Group Meeting_Fangjie_1 (2014/09 to 2015/03)

Background

The basic packaging unit of chromatin is nucleosome. TF binding, and thus the activation of transcription, is largely affected by nucleosome occupancy and its turnover rate. While many epigenetic factors, like histone modification and DNA methylation, are known to regulate nucleosome binding, *in vitro* chromatin reconstitution and *in vivo* mapping also suggested that nucleosome occupancy is partially determined by the DNA sequence itself.

Nonetheless, due to the limited complexity of genome sequences used in reconstitution, the mapping or chromatin reconstitution approaches were not capable to capture a long, affinity-based binding model applicable to all species. In addition, only 1 round of assembly selection cannot suffice to tell the affinity difference between high-affinity sites. As such, the current project aims to capture a more precise model for nucleosome binding sequences by using long random ligands in SELEX.

The binding specificity of nucleosome to DNA is lower than that of a typical TF. Unlike DNA-binding proteins that achieve specificity by virtue of direct and strong interactions between a few base pairs and amino acids, the specificity of nucleosome formation largely reflects the overall ability of a specific 147 bp DNA sequence to bend sharply around the histone octamer. Optimal nucleosome formation occurs when bendable dinucleotides (AT and TA) occur on the face of the helical repeat (every 10 bp) that directly interacts with histones **(Fig. 1)**. The homopolymeric sequences poly(dA:dT) and poly (dG:dC) are intrinsically stiff. Their presence tends to deplete nucleosomes in the vicinity. The 10 bp periodicity of *in vitro* reconstituted nucleosomes is more distinct than that of the nucleosomes formed *in vivo*. Hopefully after several rounds of SELEX, we can observe an even stronger periodicity in the enriched ligands.

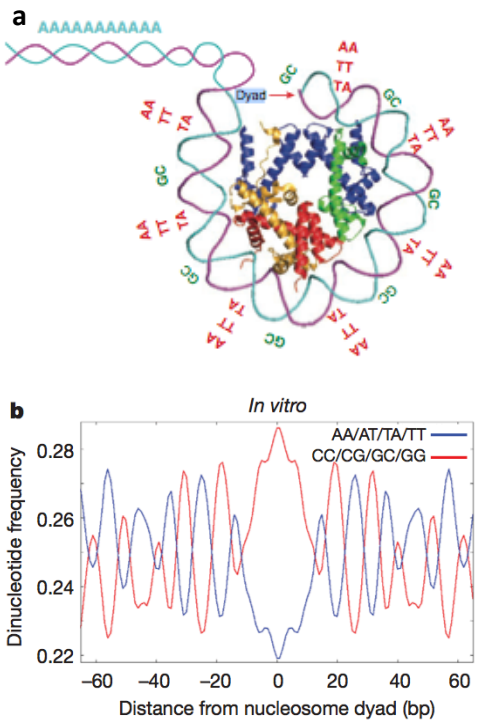
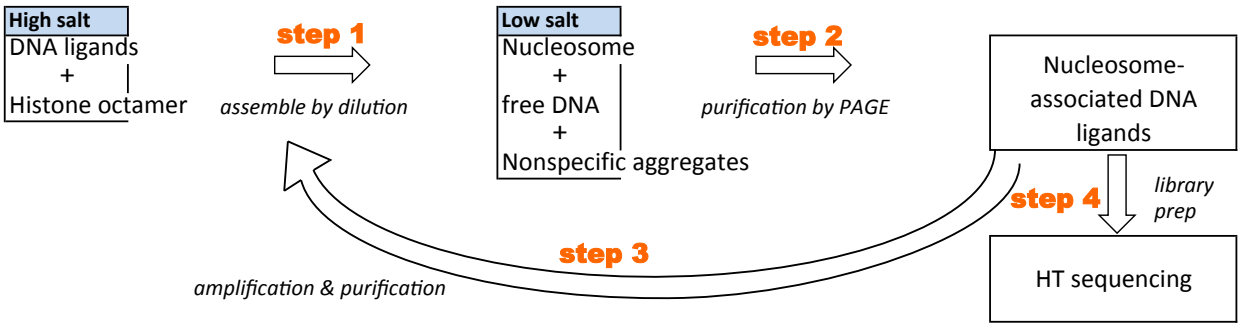


