UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE FÍSICA DE SÃO CARLOS

ALFREDO JOSE FLOREZ ARIZA

Structural analysis of DNA wrapping in bacterial transcription initiation complex by transmission electron microscopy and single particle analysis

São Carlos

2018

ALFREDO JOSE FLOREZ ARIZA

Structural analysis of DNA wrapping in bacterial transcription initiation complex by transmission electron microscopy and single particle analysis

Dissertation presented to the Graduate Program in Physics at the Instituto de Física de São Carlos, Universidade de São Paulo to obtain the degree of Master of Science.
Concentration area: Applied Physics
Option: Biomolecular Physics
Advisor: Dr. Rodrigo Villares Portugal

Corrected Version

(Original version available on the Program Unit)

São Carlos
2018

Florez Ariza, Alfredo Jose
Structural analysis of DNA wrapping in bacterial transcription initiation complex by transmission electron microscopy and single particle analysis / Alfredo Jose Florez Ariza; advisor Rodrigo Villares Portugal - revised version -- São Carlos 2018.
93 p.

Dissertation (Master's degree - Graduate Program in Física Aplicada Biomolecular) -- Instituto de Física de São Carlos, Universidade de São Paulo - Brasil , 2018.

1. Transcription initiation complex. 2. RNA polymerase. 3. DNA wrapping. 4. Transmission electron microscopý. 5. Single particle analysis. I. Portugal, Rodrigo Villares, advisor. II. Title.

With love to my parents, Norma and Enrique, my brother Diego and my sister Nicole,
and to my lovely Annar.

ACKNOWLEDGMENTS

I thank my parents, Norma and Enrique, for all this time of sacrifice. They are a big part of my motivation, and although life has separated us, I am sure that someday we will be able to meet again and be happy. I love them incalculably.

To my brothers, Diego and Nicole, who with their joy and their company, made my days happier. They motivate me to be a better person; they encourage me to take on new challenges. I miss them every day.

To my love Annar, for these more than five years we are together. It has been complicated to be separated, but your unconditional support, love and great friendship, makes me feel that soon we will be together again. I love you.

To my good friend Juan, his brother Sebastian, and to his parents Cesar and Esperanza. Thanks for your constant support offered. All your affection will be engraved in my memory.

To my advisor, Rodrigo Portugal, thanks for all the support and trust placed. Although more than one problem occurred, he always motivated me constantly to continue. Without a doubt, his orientation in addition to his excellent human and academic quality, is a reason for reference and admiration for me.

To Professor Marin Van Heel, for his willingness and academic advice. The minutes that he uses for us are of great help to improve daily. It is really very rewarding and valuable to be able to count with him in the laboratory.

To Professor Daniel Guerra, for all the confidence placed in me during the stage in which I was his student. Undoubtedly, that stage under his academic guidance, allowed me to acquire new knowledge that was vital to get involved in this work. Thank you very much for the opportunity, the support, and the friendship offered.

I would like to thank the new friends I met here. To my good friend Marcelo, thank you very much for so much support during these years, for your unconditional friendship and for your good advice. To Alexandre Cassago, for his dedicated support during the beginning of this work, for his dedication and complete willingness to develop this research. To my colleagues and new friends of the LNNano: Antonio, Débora, Daniela, Karo, Juliana, Katye, and Rafael, for the pleasant shared moments and scientific discussions.To the Institute of Physics of Sao Carlos (IFSC), to the National Nanotechnology Laboratory (LNNano) and to the CNPEM for all the infrastructure offered, and to the CAPES agency for its financial support.

“Science is not only compatible with spirituality; it is a profound source of spirituality.”

— Carl Sagan, TheDemond-Haunted World: Science as a Candle in the Dark.

ABSTRACT

FLOREZ ARIZA, A. **Structural analysis of DNA wrapping in bacterial transcription initiation complex by transmission electron microscopy and single particle analysis.** 2018. 93p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2018.

The transcription initiation is the first step in gene expression and an important regulation step in all living organisms. In bacteria, it has been proposed that DNA bending and its wrapping on the surface of *E. coli* RNAP might facilitate the opening of the transcription bubble, which is necessary for the initiation of gene transcription. In this work, it is shown the first structural study to evaluate a DNA wrapping model, including its length and the relative position in the bacterial transcription initiation complex (RP complex), assembled between RNA polymerase-σ^{70} holoenzyme (RNAP) and a λPR promoter (-100 to +30 wild type). RP complex was prepared and negatively stained with 2% uranyl acetate on a thin-carbon coated grid and the data acquisition of 500 images was performed in a JEM-2100 (JEOL, Japan) microscope equipped with an F-416 CMOS camera (TVIPS, Germany). Single particle analysis of 16,015 particles, grouped in 666 class-averages, was conducted using IMAGIC 4D software (Image Science, Germany) to obtain a three-dimensional model of the RP complex at 20Å resolution. After the rigid-body fitting of the RNAP crystallographic structure (PDB 4YG2) and the modeled DNA promoter, it was observed that the regions 1.2 and 4.2 of the σ^{70} subunit interacts with the consensus zones, -10 and -35 hexamers of the promoter. Furthermore, it was possible to observe that αCTDs (C-terminal domain) in both alpha subunits would be oriented to facilitate the interaction with the first and second UP-elements regions, respectively (centered around –50 and -75 positions in the promoter). These was enabled by the presence of the characteristics motifs helix-hairpin-helix in these domains. In addition, the downstream DNA, from the transcription bubble, appears to be inside the protein main channel, oriented in a way to enable interactions with the RNAP clamp and jaws. Finally, it was observed that the DNA wrapping has ~32 nm of total length and involves a promoter bent of ~255° around the RNAP surface. The 3D-model obtained in this study is the very first direct structural confirmation of the DNA promoter wrapping in a bacterial transcription initiation complex.

Keywords: Transcription initiation Complex. RNA polymerase. DNA wrapping. Transmission electron Microscopy. Single particle analysis.

RESUMO

FLOREZ ARIZA, A. **Análise estrutural do enovelamento do DNA no complexo da iniciação de transcrição bacteriano usando microscopia eletrônica de transmissão e análise de partículas isoladas.** 2018. 93p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2018.

