused on protein interfaces involved in DNA recognition that would alter their genomic binding profile. Still, several swap experiments that install new functions and create engineered reprogramming TFs have been successful. The reprogramming pioneer Weintraub had provided the first evidence that lineage conversion activity of reprogramming TFs can be radically interchanged with strategically placed point mutations at the DNA contact interface. Following the seminal discovery that MyoD alone can induce a myogenic program in fibroblasts, Weintraub et al. continued to dissect the molecular basis of its specific activity. MyoD belongs to the bHLH family of TFs whose members bind to short palindromic CANNNTG E-box motifs as homo- or heterodimers. By adopting a scissor-like architecture, bHLH TFs bind the major groove of the DNA through the basic regions of helix

![Figure 2](image-url)
specific functions (Figure 2a). Rather, just three amino acids at the DNA interaction surface specify their functional diversity. Subsequent studies suggested that the Ala-Thr dipeptide affects the conformation of Arg-111 in the basic region and thereby modulates the access of Arg-111 to the major groove of the DNA binding site.124 Those rearrangements at the DNA-binding interface could translate into allosteric events at other interfaces, such as binding sites for chromatin modifiers, and thereby influence the functional consequences of the binding event.125

**Turning Sox17 into a pluripotency inducer**

Our laboratory has made efforts to scrutinize the mechanism how proteins of the 20-member Sox family recognize their DNA target sites. Confusingly, all Sox proteins bind a near-identical CATTGT-like sequence126,127 and engage DNA by binding to the minor groove to induce a 70° kink using a conserved set of amino acids.128–130 How then can individual members select specific gene-sets to initiate characteristic cell fate decisions? The DNA binding HMG domain of Sox protein not only mediates sequence-specific DNA recognition but is also the main determinant of a partner code enabling selective interactions with other TFs.131–141 By conducting quantitative electrophoretic mobility shift assays to study the HMG mediated partnership with the Pit1-Oct-Unc-86 (POU) domain of Oct4, we observed different propensities of Sox-family members to heterodimerize with Oct4 on a series of differently configured composite sox-oct binding sites.132,133 In particular an unusual “compressed” element—where one nucleotide separating the sox and oct half sites is removed — still recruits Sox17/Oct4 dimers, whereas Sox2/Oct4 dimers can no longer assemble (Figure 2b). Conversely, the Sox2/Oct4 pair dimerizes markedly better on the canonical motif than the Sox17/Oct4 pair. In the search for the structural basis for these differences, a single amino-acid at position 57 of the HMG caught our attention. This residue, a Lys in Sox2 and a Glu in Sox17, shows a high degree of sequence variation amongst paralogous Sox proteins although it occupies a critical position at the Oct4 interaction interface.125,129,130,132,133 By exchanging this residue between Sox2 and Sox17 to produce Sox17EK and Sox2KE proteins, highly cooperative dimer formation of the Sox17EK/Oct4 complex on the canonical motif is installed. The wild type Sox2 normally partners with Oct4 in OKSM143 or OSNL (OS plus Nanog and Lin28)9 cocktails to activate the pluripotency circuitry. By contrast, the wild-type Sox17 induces endoderm differentiation when overexpressed in ESCs. The activity of the engineered factors was, therefore, studied in iPSC generation assays.142,143 When we replaced Sox2 with Sox17EK in OSKM cocktails, we could induce iPSCs with improved efficiency in both mouse144 and human cells.95 An analogously modified Sox7EK protein showed a similar behavior, whereas Sox4 and Sox18 needed additional TAD engineering for their conversion into reprogramming TFs.95 Using chromatin immunoprecipitation followed by high-throughput sequencing, we found that Sox17EK and Sox2 show a very similar binding profile when overexpressed in mouse ESCs.144 Both proteins pair with Oct4 on many genomic loci earmarked with canonical sox-oct motifs. By contrast, Sox17 partners with Oct4 on enhancers containing the compressed motif. Apparently, a single point mutation drastically changed how Sox proteins co-select their target genes by partnering with Oct4. Yet, the converse Sox2KE mutant could neither effectively dimerize with Oct4 on the canonical nor on the compressed sequence. This puzzle was resolved more recently when a novel Oct4 crystal structure and molecular dynamics simulation suggested an additional discriminatory interaction between residue 46 of Sox proteins with an Oct4 specific helix in the POU linker.145,146 Indeed, a rationally designed Sox2E46LK57E double mutant now cooperatively dimerizes with Oct4 on the compressed motif. It will be of interest to explore the activity of this engineered Sox factor in lineage conversion experiments.