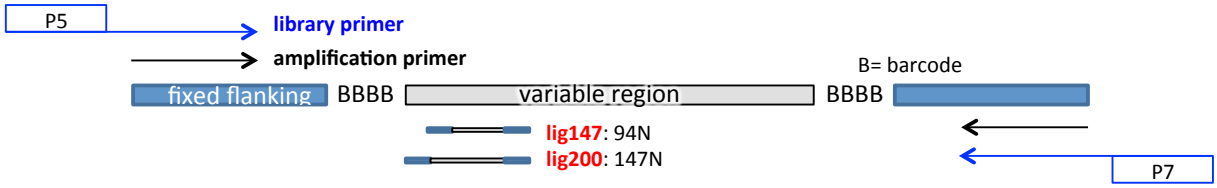
Figure 1 *Nucleosome sequence preferences. A preference for distinctive dinucleotides recurs periodically at the DNA helical repeat and is known to facilitate the sharp bending of DNA around the nucleosome. The linker regions exhibit a strong preference for sequences that resist DNA bending and thus disfavor nucleosome formation.*

EMSA SELEX for nucleosome binding ligands (with Bei's help)

*** Flowchart of EMSA-SELEX**



*** DNA ligand design (based on Arttu's design)**



*** Optimization: Eliminating the single-stranded (for synthesis of initial DNA ligands and step 3)**

_ Single-stranded DNAs may show similar shifts as assembled nucleosomes in EMSA SELEX. Presence of them will result in a decrease in S/N ratio (not a problem for HT SELEX), especially when the histone octamer concentration is low (high selectivity conditions).

_ If nucleosome binds ssDNA, the presence of ssDNA may generate erroneous binding signal.

_ Because of the high complexity of SELEX library, ssDNA is easily generated in PCR reactions.

_ For the synthesis of initial DNA ligands, Bei's protocol did not work well for my oligos (**Fig. 2a**). Adjustment of annealing temperature, primer concentration, final extension time, rounds of amplification failed to alleviate ssDNA contamination. Occasionally I noticed that the amplification efficiency of the PCR reaction was low (DNA amount only doubled after a total 4 rounds of PCR), and that my long oligos were intrinsically less pure than previous SELEX ligands. Then after the optimization of template amount (Bernie's suggestion), PCR rounds, dNTP concentration, and Mg2+ concentration, relatively clean dsDNA ligands were obtained **(Fig. 2b)**.

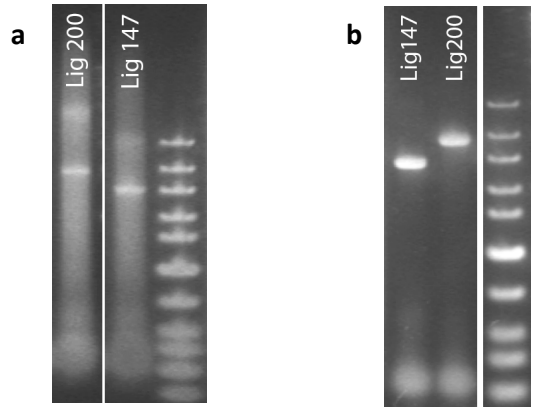


Figure 2 *Optimization of dsDNA ligand synthesis. (a),before optimization, smears contain ssDNA and contamination of shorter fragments; (b), after optimization*

_ For the amplification of ligands between rounds (step 3), because it is not practical timewise to adjust the template concentration or PCR cycles for each reaction, generally the PCR amplification is carried on until saturation. However, even after a final round, we may still see some additional bands **(Fig.3,** left**)**, which are less well-defined than the product bands but are also not a complete smear. qPCR results indicate that all these bands are actually the ligands (**Fig. 3**, middle). Probably they are partially single-stranded ligand pairs with the randomized regions forming bubbles. Some of these additional bands shift similarly as the assembled nucleosome. It could be preferable to reduce their amount. I found that extending the final extension time of the final round could help (**Fig.3**, right). (Recently I tested Jian's digestion method with S1 nuclease, it worked very well to clean up the product. Since Exonuclease I cannot digest such additional bands, these bands should really be the "bubble" structures).