A iniciação da transcrição é o primeiro passo na expressão gênica e importante ponto de regulação em todos os organismos vivos. Em bactérias, foi proposto que o enovelamento do DNA na superfície da RNAP de *E. coli* pode facilitar a abertura da bolha de transcrição, necessária para o início da transcrição gênica. Neste trabalho, é apresentado o primeiro estudo estrutural direto para avaliar o comprimento do enovelamento do DNA e sua posição no complexo de iniciação da transcrição bacteriana (complexo RP), montado entre a holoenzima RNA polimerase-σ70 (RNAP) e um promotor λPR (-100 para +30, tipo selvagem). Amostras do complexo RP foram preparadas e contrastadas negativamente com 2% de acetato de uranila em uma grade com filme fino de carbono e a aquisição de 500 imagens foi realizada em um microscópio JEM-2100 (Jeol, Japão) equipado com uma câmera CMOS F-416 (TVIPS, Alemanha). A análise de partículas isoladas de 16.015 partículas, agrupadas em 666 médias de classe, foi conduzida usando o software IMAGIC 4D (Image Science, Alemanha) para obter um modelo tridimensional do complexo RP, a 20Å de resolução, estimado pelo critério de ½ bit. Após o ajuste de corpo rígido da estrutura cristalográfica da RNAP (PDB 4YG2) e do promotor de DNA modelado, observou-se que as regiões 1.2 e 4.2 da subunidade σ^{70} interagem com as zonas de consenso, hexâmeros -10 e -35, do promotor. Além disso, foi possível observar que os αCTDs (domínio C-terminal) em ambas as subunidades alfa estariam orientados para facilitar uma possível interação com a primeira e segundas regiões dos elementos UP, respectivamente (centradas em torno das posições –50 e -75 do promotor). Estas seriam possíveis devido à presença de alguns motivos de características hélice-grampo-hélice nesses domínios. Além disso, a região do promotor, *downstream* da bolha de transcrição, parece estar dentro do canal principal da proteína, orientado de forma a possibilitar interações com o *clamp* e *jaw* da RNAP. Finalmente, foi observado que o comprimento total do enovelamento de DNA envolve cerca de 32 nm e 255° de rotação do DNA ao redor da superfície da RNAP. Portanto, este modelo 3D é a primeira confirmação estrutural direta do enovelamento de DNA em um complexo bacteriano de iniciação da transcrição.

Palavras-chave: Complexo da iniciação da transcrição. RNA polimerase. Enovelamento do DNA. Microscopia eletrônica de transmissão. Análise de partículas isoladas.

LIST OF ABBREVIATIONS AND SYMBOLS

CCD	Charge-coupled Device
CTF	Contrast Transfer Function
TEM	Transmission electron microscopy
SPA	Single Particle Analysis
MRA	Multi-reference Alignment
MAS	Multivariate Statistical Analysis
ABC	Alignment by Classification
PDB	Protein Data Bank
αCTD	alpha subunit Carboxi-terminal domain
DNA	Deoxyribonucleic acid
RNAP	RNA polymerase
RP	Transcription inititation complex

CONTENTS

1 INTRODUCTION

1.1 Bacterial transcription initiation complex

1.1.1 General aspects

The assembly of transcription initiation complex (RP) is the first step for the gene expression and an important regulation point for all living organisms. One of the major players in the formation of this complex is the RNA polymerase holoenzyme (RNAP), which is formed from the binding of the core enzyme with the sigma factor. Bacterial core RNA polymerase consists of five subunits, $\alpha_2\beta\beta'\omega$, with a molecular mass of 370 kDa (Figure 1).[1] This core enzyme is capable of non-specific DNA binding and starts the RNA synthesis from DNA ends or nicks. The transcription machinery must recognize with high specificity the sites where genes begin, called promoters, among a huge extension of non-specific binding sites offered by genomic DNA. The core enzyme needs to bind the sigma factor (σ) to form the transcriptionally active holoenzyme. This σ factor has the capability to specificity recognize some consensus regions on the DNA promoter and thus start the process of forming a stable RP complex.[2,3] The nature of the sigma factor with which the core enzyme would interact depends on the environmental conditions and the identity of the promoter. Bacteria possess an essential housekeeping σ factor that controls many different promoters. In *Escherichia coli* it is the σ^{70} (molecular weight of 70 KDa) and in Bacillus subtilis and other Gram-positive bacteria it is the σ^{A}.[4] In the case of the σ^{70} factor, it recognizes the -35 and -10 hexamers consensus zones, upstream from the transcriptional start site (+1) of the promoter (Figure 1).[5] The sigma factor binding to the consensus zone results in the formation of the closed complex (RPC).That induces significant conformational changes in both, the RNAP and the DNA, including the transcription bubble formation, to form a stable transcription initiation open complex (RPo).[2,5]

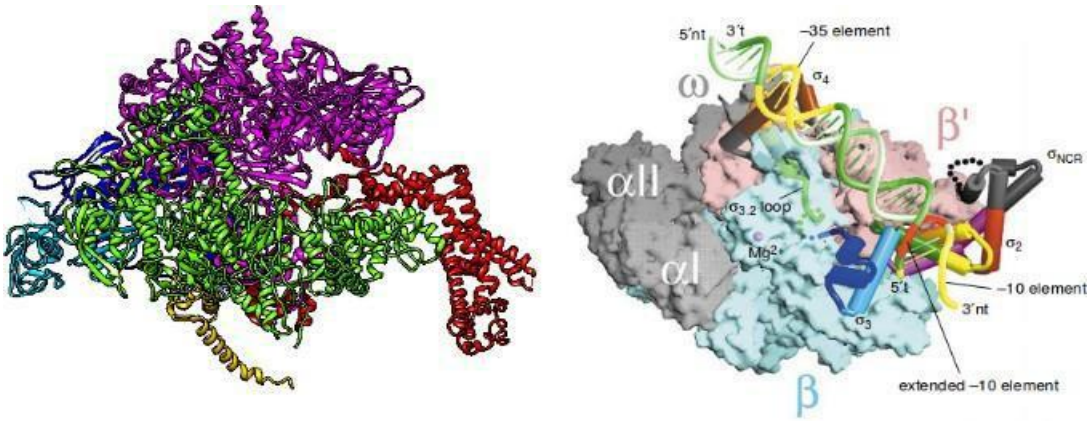


Figure 1 - Crystallographic structure of E. coli RNA polymerase sigma 70 holoenzyme (left: PDB 4YG2). The σ^{70} is in red, ω in yellow, β in magenta, β' in green, and the alpha subunits are shown as follow: α^{I} in blue and α^{II} in light-blue. Structure of the close RP complex between the RNAP holoenzyme and fork-junction DNA. The core enzyme is shown in surface representation and the sigma factor is shown as a Cα backbone worm. It can be observed the -35 and the -10 elements (in yellow) are recognized by the region 4 and 2 of the sigma subunit respectively.