Collectively, the MyoD and Sox17EK examples show that the cell fate conversion potential of reprogramming TFs can be drastically changed with rather minimal modifications at structurally critical interfaces. We surmise that these insights could be utilized to engineer more potent and safer reprogramming TFs. Contrary to the TAD-TF and the TALE-TF approach; TFs with engineered DNA-binding domain likely engage the genome in a new manner (Figure 3). This way, it could be possible to break reprogramming barriers more effectively and to direct cells trapped in a local minimum of the Waddington landscape towards a desired state of differentiation.

**OUTLOOK — NOVEL APPROACHES FOR REPROGRAMMING FACTOR DESIGN**

To produce cells for clinical applications, the process should be tightly controlled, fast and exclude undesired by-products. In particular,
reprogramming to pluripotency has witnessed a multitude of studies aimed to improve the efficiency of iPS cell generation (excellent reviews by Papp and Plath19 and Soufi20). Initially, iPSC generation was rather slow and only a small number of cells transfected with a cocktail of reprogramming TFs could be reprogrammed.4 Confusingly, it appeared that there is a high degree of randomness in cell populations and by simple chance a small subset of cells enters a path leading to the successive progression towards pluripotency in a more deterministic fashion.14,15 Yet, as roadblocks toward the pluripotency continue to be removed; fully controlled and efficient iPSC generation could soon be achieved.8,21 As the quality of iPSCs produced by engineered reprogramming factors was validated by examining their contribution to embryonic development and the capacity for germline transmission, synthetic TFs could still contribute to the ultimate cocktail.34,37,103,120 However, iPSCs are only an intermediary by-product on the way towards transplantation-grade functional cells. To lower the risk of cancerogenesis, a pluripotent intermediate should be avoided altogether or, minimally, complete differentiation of formerly pluripotent cells has to be ensured. Reproducibly generating functional cells to cure degenerative diseases will remain a challenge in the years to come. We anticipate that protein engineering techniques can help to overcome reprogramming barriers and better control cell lineage conversions to produce functional cells more safely and with properties more closely matching their in vivo counterparts (Figure 3). While widely used in fields such as enzymology and immunology, protein engineering is still in its infancy in cellular reprogramming. This is partly because of our incomplete understanding of how TFs work. Our structural knowledge is mostly restricted to isolated domains bound to short stretches of DNA. The mechanism of DNA target site selection, chromatin opening and how TFs stimulate mRNA synthesis remains largely unclear. Nevertheless, the studies highlighted in this review testify the promise of the approach and warrant further exploration as to whether protein engineering can bring stem cell biology closer to the bedside.

COMPETING INTERESTS

The authors declare that they have no competing interests.

ACKNOWLEDGMENTS

Our laboratory is supported by the people’s government of Guangzhou municipality Science and Technology Project 2011Y2-00026 and by a 2013 MOST-China-EU Science and Technology Cooperation Program, Grant No. 2013DFE33080. VM. thanks the CAS-TWAS President’s Fellowship and UCAS for financial and infrastructure support. We thank Andrew Hutchins, Iekai Chen, Ajaybabu Pobbati, Swaine Chen and Jörg Kahle for insightful comments.

