i. ChIP-seq occupancy mapping of the archaeal transcription machinery

Fabian Blombach*, Kathy L. Smollett and Finn Werner*

Institute of Structural and Molecular Biology, Division of Biosciences, University College London, London WC1E 6BT, United Kingdom

* Corresponding authors f.blombach@ucl.ac.uk, or f.werner@ucl.ac.uk

Running title: “ChIP-seq of the archaeal transcription machinery”

ii. Abstract

Genome-wide occupancy studies for RNA polymerases and their basal transcription factors deliver information about transcription dynamics and the recruitment of transcription elongation and termination factors in eukaryotes and prokaryotes. The primary method to determine genome-wide occupancies is chromatin immuno-precipitation combined with deep sequencing (ChIP-seq). Archaea possess a transcription machinery that is evolutionarily closer related to its eukaryotic part but it operates in a prokaryotic cellular context. Studies on archaeal transcription brought insight into the evolution of transcription machineries and the universality of transcription mechanisms. Because of the limited resolution of ChIP-seq, the close spacing of promoters and transcription units found in archaeal genomes pose a challenge for ChIP-seq and the ensuing data analysis. The extreme growth temperature of many established archaeal model organisms necessitates further adaptations. This chapter describes a version of ChIP-seq adapted for the basal transcription machinery of thermophilic archaea and some modifications to the data analysis.

iii. **Keywords:** ChIP-seq, chromatin immuno-precipitation, ChIP-exo, Archaea, *Saccharolobus*, *Sulfolobus*, transcription, RNA polymerase, TFB, TFE

1. Introduction

Chromatin immuno-precipitation combined with deep sequencing (ChIP-seq) is a widely used method to study the genome-wide distribution of transcription factors and chromatin proteins. The method relies on the fixation of DNA-protein interactions by formaldehyde cross-linking. DNA is then fragmented by mechanical shearing through sonication or enzymatic digestion and the DNA-protein complexes of interest are enriched through binding on immobilised antibodies against the target protein.

Formaldehyde is the standard cross-linker used in ChIP-seq because of its short cross-linking distance of about 2 Å, rapid permeability of cells, and fast reactivity. These properties allow formaldehyde cross-linking to preserve physiologically relevant DNA-protein and protein-protein interactions [1]. Standard cross-linking protocols recommend temperatures of 25 °C to at 37 °C combined with incubation times of several minutes although prolonged incubation times favour the formation of unspecific DNA-protein cross-links [2]. The cross-linking protocol poses a particular challenge for the adaptation of ChIP-seq to thermophilic prokaryotes. For the basal transcription machinery of thermophiles to be caught in the act of transcription, such long incubations with the cross-linker far from physiologically relevant temperatures appear to be unsuitable. Alternative cross-linking protocols with shorter incubation times at the relevant growth temperature have been therefore employed for ChIP-seq in the euryarchaeal species *Pyrococcus furiosus* and *Methanocaldococcus jannaschii* [3,4]. We further developed ChIP-seq for thermophilic archaea by introducing an indirect immuno-precipitation method into the protocol. Indirect immuno-precipitation allows free antibodies to bind to their target proteins before the antibodies are immobilised on magnetic beads in

contrast to the direct approach where the antibodies are immobilised first. The indirect method dramatically reduces the background of genomic DNA in the immuno-precipitated samples in our hands. The method makes it thereby possible to obtain a good signal to noise ratio even when the DNA recovery is reduced under these modified, more specific cross-linking conditions.

Several other protocols for ChIP-seq in archaea have been published [5-7]. Our protocol described here is designed to work across different target proteins such as RNA polymerase, transcription initiation, elongation and termination factors as well as chromatin proteins with high spatial resolution. We have successfully applied this protocol to three Sulfolobales species (*Sulfolobus islandicus, Sulfolobus acidocaldarius,* and *Saccharolobus solfataricus*) and we have also adapted this protocol successfully for the euryarchaeon *Methanocaldococcus jannaschii* ([8] and unpublished data). The basic ChIP protocol is also suitable for more advanced ChIP-based methods such as ChIP-exo in order to further increase the spatial resolution of the data [8].