Source: Adapted from MURAKAMI;[6] MURAKAMI; DARST.[2]

All these conformational changes are associated with kinetic intermediates which vary in stability and lifetime among different promoters, suggesting that the DNA sequences determine different configurations of the RP complex.[7]

1.1.2 The σ^{70} factor

The sigma factor plays a crucial role in the specific promoter recognition. The σ70 subunit has four different domains or regions (from 1 to 4) and one non-conserved region (NCR). Each of these regions have different sub-domains (Figure 2).

There are several biochemical and structural studies supporting the idea that the DNA recognition and binding, is accomplished by the interaction of the $\sigma_{4.2}$ sub-domain with the -35 hexamer, while $\sigma_{3.0}$ sub-domain binds to the -10 hexamer. The $\sigma_{1.2}$ and $\sigma_{2.1}$ – $\sigma_{2.4}$ regions interact with the transcription bubble[8] and the σNCR interacts with the promoter DNA upstream of the -10 element.[9]

The TGn element on the promoter, a consensus region which has been reported be crucial for an optimal activity for different promoters is recognized by the $\sigma_{3.0}$ subdomain, possibly through E458 and H455 residues as was shown by some mutagenesis assays.[10] This interaction was also shown to be present in *T. aquaticus* RNAP holoenzyme-fork DNA complex, through analogous residues in the σ^A factor.[11]

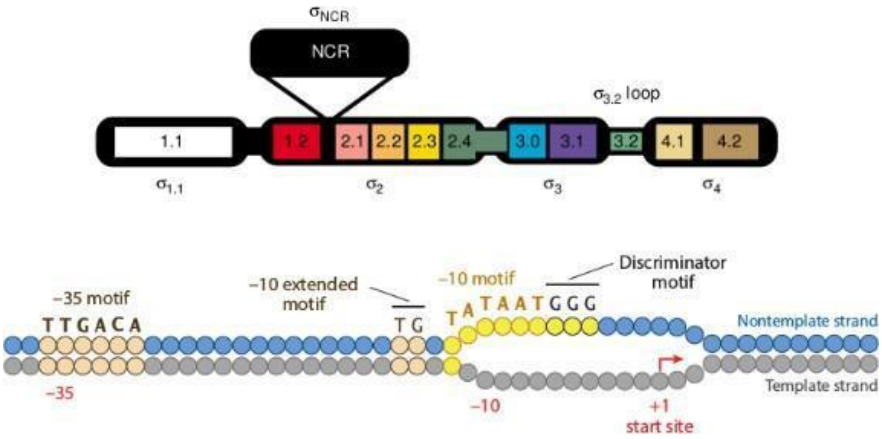


Figure 2 - (A) Sketch of the *E. coli* σ^{70} sub-domains. (B) Schematic representation of the DNA canonical promoter regions recognized by the σ^{70} sub-domains to form the RP complex. Source: Adapted from FEKLÍSTOV;[5] MURAKAMI; DARST.[2]

Furthermore, the $\sigma_{1.1}$ sub-domain seems to perform a regulatory role. Studies using NMR spectroscopy showed that in absence of RNA core enzyme, the σ^{70} is auto-inhibited to DNA binding. It is suggested that there is an "intra molecular interaction" between the 4.2 sub-domains, which should bind to -35 hexamer, with the region $\sigma_{1.1}$.[12] Moreover, there is the hypothesis that when the σ^{70} binds to the DNA, the $\sigma_{1.1}$ blocks the “primary” or main channel of the enzyme, during the formation of the transcription initiation complex. After major conformational changes, the $\sigma_{1.1}$ is expelled, allowing the DNA promoter to completely enter the primary channel (Figure 3). This hypothesis is supported by the structural study of the Phage T7 Gp2 protein inhibition of the RNAP-σ^{70}, which shows that there is a ternary interaction between the $\sigma^{1.1}$-Gp2-β'jaw, that prevent the $\sigma^{1.1}$ to leave the main channel inhibiting the function of the enzyme.[13]

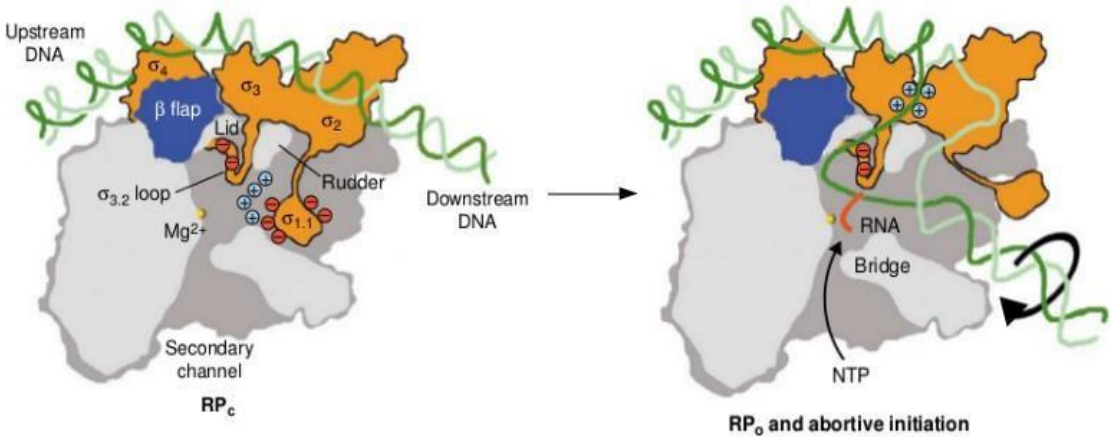


Figure 3 - (A) Cartoon representation of conformational changes in the transition from RPc to the RPo complex. In the figure the core E. coli RNA polymerase is shown in grey – except the β-flap shown in blue- and the σ^{70} factor is in orange. Each σ domain is indicated by name. Once the RPo complex is formed, the $\sigma_{1.1}$ is out of the main channel.
Source: Adapted from MURAKAMI; DARST. [2]

Some of the conformational changes, occurred during the formation of the transcription initiation complex, involve portions of DNA that are considerably distant to the transcription-starting site. These regions on the promoter, known as UP-elements, are upstream regions to the -35 hexamer, and generally possess phased A-T tracks with a certain length that could interact with the α subunits of RNAP.