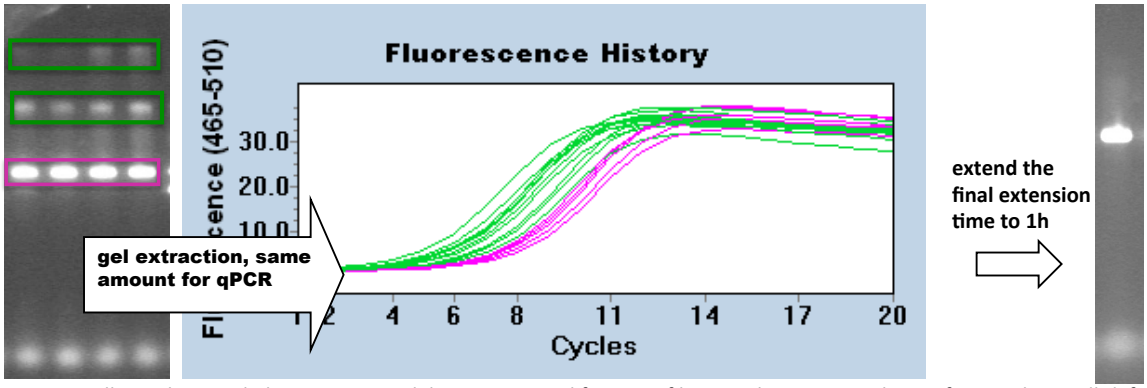


Figure 3 *Partially single stranded DNA generated during PCR amplification of lig200. They migrate slower, forming less well-defined bands.*

_ Step 4 generates even more partially single-stranded DNA than step 3, possibly because the common flanking segment of seq library is longer than the SELEX ligand. Thus the bubble structure between partially mismatched pairs could be more stable. I did not tried to optimize this step because ssDNA will not do harm to sequencing, and that a gel purification step is included in the current workflow to prepare seq library.

*** Optimization: nucleosome reconstitution (Step 1)**

_ Nucleosomes were successfully reconstituted in the first place (**Fig. 4**)

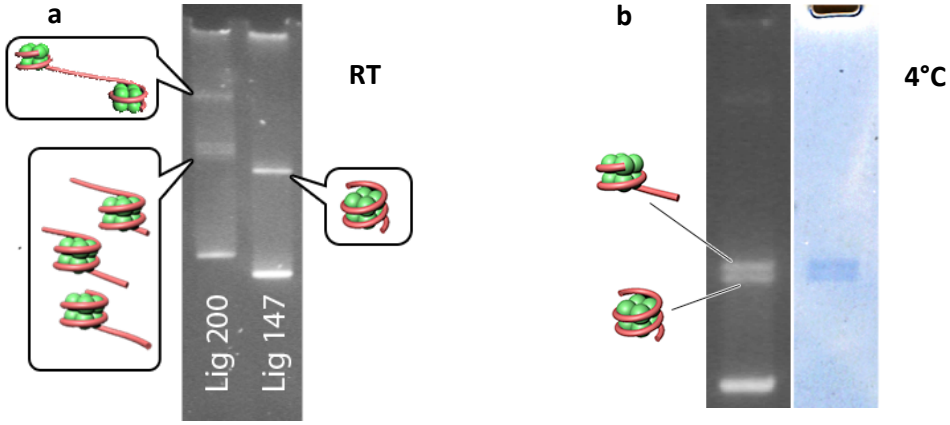


Figure 4 *The nucleosomes reconstituted with lig200 and lig147 by dilution. a), reconstituted under room temperature; b) reconstituted at 4°C, the right lane showing Coomassie blue staining of the same gel*