2. Materials

2.1. Antibody purification

1. 2 mL Rabbit antiserum against protein of interest and 0.5 mL pre-immune serum
2. 0.2 µM syringe filter and 2 mL syringe
3. Disposable 2 mL gravity-flow Columns
4. Protein G agarose resin, 50% suspension
5. PD-10 desalting columns (Cytiva)
6. Amicon Ultra 0.5 mL 100K ultrafiltration device (Merck)
7. 10x PBS buffer: 1.37 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 18 mM KH_2PO_4, pH 7.4

8. Glycine elution buffer: 0.1 M glycine, pH 2.7
9. Neutralisation buffer: 1 M Tris-HCl pH 8.0
10. 1x PBS/glycerol: 1x PBS buffer supplemented with 10% glycerol
11. Nanodrop spectrophotometer (ThermoFisher)

2.2. Cell cross-linking

1. Brock medium supplemented with 0.2% glucose and 0.1% tryptone [9]
2. *Saccharolobus solfataricus* P2 stock (formerly *Sulfolobus* genus) [10,11]
3. Magnetic Hotplate Stirrer and large stir bar.
4. 36.5-38% Formaldehyde solution containing 10-15% methanol as stabiliser (Merck).
5. 1 M Tris-HCl pH 8.0
6. 1x PBS buffer: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4 (or dilute 10x PBS buffer, *see* Subheading 2.1, **step 7**)
7. Cooling centrifuge suitable for 50 mL tubes.

2.3. DNA shearing

1. cOmplete™ Protease Inhibitor Cocktail (Merck).
2. Lysis buffer: 50 mM HEPES-NaOH pH 7.5, 140 mM NaCl, 1 mM Na_2 EDTA, 0.1% Sodium Deoxycholate, 1% Triton-X-100.
3. Q700 sonicator with cup horn and recirculating chiller (Qsonica) plus the 1.5 mL sonication tube rack for the cup horn
4. 1.5 mL Polystyrene sonication tubes (ActiveMotif) (*see* **Note 1**)
5. Cooling microcentrifuge.
6. 1.5 mL DNA LoBind microcentrifuge tubes (Eppendorf)

7. ChIP elution buffer: 50mM Tris-HCl pH 8.0, 10 mM Na_2 EDTA, 1% Sodium dodecyl sulfate
8. Heat block at 65 ^oC
9. 20 mg/mL Proteinase K
10. 10 mg/mL RNase A
11. QIAquick PCR purification kit
12. 50% glycerol prepared with nuclease-free water

2.4. Chromatin immuno-precipitation

1. Lysis buffer: 50 mM HEPES-NaOH pH 7.5, 140 mM NaCl, 1 mM Na_2 EDTA, 0.1% Sodium Deoxycholate, 1% Triton-X-100 (same buffer as Subheading 2.3., **item 2**)
2. Lysis buffer 500: 50 mM HEPES-NaOH pH7.5, 500 mM NaCl, 1 mM Na_2 EDTA, 0.1% Sodium Deoxycholate, 1% Triton-X-100
3. LiCl wash buffer: 10 mM Tris-HCl pH 8.0, 100 mM LiCl, 1 mM Na_2 EDTA, 0.5% Sodium Deoxycholate, 0.5% Nonidet P-40
4. TE buffer: 10 mM Tris-HCl pH 8.0, 1 mM Na_2 EDTA
5. ChIP elution buffer: 50mM Tris-HCl pH 8.0, 10 mM Na_2 EDTA, 1% Sodium dodecyl sulfate
6. Dynabeads™ M-280 Sheep Anti-Rabbit IgG (ThermoFisher) (*see* **Note 2**)
7. Magnetic rack suitable for 2 mL microcentrifuge tubes, e.g. MagnaRack (ThermoFisher)
8. 1.5 mL and 2 mL DNA LoBind microcentrifuge tubes (Eppendorf)
9. Rotator for microcentrifuge tubes at 4 ^oC (by placing the apparatus in a cold room or fridge)
10. Heat block at 65 ^oC
11. 20 mg/mL Proteinase K
12. 10 mg/mL RNase A

13. QIAquick PCR purification kit (Qiagen)
14. Qubit™ dsDNA HS Assay Kit (ThermoFisher)
15. Qubit™ Assay Tubes (ThermoFisher)
16. Qubit Fluorometer (ThermoFisher)

3. Methods

3.1. Antibody purification

Purification of antibodies from antisera is required to accurately assess the amount of antibody used in the ChIP experiments. In addition, the purification step will remove components that potentially interfere with the ChIP experiment. The purification can be carried out on the bench at room temperature, but we advise to cool all the buffers and avoid keeping the antisera and purified antibodies at room temperature for any longer time.