1.1.3 Alpha subunits and UP-elements

The RNAP has two alpha subunits both have the same polypeptide chain. Each α subunit has two independent folded domains: the amino-terminal domain (αNTD, residues 8–233), responsible for dimerization and interactions with β and β', and the carboxyl-terminal domain (αCTD, residues 245–329). A flexible tether, known as α-linker, of about 18 residues length, connects the two domains.[14]

There is experimental evidence that both αCTDs have a key role in the formation of transcription initiation complex. These are capable of binding DNA specific sequences, commonly called UP-elements. These UP-elements, located upstream the -35 hexamer, are not conserved as high as the -10 and -35 hexamers, but exhibit interesting features, as to own A-T rich regions (of ~10 bp length or more,

separated with a periodicity of ~10.3 bp), providing a particular curvature to the DNA, which seems to be related with the promoter strength and transcription regulation.[15,16]

In this sense, biochemical and genetic screening studies showed that the αCTDs can bind UP-elements mainly through seven residues: L262, R265, N268, C269, G296, K298, S299, all of which belong to a helix-harping-helix motif present in this domain.[17] Even more, X-ray crystallography studies showed that αCTD is oriented to interact with the DNA minor-groove and backbone of the UP-elements. Moreover, because these UP- elements can be in close proximity to the -35 hexamers, the αCTD could be possibly interacting with the region 4 of the σ factor, aiding to stabilize the RP complex.[18,19]

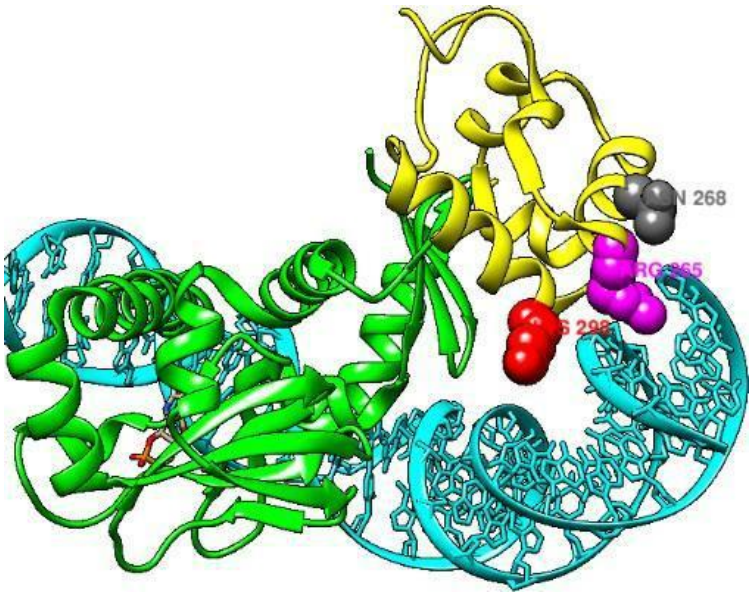


Figure 4 – Crystallographic structure of the ternary complex CAP-DNA-αCTD. The CAP protein is in green, the αCTD in yellow and the DNA UP-element region is in cyan. In the figure can be observed that the residues (in spheres) lys-298 (red), arg-265 (magenta) and asn-268 (grey), are oriented to interact with the minor groove and backbone of the DNA UP-element.
Source: Adapted from BENOFF.[18]

The effects of the UP-elements on promoter strength were established by different studies that evaluated how different upstream sequences can affect the relative activity among promoters. In this sense, Ross and co-workers showed that the two UP- elements (proximal UP-element, -38 to -41 and distal UP-element, -51 to -60) of the wild-type rrnB P1 promoter, improved the *in vivo* transcription activity of the lacZ core promoter (+52 to -37) up to ~33 fold when having both UP-elements.[20] But even each one of the UP- elements, separately, were able to improve the transcription activity. Moreover, it was reported in the same study that *E. coli* RNAP holoenzyme lacking αCTDs was not able to interact with the UP-element regions (confirmed by foot-printing assays), and that the transcription activity was completely dependent on this interaction.[20] Furthermore, this kind of UP-elements/αCTDs interactions seems to be present in different bacterial phyla. For example, a study performed in *Helicobacter pylori*, by Nuclear Magnetic Resonance (NMR), showed that both αCTDs interact directly with the UP-element of the DNA promoter and NikR protein to form a stable ternary transcription initiation complex.[21]

The interactions described above, involve DNA regions centered on -45 to -60 bp upstream the transcription start site. However, for some promoters, like the λ PR promoter, it was reported that the RNAP-αCTDs was able to interact with far upstream zones of the DNA, up to -90 bp upstream. This kind of new interaction between those UP- elements and RNAP would only be possible if the DNA promoter underwent a significant curvature around the RNAP surface. This interesting and relatively new phenomenon, known as *DNA wrapping*, will be detailed in the following section.

1.1.4 The "wrapping" phenomenon

As mentioned before, the interaction between those upstream promoter regions and RNAP would only be possible if the DNA promoter underwent a significant curvature around the RNAP surface, hinting a DNA wrapping formation. In this sense, experimental evidence suggest that DNA wrapping would entail a DNA curvature of ~300° around the RNAP surface (Figure 5).[22] Concerning the λPR promoter, different studies by Atomic Force Microscopy (AFM), proposed that in the RP complex, the promoter was forming a "wrapping" of ~30 nm (~90 bp)

around the RNAP surface (Figure 6). Also, it was reported that this DNA wrapping is reduced to ~ 3 nm when the complexes are formed with RNAP lacking the αCTDs, and to ~15nm when the α-linker is missed. Those observations seemed to be according with different DNase and OH radical foot-printing assays, performed in previous studies. [23–26]

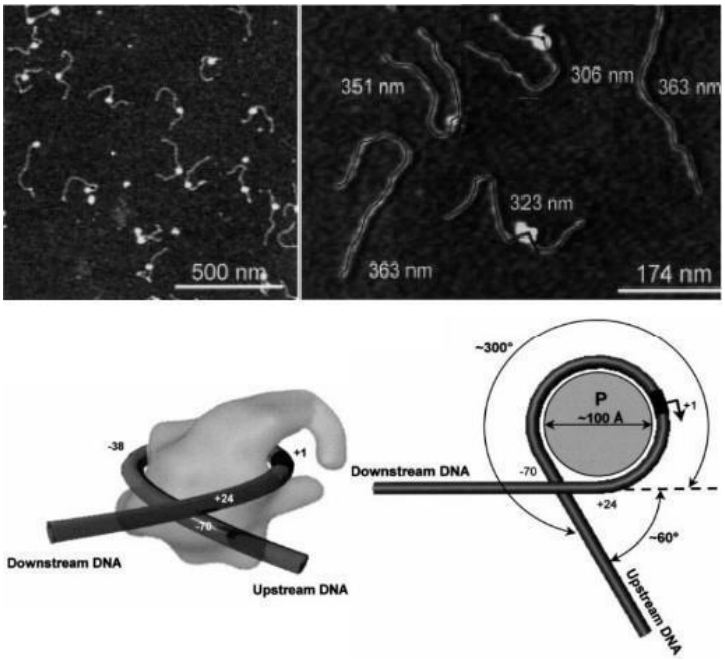


Figure 5 - A square patch of one AFM micrograph of RP complexes (top left) assembled between E. coli RNAP and a DNA fragment containing the λ PR wild type promoter (-100 to +34). A zoom in image (top right) showing the RP complexes and the measures of the DNA contour length free and in complex. Schematic representation of the RP complex (bottom), showing the RNAP in light grey and the DNA promoter in dark grey. The DNA trajectory has proposed around the RNAP surface, with a bend angle of 300°.