_ The reconstitution of nucleosome seems delicate, for a long period, I could not repeat the initial result but instead obtained a double band or a smear (**Fig. 5,** middle). I lowered the voltage of electrophoresis, improved the cooling during EMSA, and operated more carefully during the reconstitution but still could not get rid of such smear. I started to doubt if the histone octamer deteriorated during storage. If this is true, there could be little point to go forth with the SELEX because denatured or decomposed protein probably would not give the correct signal. However, when I compared octamers stored under -20°C, -80°C, and 4°C, they make no difference in reconstitution. Finally and fortunately, I found that it is not the protein but the electrophoresis conditions that is problematic. Longer immersion of nucleosome particles in even 0.2x TBE will decompose gradually during the electrophoresis. A higher voltage but shorter time of electrophoresis is able to reproduce the initial result more stably (**Fig. 5**, right). In addition to the electrophoresis condition, the operation details in reconstitution do affect the yield.

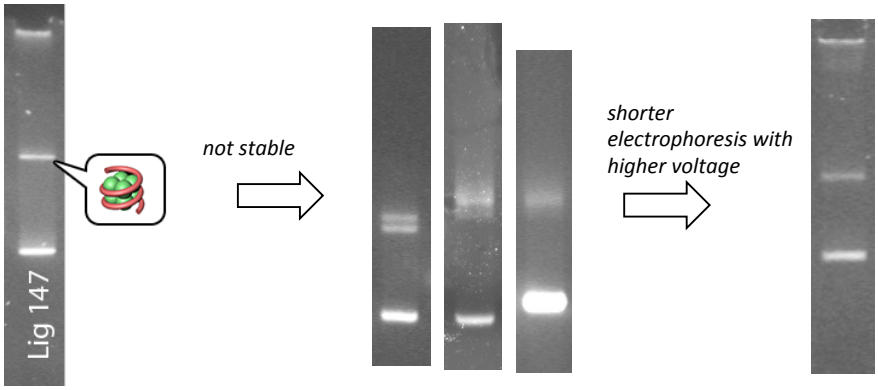


Figure 5 *The nucleosome reconstitution easily yields smears. A shorter time for electrophoresis is important.*

*** 4 cycles of SELEX finished with 24 barcodes (12 of lig147, 12 of lig200), libraries were sent for sequencing**

Lig 147		Lig200	
no Mg2+, 100 mM final KCl conc.	2.5 mM Mg2+, 150 mM final KCl conc.	no Mg2+, 100 mM final KCl conc.	2.5 mM Mg2+, 150 mM final KCl conc.
6 different octamer/lig147 ratio	6 different octamer/lig147 ratio	6 different octamer/lig200 ratio	6 different octamer/lig200 ratio

*** Problems remaining**

_ Although I switched to staining the gel using a spray, cross-contamination still occurs between wells, especially when lig147 is contaminated by lig200 (**Fig.6**, red arrows). The lig200 will enrich after several rounds of SELEX because the partially ssDNA band of lig200 shifts the same as the lig147-nucleosome. However, since we can separate different lanes by barcode, and the ssDNA formation is less likely to show sequence preference, maybe such cross-contamination will only add a bit to the background and should not be a big issue.