1. Place all buffers on ice.
2. Prepare disposable 2 mL gravity-flow columns by placing frits into the columns (as per manufacturer's instructions). for the antiserum and pre-immune serum as control. For multiple antisera, scale up accordingly.
3. Pipette 0.4 mL of the 50% Protein G agarose suspension into the 2 mL columns (0.2 mL bed volume).
4. Equilibrate the columns with 10 column volumes (CV) of buffer A.
5. Thaw 2 mL of the antiserum and 0.5 mL pre-immune serum on ice (*see* **Note 3**)
6. Add 2 mL 1x PBS buffer to the antiserum.
7. Filter the antiserum/PBS mix through a 0.2 μm pore size syringe filter onto the column.
8. Once the antiserum/PBS mix has passed the column, wash the column with 8 mL of 1x PBS buffer.

9. For each column, prepare a 1.5 mL microcentrifuge tubes with 150 µL Neutralisation buffer.
10. Elute the antibodies from each column with 750 µL glycine elution buffer into the prepared microcentrifuge tubes with Neutralisation buffer and place tubes on ice.
11. Equilibrate two PD-10 columns with 25 mL 1x PBS/glycerol
12. Load the eluate on the PD-10 column (0.9 mL volume) and let the sample enter the column.
13. Add 1.6 mL 1x PBS/glycerol and let the buffer enter column.
14. For each column, prepare a tube rack with 8 microcentrifuge tubes.
15. Elute with 3.5 mL PBS/glycerol while collecting ~0.5 mL fractions into the microcentrifuge tubes (approximately 10 drops). Perform the elution step separately for each antibody!
16. Store the elution fractions on ice.
17. Analyse all fractions on a Nanodrop spectrophotometer by measuring A280nm absorption. Use the IgG molar extinction coefficient for conversion into protein concentration estimates.
18. If the estimated protein concentration is below 0.5 mg/mL, combine the peak fractions and concentrate the protein using a 100 kDa cut-off ultrafiltration device. This is likely to be the case for the material purified from pre-immune serum.
19. Aliquot the protein into multiple microcentrifuge tubes and store at -20 ^{o}C
20. Assess the purity of the purified antibodies by SDS-PAGE using sample buffer with reducing agent and Coomassie staining. The SDS gel should show predominantly two bands at 25 kDa and 50 kDa corresponding to light and heavy chain, respectively.

3.2. Cross-linking of cells

1. Grow *Saccharolobus solfataricus* str. P2 in 1 litre Brock medium supplemented with 0.2% glucose and 0.1% tryptone (*see* **Chapter 8**) in a 2 litre Erlenmeyer flask at 75 ^{o}C to late exponential growth phase, OD_{600nm} 0.25 to 0.4 (*see* **Note 4**).
2. Prepare the work environment for the cross-linking reaction inside a fume hood. Have all materials required in the immediate vicinity (*see* **Note 5**)
3. Preheat the magnetic stirrer.
4. Transfer flask on a heated magnetic stirrer in fume hood (*see* **Note 5**)
5. Add rapidly 11.1 mL of the stabilized formaldehyde solution for a final concentration of 0.4%. Let the cross-linking reaction proceed for exactly 1 min.
6. Stop the cross-linking reaction by the rapid addition of 100 mL 1 M Tris-HCl pH 8.0 (100 mM final concentration) (*see* **Note 6**)
7. Transfer the flask immediately onto a bucket with ice water for fast cooling.
8. Collect the cells by centrifugation for 20 min at 5,000 x *g*, 4 ^{o}C in a cooling centrifuge
9. Carefully remove the supernatant (*see* **Note 7**)
10. Resuspend the cell pellet in 12 mL 1x PBS buffer
11. Transfer the resuspended cells to a 50 mL tube and collect the cells by centrifugation for 10 min at 4,000 x *g*, 4 ^{o}C in a cooling centrifuge.
12. Decant supernatant.
13. Repeat steps 10 to 12 two more times for a total of 3 washes.
14. Freeze the cell pellet in liquid N_2 and store at -80 ^{o}C.