Source: Adapted from RIVETTI.[22]

Recently, a study using fluorescence resonance energy transfer (FRET), on the same system, measured the distance between fluorescent labels in positions +14 and -100 relatives to the transcription starting site (+1), confirming that these locations come close to each other in a way that is consistent with the proposed DNA wrapping.[27]

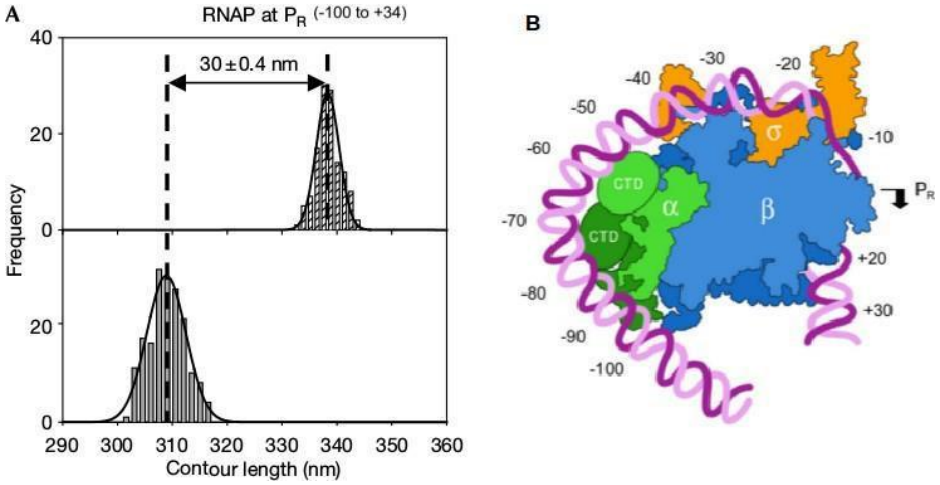


Figure 6 – (A) Contour length distributions of the DNA both in absence (dashed bars, top) and presence of the RNAP (gray bars, bottom). The DNA compaction in the RP complex is around 30 nm. (B) Cartoon representation of the RP complex, showing that UP-elements, centered around -50 and-90 bp, would be interacting with the αCTDs

Source: Adapted from MANGIAROTTI.[23]

Furthermore, another study showed that the DNA wrapping length is significantly reduced by allosteric regulators like the guanosine tetra phosphate (ppGpp) and Dksa protein.[28] The wrapping phenomenon have been also suggested to be present in other promoters like rrnB P1, hdeAB and tRNAtyr;[29–31] and also for eukaryotes, as was observed for the adenovirus promoter in complex with the RNA pol II of *S. cerevisiae*.[32,33] All of these studies, seem to indicate that the DNA wrapping is a conserved phenomenon along different kind of promoters and directly related to the promoter strength and stability of the transcription initiation complex.

However, even with all the studies mentioned above, there is no direct structural evidence of the DNA wrapping. Until now, there are questions that remains without answer like which could be the protein domains possibly interacting with all the UP-elements? How could this be related to the formation of a stable RP complex? Which would be the real extension of the DNA wrapping, and the promoter regions involved? In the present work, we present the first direct structural study of DNA wrapping in the RP complex assembled between the *E. coli* RNAP holoenzyme (RNAP-σ^{70}) and the λ PR wild- type promoter (-100 to +30 bp). This study was carried out by transmission electron microscopy (TEM) and single particle analysis (SPA).

1.2 Technique: Transmission Electron Microscopy

1.2.1 Overview

The relationship between structure and function is evident at all levels of the life sciences. Just as evolution has spawned fins, legs, wings, and hands as divergent variations of the same pattern, the bio-machines that make the cell work, are each one made up of molecules with specific structures for their purposes. The enzymes receive the substrate and accommodates it oriented to the functional groups that ensure that the desired chemical reaction takes place. Cytoskeletal proteins are formed as large girders that can assemble networks with the rigidity or flexibility necessary to support the structure and movement of the cell. Molecular motors couple chemical reactions and the use of thermal energy to move in a specific direction to, for example, transport vesicles between organelles. Structural biology proposes to elucidate these and many other processes by solving the functional structure of these macromolecules of life, with atomic resolution.

Since the 1950s, X-ray diffraction has been the predominant technique to carry out this task. Providing a crystal formed by periodic and well-organized repetitions of the molecule under study, it is possible to elucidate its 3D structure with atomic resolution. However, many of these molecules may not form crystals. There are also some limitations to elucidate flexible domains of the molecule under study.[34]

On the other hand, nuclear magnetic resonance (NMR) is based on the detection of the atoms chemical displacement, and therefore their bounds to each other, allowing for the 3D reconstruction. However, when molecules like proteins, are larger than 35-50 kDa, the chemical shift signals may overlap each other, making it challenging or impossible to differentiate them.[35]

Currently, Single Particle Cryogenic Electron Microscopy (Cryo-EM) is capable to achieve resolutions comparable to X-ray crystallography, with the advantage of not requiring the crystallization of the sample and the capacity, in some cases, of giving direct information about different conformational states associated to the biological function of a molecular complex.