_ I got some smears during the amplification of nucleosome-bound DNA ligands (cut and extracted from the PAGE gel) in the later rounds (**Fig.6**). As the smears are more sever for samples with more diluted DNA, they seems to come from the primer assembly. (Probably I used too long time for the final extension in order to get rid of ssDNA)

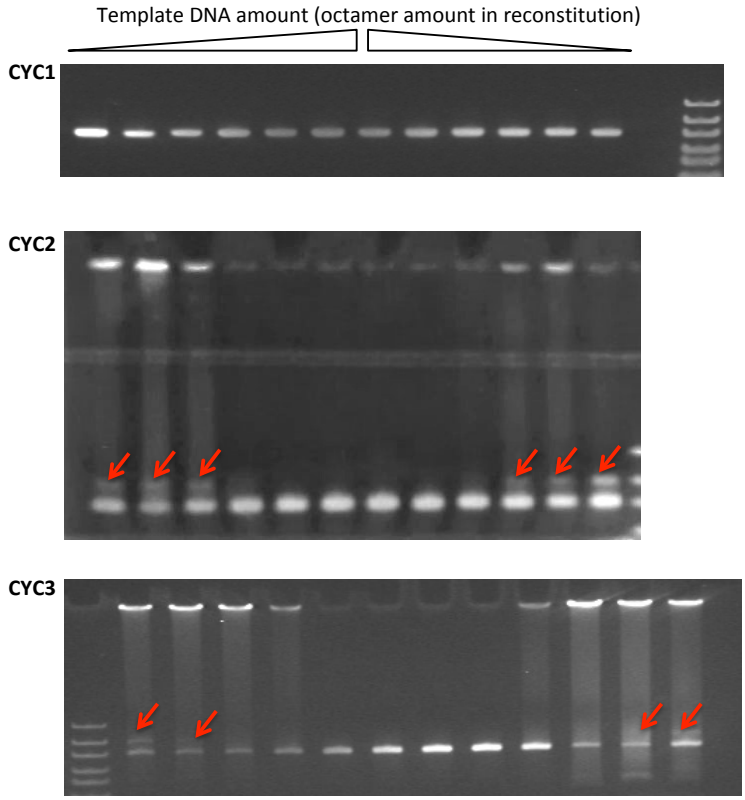
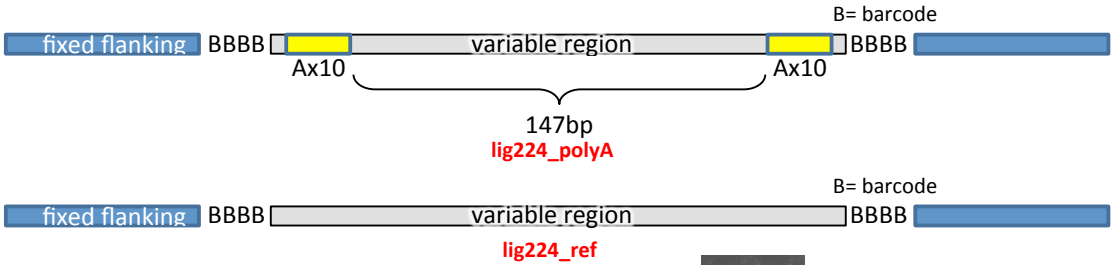


Figure 6 *Amplification results of the nucleosome-bound lig147 (used as ligand pools for the next round). Cross-contamination from lig200 are indicated with red arrows. Smears are seen in the amplification products for which the template concentrations are low.*

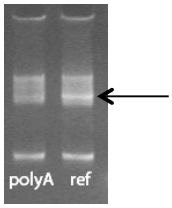
*** Effects of polyA tract on nucleosome positioning**

_ We hope to confine the nucleosome onto a 147bp stretch which is completely random. This could facilitate the alignment of sequences during data analysis.

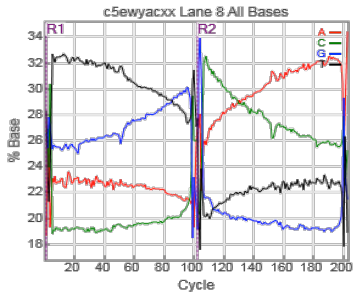
_ One of Jussi's idea is to use the polyA tract, which disfavors nucleosome occupation, to restrict the nucleosome's position. The design of ligands are as follows:



_ Nucleosome reconstitution using these ligands yields



Sequencing of cyc0 ligand library (20 barcodes)