3.3. DNA shearing by sonication

1. Switch on the recirculating chiller for the sonicaotr and cool the cup sonicator to 4 °C.
2. Take cross-linked cell material from the -80 °C freezer and thaw on ice.
3. Resuspend in 6 mL ChIP lysis buffer + Protease inhibitor complete (ROCHE).
4. Aliquot 8x 500 µL into 1.5 mL polystyrene sonication tubes (Active Motif).
5. Store remaining material at -80 °C (*see* **Note 8**).
6. Mount the rack. Take care not to use to much force when tightening the stem.
7. Adjust water level in cup horn so that it matches the filling level in the tubes.
8. Sonicate for a total sonication time of 40 min at Amplitude 16 with 10s pulses and 20s pause intervals, 120 min total time.
9. Transfer the lysate into 1.5 mL microcentrifuge tubes.
10. Centrifuge the lysate in a cooling microcentrifuge at 20,000 x *g* for 15 min at 4 °C.
11. Retrieve supernatant carefully leaving about 50 µL remaining at the bottom of the tube. We refer to this sample as chromatin input from here on.
12. Remove 25 µL of the chromatin input into a 1.5 mL LoBind microcentrifuge tube to test shearing efficiency
13. Freeze the remaining chromatin input in liquid N_2 and store at -80 °C
14. Mix the 25 µL aliquot of chromatin input with 200 µL ChIP elution buffer in a 1.5 mL LoBind microcentrifuge tube.
15. Add 2.5µL 20 mg/mL Proteinase K and 1µL 10 mg/mL RNase A.
16. Incubate at 65 °C overnight.
17. Clean up the DNA using the QIAquick PCR purification kit. Add 1,000 µL PB buffer (part of the kit) to the sample and follow the routine protocol. Elute the DNA with 50 µL EB buffer (part of the kit).

18. Quantify the DNA concentration on a Nanodrop spectrophotometer. A DNA concentration are around 100 ng/µL. Because of the sample to elution volume ratio, the effective concentration in the sheared chromatin sample is two-fold higher.
19. Into a new microcentrifuge tube, pipette 8 µL of sheared DNA and 2 µL 50% glycerol on 2% agarose gel alongside a DNA size standard covering the 100 to 500 bp region to check the fragment size range. Stain the gel with ethidium bromide after electrophoresis. The sheared DNA should show a relatively narrow distribution with maximum intensity at ~150 nt (*see* **Fig. 1a**) (*see* **Note** 9).

3.4. Immuno-precipitation

1. Thaw an aliquot of sheared chromatin on ice.
2. Prepare four 2 mL LoBind with 10 µg of sheared chromatin each and dilute with Lysis buffer to 20 ng/µL final concentration, i.e., 500 µL volume.
3. To optimize the immuno-precipitation conditions, set up four different samples with 1 to 4 µg Protein G-purified antibodies (Subheading 3.1). Mix purified antibodies with lysate in 2 mL DNA LoBind tube (*see* **Note 10**).
4. Place the tubes into a rotator at 4 ^{o}C for overnight incubation.
5. The following day, add 50 µL resuspended Dynabeads to a new 2 mL LoBind tube for each ChIP experiment and place the tubes on the magnetic rack.
6. Once the beads are collected, remove the supernatant.
7. Resuspend the beads in 500 µL 1x PBS buffer and place the tube back into the rack.
8. Once the beads are collected, remove the supernatant and repeat the washing step twice with 500 µL ChIP lysis buffer.
9. Once the beads are collected and transfer the lysate/antibody mix to the beads.

10. Place the tubes into a rotator at 4 °C for 1 hr (*see* **Note 11**).
11. Transfer the tubes onto the magnetic rack to separate
12. Remove and discard the supernatant including residual fluid trapped in the lid of the tubes.
13. Resuspend the beads in 1,000 µL lysis buffer and place into the rotator for 15 min at 4 °C
14. Repeat the lysis buffer washing once (**steps 11 to 13**)
15. Wash the beads likewise with 1,000 µL lysis buffer 500 for 15 min at 4 °C
16. Wash the beads likewise with1,000 µL LiCl buffer for 15 min at 4 °C
17. Resuspend beads with 1 mL TE buffer and transfer to new 2 mL Lobind tube. Wash 1x 15 min 4 °C (*see* **Note 12**).
18. Resuspend the beads in 200 µl ChIP-EB buffer and incubate at 65 °C for 10 min in a heat block.
19. Allow tubes to cool to room temp (~5 min) and place them on the magnetic rack.
20. Transfer the supernatant into a 1.5 mL LoBind microcentrifuge tube (*see* **Note 13**).
21. Add 2 µL 10 mg/mL RNase A and 5 µL 20 mg/mL proteinase K.
22. Incubate the samples at 65 °C overnight.
23. Clean up the DNA using the QIAquick PCR purification kit. Add 1,000 µL PB buffer (part of the kit) to the sample and follow the routine protocol. Elute the DNA with 50 µL EB (part of the kit).
24. Quantify 5 µL IP samples and 1 µL of the chromatin input using the Qubit dsDNA HS assay. The expected concentration range is up to 1,000 ng/mL for more abundant targets such as RNA polymerases. However, some less efficient ChIP experiments might be close to the detection limit of 20 ng/mL. The pre-immune serum control sample should be generally below the detection limit. If possible, qPCR experiments (*see* Subheading 3.5) show

greater sensitivity and specificity and are therefore useful to verify the specificity of low efficiency ChIP experiments compared to pre-immune serum controls and chromatin input.