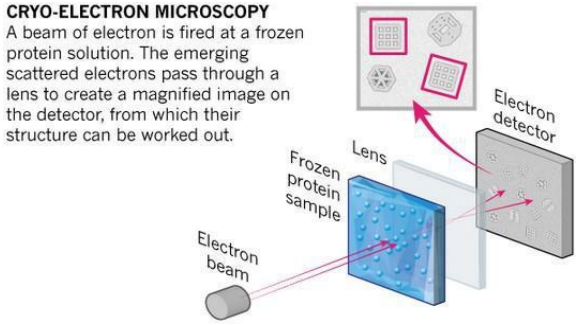


Figure 7 - Schematic representation of how Cryo-EM works. An electron beam, accelerated with 100-300 kV, interacts with frozen protein sample, and generate 2D projections that shown the protein in different random orientations, which will use then to generate a 3D model.
Source: Adapted from CRESSEY; CALLAWAY.[36]

The development of Cryo-EM technique started in the 1980s,[37,38] and for many decades it was referred as a "blobology" science, since it was only able to produce low-resolution maps (15-20 Å), unlike the high-resolution achieved by X-ray crystallography. However, the continuous improvements in methods and instrumentation led to a key situation in 2013 (Figure 8). It was then that the so-called "resolution revolution" was initiated, driven by the new generation of direct electron detectors that present a very high speed and sensitivity, superior to previously used CCD devices.[39–41] With a better signal-to-noise ratio and the ability to produce films instead of single shots, in recent years, Cryo-EM has reached the resolution of traditional X-ray crystallography.[42,43]

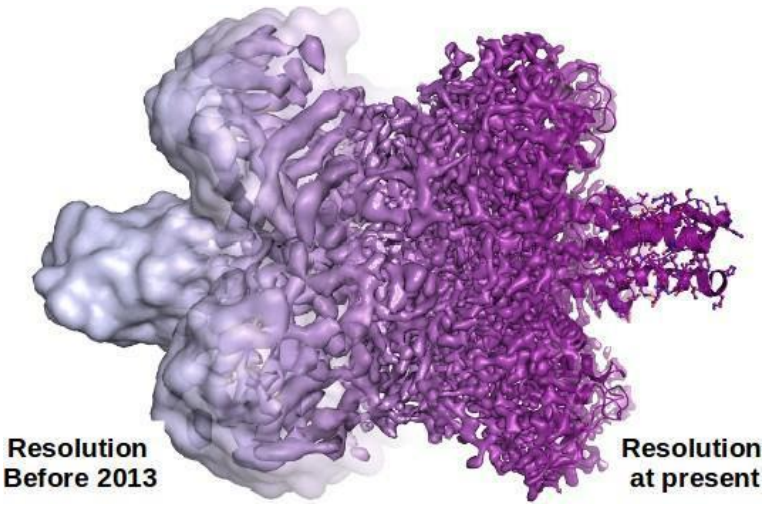


Figure 8 - Improvement of the resolution achieved by Cryo-EM in last few years. From ~20Å before the 2013 year until better than 2Å up today.
Source: Adapted from CRESSEY; CALLAWAY.[36]

In Figure 9, it is shown a graphic summary of a recently reported structure of the hisPEC-NusA complex, obtained by Single particle Cryo-EM with an overall resolution of 3.6 Å. [44]

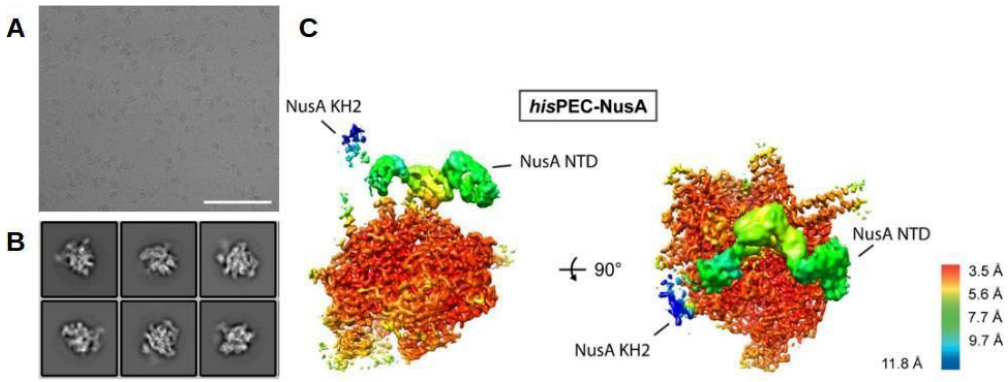


Figure 9 - A micrograph patch of the hisPEC-NusA complex obtained by Cryo-EM (A). Some representative class- averages, showing the complex in different orientations (B). A 3D model reconstructed at ~3.6 Å of global resolution in which is clearly possible observe the amino-acids backbone and nucleic acids electron density (C). The scale bar is 100 nm
Source: Adapted from GUO. [44]

During the last 15 years, thousands of structures have been produced by Cryo-EM. Figure 10 shows the number of density maps for different macromolecular complexes, deposited per year (from 2002 to date) in the public repository “The Electron Microscopy Data Bank” (EMDB).[45] In addition, the resolution achieved for the density maps has been improving year after year, as shown in Figure 11. The improvement trend can be observed in the curves of highest resolution and average obtained during each year (from 2002 to today), for the different density maps deposited in the EMDB. [45]

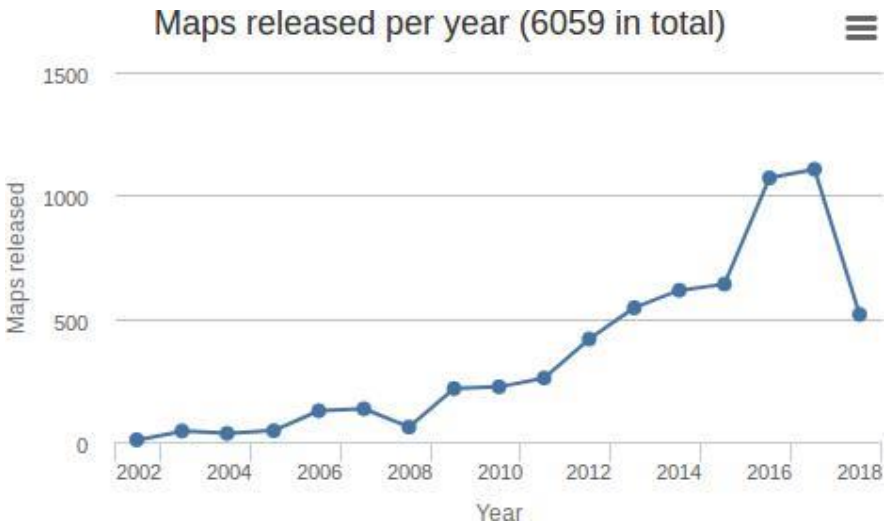


Figure 10 - The cumulative number of 3D maps released (2002 to 2018)
Source: Adapted from EMDB STATISTICS.[45]