REFERENCES

1. Haecckl EH. Zur Entwicklungs geschichte der Siphonophoren. Utrecht: C. Van Der Post Jr; 1869.
2. Gurdon JB. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. J Embryol Exp Morphol 1962; 10: 622–40.
3. Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts into embryonic Sertoli‑like cells by defined factors. Cell 2006; 126: 663–76.
4. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, ect al. Induction of functional heart muscle from mouse fibroblasts by defined factors. Cell 2010; 143: 508–25.
5. Wang H, Cao N, Spencer CI, Nie B, Ma T, et al. Small molecules enable cardiac reprogramming of mouse fibroblasts with a single factor, Oct4. Cell Rep 2014; 6: 951–60.
6. Wu G, Schöler HR. Role of Oct4 in the early embryo development. Cell Regen (Lond) 2014; 5: 7.
7. Cheng S, Hu W, Qi, Liu B, Yuan B, et al. Generation of functional neural cells from mouse embryonic stem cells. Cell Mol Biol 2010; 56: 515–22.
44 Jin ZB, Takahashi M. Generation of retinal cells from pluripotent stem cells. Proc Brain Res 2012; 201: 171–81.
45 Okita K, Ichisaka T, Yamanaka S. Generation of germine-competent induced pluripotent stem cells. Nature 2007; 448: 313–7.
46 Ohnishi K, Sermi K, Yamamoto T, Shizumiz M, Tanaka A, et al. Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation. Cell 2014; 156: 663–77.
47 Patel PH, Katweke H, Adham E, Aishbach M, Loeb LA. A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. J Biol Chem 2001; 276: 5044–51.
48 Summerer D, Rudinger NZ, Detmer I, Marx A. Enhanced fidelity in mismatch extension by DNA polymerase through directed combinatorial enzyme design. Angew Chem Int Ed Engl 2005; 44: 4712–5.
49 Lecorte AM, Patel MP, Sess LE, McInerney P, Jarroz M, et al. Directed evolution of a polymerase II for next-generation sequencing. Angew Chem Int Ed Engl 2010; 49: 5921–4.
50 Li Y, Gembhrd MC, Li Q, Taff JM, Georgiou G, et al. Engineering of TEV protease variants by yeast ER sequestration screening (YES) of combinatorial libraries. Proc Natl Acad Sci U S A 2013; 110: 7229–34.
51 Cormack BP, Valdivia RH, Faloek S. PACE-optimized mutants of the green fluorescent protein (GFP). Gene 1996; 173: 33–8.
52 Heim R, Cullibb AB, Tien RE. Improved green fluorescence. Nature 1995; 373: 663–4.
53 Fogarty PF. Biological rationale for new drugs in the bleeding disorders pipeline. Hematol Am Soc Hematol Educ Program 2011; 2011: 397–404.
54 Peters RT, Low SC, Kamphaus GD, Dumont JA, Amari J, et al. Prolonged activity of factor IX as a monoclonal Fc fusion protein. Blood 2010; 115: 2057–64.
55 Swaroop M, Moussalli M, Pipe SW, Kaufman RJ. Mutagenesis of a potential immunoglobulin-binding protein-binding site enhances secretion of coagulation factor VIII. J Biol Chem 1997; 272: 24121–4.
56 Pipe SW, Kaufman RJ. Characterization of a genetically engineered inactivation-resistant coagulation factor Villa. Proc Natl Acad Sci U S A 1997; 94: 11851–6.
57 Schiel-Bengtsdor B, Dure P. Pathway engineering and synthetic biology using acetogens. FEBS Lett 2012; 586: 2191–8.
58 Rohnin L, Ol MK, Liao JC. Microbial pathway engineering for industrial processes: evolution, combinatorial biosynthesis and rational design. Curr Opin Microbiol 2001; 4: 330–5.
59 Kirchhofer A, Helma J, Schmidtths K, Fraurer C, Cui S, et al. Modulation of protein properties in living cells using nanobodies. Nat Struct Mol Biol 2010; 17: 133–8.
60 Zhang F, Cong L, Kosuri S, Church GM, et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 2011; 29: 149–53.
61 Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. Rev Mol Biol Cell Biol 2009; 10: 866–76.
62 Cadwell RC, Joyce GF. Randomization of genes by PCR mutagenesis. PCR Methods Appl 1992; 1: 28–33.
63 Iragu R, Tarquini L, Andrea N, Moulis C, Morel S, et al. Combinatorial engineering of dextranuse specificity. PLoS One 2013; 8: e77837.
64 Hiraga K, Arnold FH. General method for sequence-independent site-directed chimeragenesis. J Mol Biol 2003; 330: 287–96.
65 Simon MD, Sato K, Weiss GA, Shokat KM. A phage display selection of engrafted homodomain mutants and the importance of residue G50. Nucleic Acids Res 2004; 32: 3623–31.
66 Hanes J, Plöckthun A. In vitro selection and evolution of functional proteins by using ribosome display. Proc Natl Acad Sci U S A 1997; 94: 4937–42.
67 Guo J, Gai T, Barbas CF 3*. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. J Mol Biol 2010; 400: 96–107.
68 Asial I, Cheng YX, Engman H, Stallhoff M, Wu B, et al. Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell. Nat Commun 2013; 4: 2901.
69 Suzuki M, Baskin D, Hood L, Loeb LA. Random mutagenesis of Thermus aquaticus thermostability using a generic activity-independent biophysical screen inside the ribosome display. Hanes J, Plückthun A.
70 Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. Rev Mol Biol Cell Biol 2009; 10: 866–76.
71 Cadwell RC, Joyce GF. Randomization of genes by PCR mutagenesis. PCR Methods Appl 1992; 1: 28–33.
72 Iragu R, Tarquini L, Andrea N, Moulis C, Morel S, et al. Combinatorial engineering of dextranuse specificity. PLoS One 2013; 8: e77837.
73 Hiraga K, Arnold FH. General method for sequence-independent site-directed chimeragenesis. J Mol Biol 2003; 330: 287–96.
74 Simon MD, Sato K, Weiss GA, Shokat KM. A phage display selection of engrafted homodomain mutants and the importance of residue G50. Nucleic Acids Res 2004; 32: 3623–31.
75 Hanes J, Plöckthun A. In vitro selection and evolution of functional proteins by using ribosome display. Proc Natl Acad Sci U S A 1997; 94: 4937–42.
76 Guo J, Gai T, Barbas CF 3*. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. J Mol Biol 2010; 400: 96–107.
77 Asial I, Cheng YX, Engman H, Stallhoff M, Wu B, et al. Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell. Nat Commun 2013; 4: 2901.
78 Suzuki M, Baskin D, Hood L, Loeb LA. Random mutagenesis of Thermus aquaticus thermostability using a generic activity-independent biophysical screen inside the ribosome display. Hanes J, Plückthun A.
Yaku M, Zhang J, Smolen GA, Muir B, Li W, et al. Transforming properties of YAP, a candidate oncogene on the chromosome 11q22 amplim. Proc Natl Acad Sci U S A 2006; 103: 12405–10.