3.5. qPCR analysis and protocol optimisation

If a binding site of the target protein has been previously identified, it is useful to optimize the ChIP experiments conditions. For general transcription initiation factors, transcription start sites (TSS) of genes/operons known to be highly expressed form a good starting point. To this end, qPCR quantification over two different target regions, the binding site and a nearby control region in a distance of at least two-fold average fragment size of the sheared chromatin (~300 bp). For initiation factors such as TFB and the two TFE subunits TFEα and TFEβ, higher enrichment at the promoter is a good indicator for the optimisation of ChIP experiments (*see* **Fig. 1c**). Due to differences in qPCR instruments and compatible PCR plastics and qPCR master mixes, we only provide only guidelines for the qPCR evaluation.

1. Design primer pairs amplicon length of 100 to 120 bp using suitable software such as https://primer3.ut.ee.
2. Use the standard curve method for qPCR quantification. For convenience, DNA purified from the chromatin input (Subheading 3.3, **step 17**) can be used to generate the standard curve.
3. Dilute 5 µL of the ChIP elution samples ten-fold in nuclease-free water.
4. Dilute 5 µL of the chromatin input 1,000-fold in nuclease-free water.
5. Test 3 µL diluted ChIP elution or chromatin input samples in 10 µL qPCR reactions
6. Set up three technical replicates for all samples and standards.

7. To calculate the enrichment of DNA covering the promoter region relative to the control region, first normalise both qPCR values by the respective values obtained for the chromatin input sample. The ratio of the normalised values can then be used as indicator of the specificity of the ChIP protocol.

3.6. Recommendations for deep sequencing and data analysis

1. The ChIP-seq elution samples can be converted easily to Illumina sequencing-compatible libraries using e.g., the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina with dual indexing reagents (NEB). We generally construct libraries with the following modifications: (i) 11 cycles of PCR amplification (for a target yield of ~100 ng), (ii) the size selection step should be conducted for an average fragment size of 150 bp, (iii) because of the small average fragment size and the limited amplification by PCR, the successful removal of indexing reagents after the PCR amplification has to be verified and the last purification step might need to be repeated. Poor libraries such as those prepared from pre-immune control ChIP experiments may show significant levels of remaining indexing reagents.
2. When DNA libraries are prepared from samples with smaller DNA fragment sizes to boost the resolution of ChIP-seq, the reproducibility of the size selection step during library preparation decreases. Therefore, we generally apply an optional, additional step in the data processing: Paired-end sequencing data that allow the determination of the exact fragment size for each read pair after alignment to the reference genome. This allows us to adjust the DNA fragment size distribution computationally by sampling from the read pairs to obtain identical fragment size distributions across different ChIP samples and chromatin input. This step is performed prior to the calculation of ChIP-seq coverage

data[8]. A suitable R script can be found at https://doi.org/10.5281/zenodo.5346581 ('read_sampling.R').

3. This protocol has been optimised to yield ChIP-seq occupancy data with high resolution. Nevertheless, the identification of peaks in the ChIP-seq occupancy data representing binding sites of the target protein remains difficult in the crowded environment of archaeal genomes. To overcome these problems, dedicated peak calling software for ChIP-seq data from prokaryotes has been developed [12]. However, we have found that MACS2 [13], one of the most widely used peak calling algorithms, identifies transcription initiation factor peaks robustly in data generated with this protocol. For single-read sequencing data, MACS2 relies critically on a model building based on high confidence peaks with strong enrichment. The model building process estimates the effective average size of the immuno-precipitated DNA fragments and does only perform with relaxed parameter settings for archaeal ChIP-seq data due to lower number of high confidence peaks compared to eukaryotic ChIP-seq data. This problem can be easily circumvented by opting for paired-end sequencing instead that renders the model building process redundant. Importantly, MACS2 is able to deconvolute the signal from overlapping peaks which are frequent in archaeal ChIP-seq data. MACS2 peak data show a very tight spatial association with mapped TSSs (*see* **Fig. 1e**).

4. Notes

1. Polystyrene tubes provide better energy transfer compared to standard polypropylene tubes and therefore they yield far better sonication results. However, we noticed that these tubes get frequently damaged during sonication leading to approximately 5-15% of the samples being lost. In case biological material is limiting or difficult to obtain, it is

worth considering the use of thin-walled 0.5 mL polypropylene PCR tubes instead together with a 0.5 mL tube rack for the cup horn. You will need to increase the amplitude of the sonicator. We have tested setting the amplitude up to 50% but the resulting DNA size distribution is still broader and less reproducible.