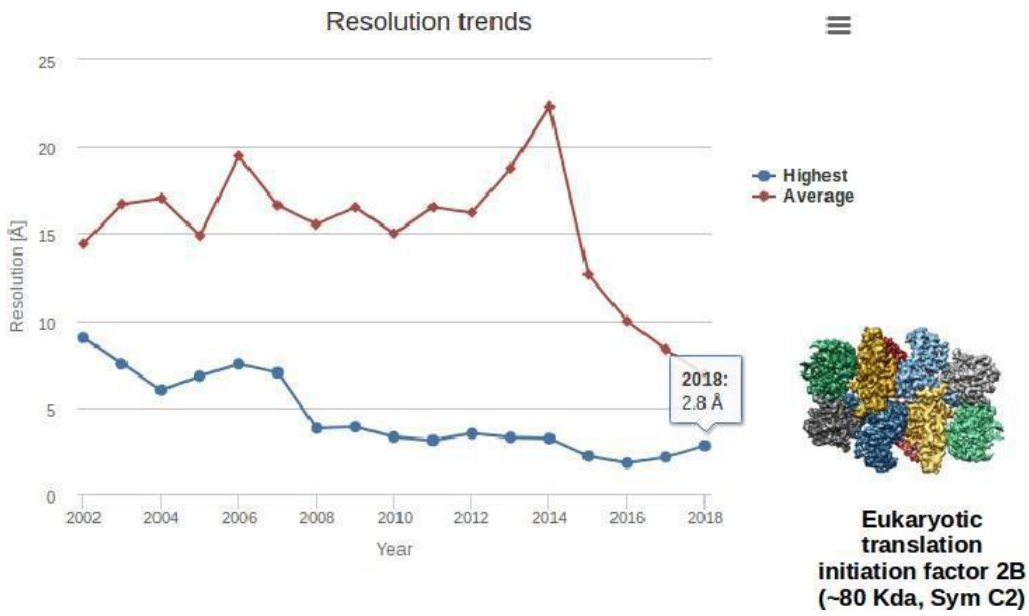


Figure 11 - The resolution trends of Single particle Cyo-EM released maps from 2002 up to April 2018. The red line indicates the average resolution attained in each year. In blue line is shown the highest resolution attained in each year. In the small panel it is shown the segmented surface map of the Eukaryotic Translation initiation factor 2B (~80 kDa), which the structure was solved at 2.8 Å of resolution in April 2018.
Source: Adapted from EMDB STATISTICS.[45]

1.2.2 Sample Preparation

1.2.2.1 Negative Stain

The negative stain method was introduced by Brenner and Horne in 1959, while they were studying the external structure of viral particles.[46] Since the beginning, this technique was widely used to obtain high contrast imaging of different biological complexes. The method consists in mixing a heavy metal salt solution, usually uranyl acetate or uranyl molybdate at 1% or 2% (m/v), with the solution containing the sample. This method provides a high contrast image, but with some downsides. For example, the sample is exposed to the hash vacuum environment of the microscope, together with the large size of the salt crystals and even the access of the staining agent to hydration channels, the structural information that can be obtained is limited to a low resolution. Nowadays, the negative stain is still used to carry out the screening of the sample (2D) and 3D analysis of complexes when the necessary information can be obtained at low resolution. As a screening tool, it is useful for assessing the sample homogeneity, purity and overall quality of the protein or biological complex.[47] Furthermore, if a single particle analysis is performed of these images, it is possible to obtain a 3D-model reconstruction at ~20 Å of resolution.

1.2.2.2 Cryogenic sample preparation

The sample preparation for Cryo-EM consists, in summary, in applying the sample solution onto the grid, removing the excess of liquid and rapidly plunging the specimen into a vessel with liquid ethane, previously thermalized by a liquid nitrogen bath. Due to the fast freezing process, the water molecules of the buffer, will not form a crystalline structure, hexagonal or cubic, forming an amorphous ice, or vitreous ice. This vitreous ice will preserve the protein in its native state and at the same time provide a partial protection against electron beam damage. In this way, the molecules under study will be "embedded" in a layer of amorphous ice, distributed in different random orientations (Figure 12). The subsequent imaging stage will produce several 2D-projections of the particle, allowing to record enough data to achieve a 3D

reconstruction of the biomolecule or complex under study. Since the first reports of vitreous ice specimen preparation, Cryo-EM has undergone a "quantum leap" progression in its attainable resolution and applicability to the study of challenging biological systems.[37,38,48–50]

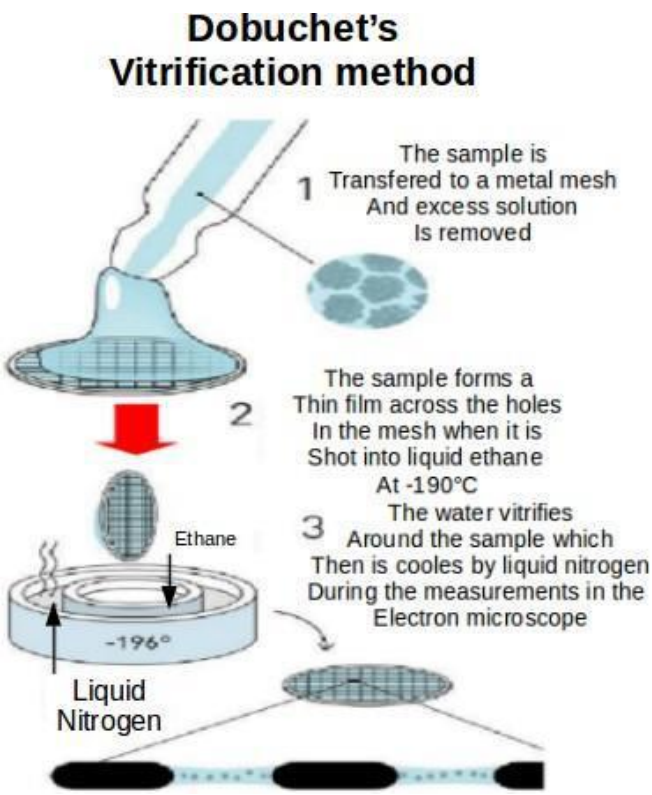


Figure 12 - Schematic representation of samples vitrification for Cryo-EM. This method was introduced by Jacques Dobuchet around 1980's.
Source: Adapted from NOBELPRIZE-ORG.[51]

1.2.3 Instrumentation

The basis of transmission electron microscopy (TEM) is the use of a high-energy electron beam (typically accelerated between 100-300 kV) passing through the sample in a way to gather information of its intern structures. Because of the particle-wave nature of electrons, these can interact with matter in different ways. The image formation arises from a combination of the interaction of elastic-scattered

electron-waves, non-scattered electron-waves and inelastic scattered electron-waves.