Yagi R, Chen LF, Shigesada K, Murakami Y, Ito Y. A WW domain-containing yes-associated protein (YAP) is a novel transcriptional co-activator. EMBO J 1999; 18: 2551–62.

Haun S, Weil A, van Drogen A, Varjosalo M, Tapon N, et al. Interaction proteme of human Hippo signaling; modular control of the co-activator YAP1. Mol Syst Biol 2013; 9: 713.

Marcus GA, Horuchi J, Silverman N, Guarente L. ADA5/SPT20 links the ADA and SPT genes, which are involved in yeast transcription. Mol Cell Biol 2001; 16: 3197–205.

Gaj T, Gersbach CA, Barbas CF 3rd. Site-specific integration of human genes with programmable DNA-specificity. Nat Biotech 2013; 31: 397–405.

Kim H, Kim JS. A guide to genome engineering with programmable nucleases. Nat Rev Genet 2014; 15: 321–34.

Suzuki K, Yu C, Qu J, Li M, Yao X, et al. Synthetic reprogramming factors: direct application to human pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. Proc Natl Acad Sci U S A 2009; 106: 21617–22.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, et al. Breaking the code of DNA binding specificity of TAL-type effectors. Science 2009; 326: 1509–12.

Bartschew VV, Bieg D, Scholze H, Altmann S, Stolzenburg S, et al. Engineering specific DNA recognition into proteins: the development of designer TALEs and inhibition of epigenetic modifiers. Proc Natl Acad Sci U S A 2013; 110: 21617–22.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, et al. Breaking the code of DNA binding specificity of TAL-type effectors. Science 2009; 326: 1509–12.

Jorge‑Moreno K, Erices R, Beltran AS, Stolzenburg S, Cuello‑Fredes M, et al. Control of the forced induced fit in a bent protein-DNA complex. J Biol Chem 2011; 286: 36787–807.

Kamachi Y, Kondoh H. Sox proteins: regulators of cell fate specification and differentiation. Development 2013; 140: 4129–22.

Kamachi Y, Uchikawa M, Kondoh H. Pairing SOX eff: with partners in the regulation of embryonic development. Trends Genet 2000; 16: 182–7.

Sarkar A, Hochedlinger K. The sox family of transcription factors: versatile regulators of embryonic development. Curr Opin Cell Biol 2013; 140: 4129–22.

Kamachi Y, Uchikawa M, Kondoh H. Pairing SOX eff: with partners in the regulation of embryonic development. Trends Genet 2000; 16: 182–7.

Sarkar A, Hochedlinger K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell Stem Cell 2013; 12: 15–30.

Wilson M, Koopman P. Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. Curr Opin Dev Biol 2002; 12: 441–6.

Bernard P, Harley VR. Acquisition of Sox transcription factor specificity through protein-protein interaction, modulation of Wnt signalling and post-translational modification. Int J Biochem Cell Biol 2010; 42: 400–10.

Jauch R, Aksoy I, Hutchins AP, Ng CK, Tian XF, et al. Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. Stem Cells Dev 2011; 20: 940–51.

Niakan KK, Ji H, Maehr R, Vokes SA, Rodolfa KT, et al. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extranephrine gene expression and indirectly antagonizing self-renewal. Genes Dev 2010; 24: 312–26.

Aksoy I, Jauch R, Chen J, Dyla M, Divakar U, et al. Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. EMBO J 2013; 32: 938–53.

Esch D, Vahokoski J, Groves MR, Pogenberg V, Cojocaru V, et al. A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat Cell Biol 2013; 15: 295–301.

Metino F, Ng CK, Veerapandian V, Schöler HR, Jauch R, et al. Structural basis for the SOX-dependent genomic redistribution of OCT4 in stem cell differentiation. Structure 2014; 22: 1274–86.

Buganin Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchical phase. Cell 2012; 150: 1209–22.

Buganin Y, Faddah DA, Jaensch R. Mechanisms and models of somatic cell reprogramming. Nat Rev Genet 2013; 14: 427–39.

İgonçalves MA, Janssen JM, Nguyen QG, Athanasopoulos T, Hauschka SD, et al. Transcription factor rational design improves directed differentiation of human mesenchymal stem cells into skeletal myocytes. Mol Ther 2011; 19: 1331–41.