2. Dynabeads Protein G (ThermoFisher) can be used alternatively. These beads have a higher binding capacity compared to Anti-Rabbit IgG Dynabeads, but they will also produce a slightly higher background in ChIP based on pre-immune controls.
3. The available volume of pre-immune serum is usually limiting and the protein yield from the purification significantly lower. We therefore include an ultrafiltration step to reach higher protein concentrations.
4. This protocol will work with a wide variety of *Sulfolobus* and *Saccharolobus* strains and different carbon and nitrogen sources. The growth conditions described here are only exemplary. If cells in later growth phase are to be used, the cross-linking conditions remain the same. However, for the DNA shearing by sonication, the volume of Lysis buffer for the resuspension of the cross-linked cells needs to be scaled up accordingly in order to achieve a similarly low viscosity as otherwise the DNA shearing will be less effective.
5. The fume hood should be placed close to the incubator where the cells are growing. If this is impossible for practical reasons, transport the culture in a polystyrene box as quickly as possible to the fume hood in order to minimise the heat loss during transfer. It is essential to prepare the experiment meticulously and proceed quickly with all steps to avoid inducing any stress response in the cells due to temperature changes.
6. This protocol uses Tris-HCl pH8 as quenching solution rather than the more commonly used glycine. Tris has the advantage of more efficient quenching [1].

7. We have observed that the first cell pellet collected after cross-linking can be unstable and form along the wall of the centrifuge bucket. Re-pelleting this cell material in 50 mL tubes during the 1x PBS washing steps yields stable pellets.
8. Frozen cell material can be used for additional sonication rounds. This is particularly useful when DNA shearing conditions still need to be optimised.
9. If the average DNA size is > 200 bp or the size distribution is very broad, the DNA sonication step needs to be repeated and optimised by (i) further dilution of the chromatin by increasing the volume of Lysis buffer, (ii) increasing the sonication amplitude to 20, or (iii) extending the total sonication time.
10. It is critical to assess the antibody amount to be used in indirect ChIP experiments because amounts exceeding the binding capacity of the Dynabeads used for antibody immobilisation are likely to decrease the DNA yield of the ChIP experiment.
11. This incubation step should not be extended because longer incubation times increase the amount of non-specific DNA co-purified with the beads. The common practice of preclearing the lysate with the same type of magnetic beads used for the ChIP experiment did not lead to a reduction in background in our hands.
12. To increase the resolution of the genome-wide occupancy maps, our cross-linking and ChIP protocol can be combined with the ChIP-exo library preparation strategy where the 5'-ends of the immuno-precipitated DNA are trimmed by λ exonuclease. We have successfully combined our ChIP-protocol with both the original ChIP-exo protocol for Illumina sequencing [14] and the streamlined protocol ChIP-exo 5.0 [15] (*see* **Fig. 1d**). It is important to notice that unlike ChIP-seq, ChIP-exo requires immuno-precipitated complexes rather than purified DNA as starting material. Thus, library preparation has to

begin immediately after the TE buffer washing step of the immuno-precipitation protocol (*see* Subheading 3.4, **step 17)**.

13. Before proceeding with DNA isolation and the ensuing qPCR analysis or ChIP-seq library preparation, it is essential at the beginning of the ChIP optimisation to verify that the antibodies are capable to immuno-precipitate the target after cross-linking. To this end, withdraw 25 µL from the elution sample (*see* Subheading 3.4, **step 20**), the chromatin input (*see* Subheading 3.4, **step 2**) and the supernatant (flow, *see* Subheading 3.4, **step 12**) and subject these samples to Western blotting and immuno-detection. Note that, depending on the choice of secondary antibodies for the immuno-detection, the light and heavy chain of the antibody present in the ChIP elution sample might also be detected (*see* **Fig.1b**).