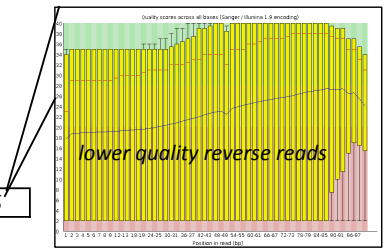


_ The sequencing results looks good, probably we can load a bit more
_ Synthesized oligos have a biased base composition for "N" position

---

_ Processing of the obtained PE-reads

100% all PE reads passed filter (116.15 million)



merging 2 pair-end reads into 1 through overlap

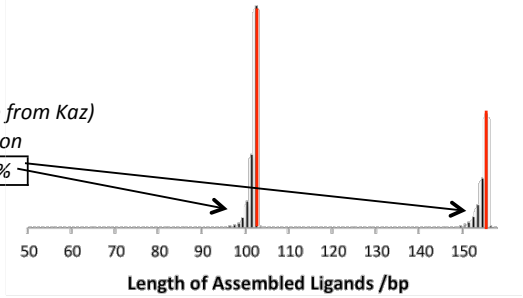
loss: 4.9%

95,10% merged long reads

(with help from Kaz)

length filtering and barcode separation

loss: 41.2%



55,92% reads files for each barcode
(3,2 M reads for 1 barcode on avg)

_ Complexity of cyc0 ligand library

In order to dilute out impurities, the synthesized DNA oligos were amplified 8 rounds before entering SELEX cycle. No significant complexity decrease can be observed according to the current sequencing depth.

Ligand	[a] Complexity	Ligand	Complexity	Ligand	Complexity	Ligand	Complexity
lig147-01	98,61%	lig147-06	98,50%	lig200-01	98,71%	lig200-06	98,57%
lig147-02	98,74%	lig147-07	98,59%	lig200-02	98,71%	lig200-07	98,66%
lig147-03	98,59%	lig147-08	98,67%	lig200-03	98,67%	lig200-08	98,67%
lig147-04	98,69%	lig147-09	98,66%	lig200-04	98,66%	lig200-09	98,67%
lig147-05	98,62%	lig147-10	98,70%	lig200-05	98,68%	lig200-10	98,75%

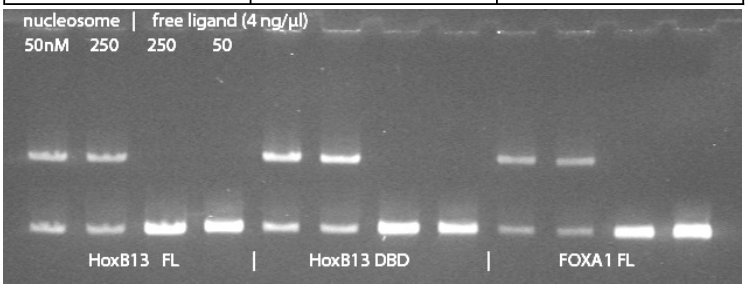
a. percentage of the remaining reads after deduplication is used here to indicate complexity

EMSA of nucleosome with TFs

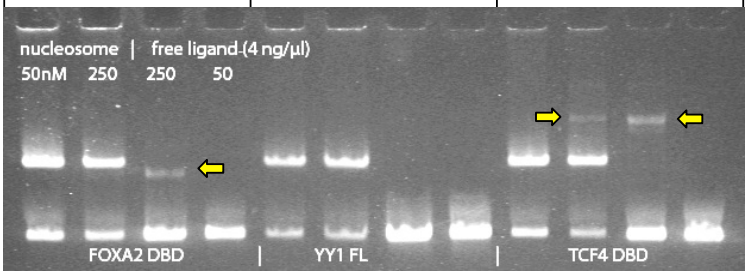
_ The binding of TFs are suggested to be in a competition with histones. Especially pioneer TFs are highlighted for their ability to bind nucleosomal DNA.