5. References

1. Hoffman EA, Frey BL, Smith LM, Auble DT (2015) Formaldehyde crosslinking: a tool for the study of chromatin complexes. J Biol Chem 290 (44):26404-26411. doi:10.1074/jbc.R115.651679
2. Baranello L, Kouzine F, Sanford S, Levens D (2016) ChIP bias as a function of cross-linking time. Chromosome Research 24 (2):175-181. doi:10.1007/s10577-015-9509-1
3. Reichelt R, Gindner A, Thomm M, Hausner W (2016) Genome-wide binding analysis of the transcriptional regulator TrmBL1 in Pyrococcus furiosus. BMC Genomics 17:40. doi:10.1186/s12864-015-2360-0
4. Smollett K, Blombach F, Reichelt R, Thomm M, Werner F (2017) A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. Nat Microbiol 2:17021. doi:10.1038/nmicrobiol.2017.21
5. Duc TN, Hassanzadeh-Ghassabeh G, Saerens D, Peeters E, Charlier D, Muyldermans S (2012) Nanobody-Based Chromatin Immunoprecipitation. In: Saerens D, Muyldermans S (eds) Single Domain Antibodies: Methods and Protocols. Humana Press, Totowa, NJ, pp 491-505. doi:10.1007/978-1-61779-968-6_31
6. Wang K, Lindås A-C (2018) Chromatin Immunoprecipitation Assay in the Hyperthermoacidophilic Crenarchaeon, Sulfolobus acidocaldarius. In: Visa N, Jordán-Pla A (eds) Chromatin Immunoprecipitation: Methods and Protocols. Springer New York, New York, NY, pp 139-146. doi:10.1007/978-1-4939-7380-4_12
7. Wilbanks EG, Larsen DJ, Neches RY, Yao AI, Wu CY, Kjolby RA, Facciotti MT (2012) A workflow for genome-wide mapping of archaeal transcription factors with ChIP-seq. Nucleic Acids Res 40 (10):e74. doi:10.1093/nar/gks063

8. Blombach F, Fouqueau T, Matelska D, Smollett K, Werner F (2021) Promoter-proximal elongation regulates transcription in archaea. Nat Commun 12 (1):5524. doi:10.1038/s41467-021-25669-2
9. Brock TD, Brock KM, Belly RT, Weiss RL (1972) Sulfolobus: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Archiv für Mikrobiologie 84 (1):54-68. doi:10.1007/BF00408082
10. Zillig W, Stetter KO, Wunderl S, Schulz W, Priess H, Scholz I (1980) The Sulfolobus-"Caldariella" group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases. Archives of Microbiology 125 (3):259-269. doi:10.1007/BF00446886
11. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, Van der Oost J (2001) The complete genome of the crenarchaeon Sulfolobus solfataricus P2. Proc Natl Acad Sci U S A 98 (14):7835-7840. doi:10.1073/pnas.141222098
12. Neches RY, Seitzer PM, Wilbanks EG, Facciotti MT (2014) In a fit of pique: Analyzing microbial ChIP-Seq data with Pique. PeerJ PrePrints 2:e290v292. doi:10.7287/peerj.preprints.290v2
13. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9 (9):R137. doi:10.1186/gb-2008-9-9-r137
14. Serandour AA, Brown GD, Cohen JD, Carroll JS (2013) Development of an Illumina-based ChIP-exonuclease method provides insight into FoxA1-DNA binding properties. Genome Biol 14 (12):R147. doi:10.1186/gb-2013-14-12-r147
15. Rossi MJ, Lai WKM, Pugh BF (2018) Simplified ChIP-exo assays. Nat Commun 9 (1):2842. doi:10.1038/s41467-018-05265-7
16. Diaz A, Park K, Lim DA, Song JS (2012) Normalization, bias correction, and peak calling for ChIP-seq. Stat Appl Genet Mol Biol 11 (3):Article 9. doi:10.1515/1544-6115.1750
17. Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R (2010) A single-base resolution map of an archaeal transcriptome. Genome research 20 (1):133-141. doi:10.1101/gr.100396.109

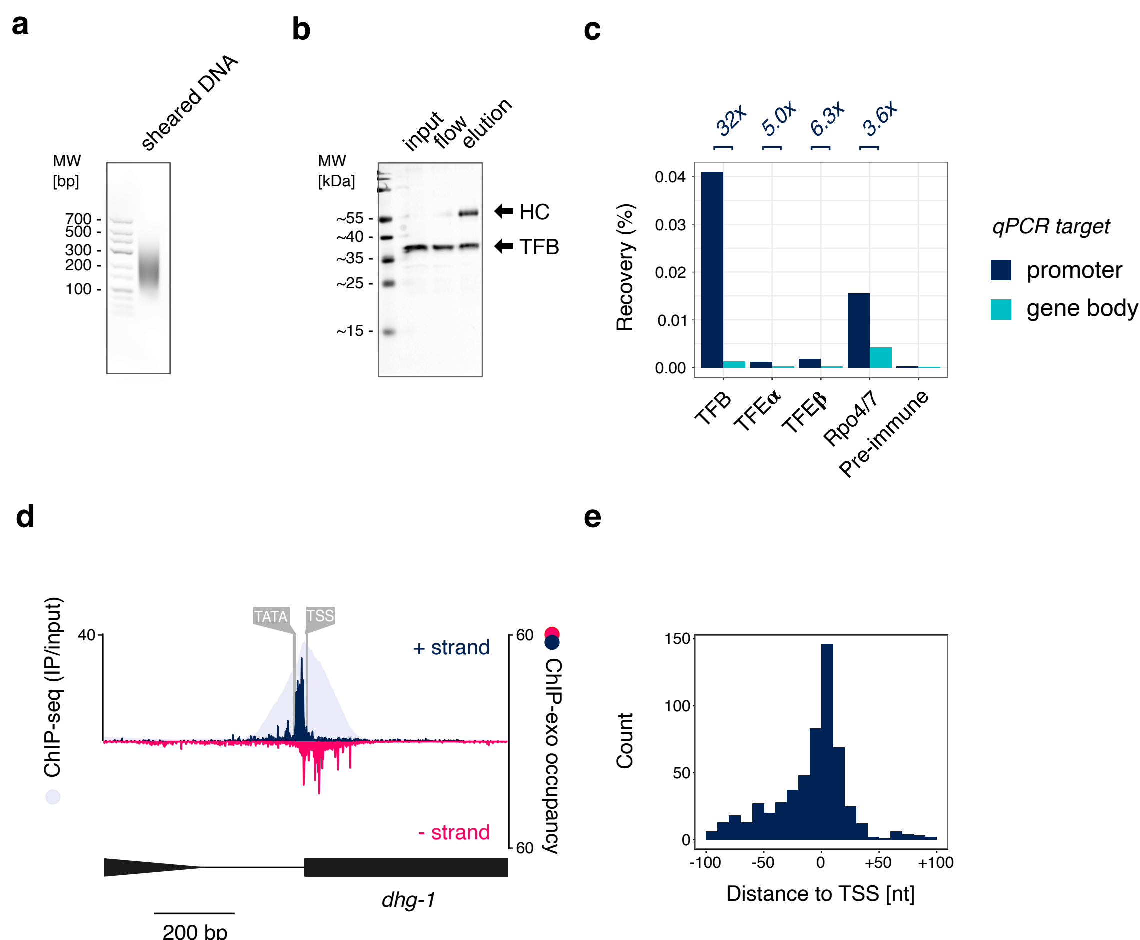

Figure 1: Key steps in the evaluation and optimisation of the ChIP-seq protocol

(a) Ethidium bromide-stained agarose gel with DNA obtained from sheared *S. solfataricus* chromatin. The expected result is a narrow size distribution with an average size of about 150 bp.

(b) Exemplary Western blot showing the detection of TFB in different fractions from a ChIP experiment. The position of the ChIP target TFB and the signal of the heavy chain from the antibody used during the ChIP experiment are indicated.

(c) Plot depicting qPCR analysis of a ChIP elution experiment with the indirect method for initiation factors TFB, TFEα, TFEβ, RNA polymerase subunits Rpo4/7 and the pre-immune control. qPCR was carried out for two genomic loci: the promoter of *gapN3* (SSO3194) and a region about 500 bp downstream of the TSS within the gene body. Quantifications were

calculated as recovery relative to chromatin input. The ratio between promoter and gene body quantifications, both normalised against chromatin input, is shown on top of the graph.

(d) ChIP-seq and ChIP-exo coverage plots for TFB obtained by indirect ChIP on the *dhg-1* gene. The plot shows the resolution difference between these two methods of library preparation. The position of the TSS and the TATA-box promoter element is indicated. ChIP-seq data represent the arithmetic mean of two biological replicates divided by their respective chromatin input signal and scaled using the signal extraction scaling method [16]. The ChIP-exo signal represent the geometric mean of three biological replicates scaled to 1x genome coverage.

(e) ChIP-seq peaks obtained with this protocol show tight spatial association with TSSs. Histogram showing the distance relationship between *S. solfataricus* TFB summit positions from MACS2 peak calling data and mapped TSSs. 450 out of 1,035 mapped primary TSSs [17] were associated with a TFB peaks summit in ≤ 40 bp distance. The skew towards positions upstream of the TSS (represented as negative distance) is due to closely spaced divergent promoters that result in overlapping TFB peaks beyond the deconvolution capability of the MACS2 peak calling algorithm. Data for panels (d) and (e) were obtained from [8], deposited as GSE141290 at NCBI GEO.