_ 6 TFs (by courtesy of Jian) were added into the assembled nucleosome (containing free ligands) to check their ability to bind nucleosomal DNA, the experiment design and results are as follows.

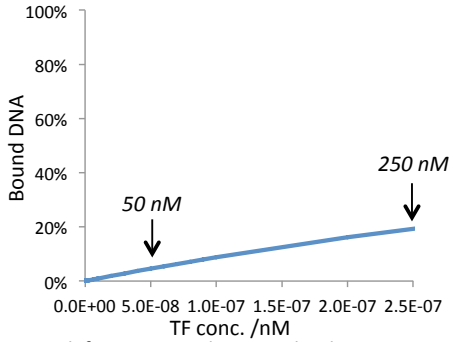
	HOXB13				HOXB13 DBD				**FOXA1 (pioneer)**			
TF conc /nM	50	250	250	50	50	250	250	50	50	250	250	50
nucleosome (with free lig147) /nM		50				50				50		
free lig147 /nM			50				50				50	



	FOXA2 DBD (pioneer)				YY1				TCF4 DBD			
TF conc /nM	50	250	250	50	50	250	250	50	50	250	250	50
nucleosome (with free lig147) /nM		50				50				50		
free lig147 /nM			50				50				50	



_ Shifts of free lig147 but not of nucleosome according to TF binding are only identified for FOXA2 and TCF4 (yellow arrows). The non-specific interaction between TF and DNA are generally with a Kd >= 1 µM, this predicts well for the observed absence of binding or weak bindings (**Fig.6**, left). However, for FOXA1, a study by EMSA (T. Sekiya, U. M. Muthurajan, K. Luger, A. V. Tulin, K. S. Zaret, Gene Dev 2009, 23) reported an apparent Kd of 1.8 nM for the binding of FOXA1 to the mononucleosome consisting of alpha-satellite DNA (no specific FOXA1 site), while no binding was detectable in my trial. I have not yet figured out what caused such difference. The differences between my binding buffer and theirs (**Fig.6**, right) do not seem to account for such a big mismatch in Kd.



my binding buffer	**binding buffer reported**
1x TE pH7.5	10mM Tris-HCl pH7,5
1 mM DTT	1mM DTT
c.a. 80 mM KCl	60 mM KCl
c.a. 1 mM imidazol	1mM MgCl2
c.a. 1% glycerol	3mg/mL BSA
	1% glycerol
	1% Ficoll

Incubation time are both 2h at RT

Figure 6 *left, anticipated TF-DNA binding percentage with a DNA ligand concentration of 50nM and Kd=1µM. Right, comparison of the binding buffers used in this study and reported before*

Future plan

*** Data analysis to identify the sequence preference of nucleosome**

*** Nucleosome assembly with histone chaperone and remodeling factors**

_ With histone chaperone (or simply polyanionic electrolytes), assembly can be completed under physiological salt concentrations and no dilution will be needed, probably more compatible with high-throughtputlization

*** Automation and high-throughputlization of nucleosome reconstitution and SELEX**

_ Because nucleosome reconstitution by dilution, EMSA, and gel cutting operations are laborious and low-throughput, and as well, EMSA is prone to cross-contaminating between wells, we hope to first set up a high throughput procedure to perform nucleosome reconstitution with lig200 and SELEX.

_ Basically we can use a good monoclonal antibody attached to beads to pull down the nucleosomes, but probably better if we can use tagged histones.

_Then we can perform CAP (consecutive affinity purification) SELEX with our TF libraries which is already tagged.

*** Relation of nucleosome positioning to DNA methylation**

_ Evidences *in vivo* show that nucleosome positioning covaries with the extent of DNA methylation. To test the effect of DNA methylation on the sequence preference of nucleosome, with the help from Yimeng, we have prepared ligands compatible with bisulfide conversion

*** Compare the *in vitro* binding preference of nucleosome with *in vivo* preference**

*** Compare the *in vitro* TF binding preference on nucleosome with *in vivo* TF preference**

*** More data analysis work if possible**