It is important to mention that the wavelength of an electron accelerated at 300 kV -like in TEM- is around 2 pm.[52] Therefore, the attainable resolution, typically around a few angstroms, is not limited by the wavelength of the radiation. The resolution limitation is a result of many different aspects, but mainly the low electron doses used during the image acquisition. This use of low electron-doses causes the shot-noise, or quantum-noise, to be the main reason of having images with a signal/noise ratio much smaller than the unit, limiting in this way, the possible resolution to reach. Obviously, through modern methods of images analysis as well as the recent developments of instrumental improvements, like the direct electron-detectors, these limitations can be partially overcome.[53,54]

The typical transmission electron microscope is generally composed by an electron gun, a set of condenser and objective lenses, intermediate lenses, projector lenses, a fluorescent screen on the bottom and detectors. In the figure 13, it is shown an overview diagram of the internal structure of a transmission electron microscope. Some of the components will be discussed later.

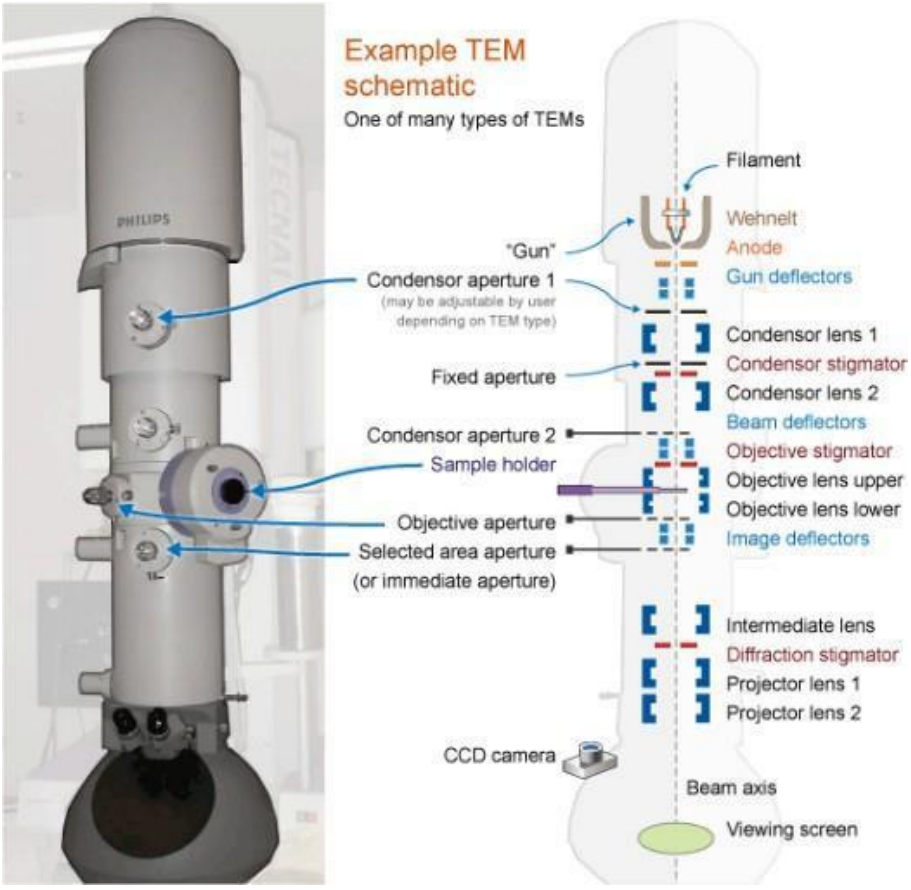
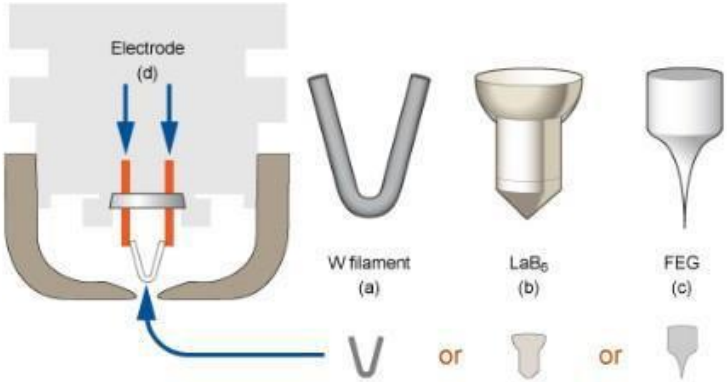


Figure 13 – An image of an Electronic transmission microscope (left) and a generalized design showing its internal structure (right).
Source: Adapted from AMMRF.[55]

1.2.3.1 Electron gun

The electron gun or *cathode*, function as an electron beam emitter, seating up on the top of the microscope column. It is kept at the acceleration voltage, for instance 300 kV, while the rest of the microscope is kept at ground potential. Since the electrons are emitted in a divergent way, the cathode filament is surrounded by a cup-shaped electrode, with a central of approximate 1 mm bore, called *Wehnelt cylinder,* which has the function to conduct the electron beam emitted from the cathode toward the optic axis. It is the tip of this cloud, closest to the Wehnelt cylinder opening that acts as the actual electron source. The electron guns are classified in two types, the thermionic sources (tungsten filament and lanthanum hexaboride, LaB_6) and the field emission guns (FEG). The principal difference between the sources are the brightness and energy spread (Figure 14), which are both directly

related to the resolution that can be reached.[56]



Emission	Thermionic		Field Emission
	W	LaB6	FE
Size (nm)	1×10^5	2×10^4	0.2
Brightness (A/cm2.steradian)	10^4 - 10^5	105 - 10^6	10^7 - 10^9
Energy Spread (eV)	1 - 5	0.5 – 3.0	0.2 – 0.3
Operating Lifetime (hrs)	>20	>100	>300
Vacuum (torr)	10-4 – 10^{-5}	10-6 – 10^{-7}	10-9 – 10^{-10}

Figure 14 - **Top:** Schematic representation of the electron gun, with diverse types of emitters (source), tungsten filament (a), LaB_6 (b) and FEG (c). **Bottom:** a table showing the principal differences between the three electron sources.
Source: Adapted from AMMRF.[55]

1.2.3.2 Wehnelt cylinder

To drive the electrons emitted by the gun toward the optical axis, the cathode filament is surrounded by a cup-shaped electrode that is kept at a slightly more negative potential that the electron source (~500V less). This electrode, called Wehnelt cylinder, has a small ~ 1 mm central bore, through which the electrons will be ejected.[56]

Figure 15 - The Wehnelt cylinder after to be removed from an electron gun Source: Adapted from AMMRF.[55]

1.2.3.3 Condenser lens

The condenser lenses purpose is to manage the electron beam towards the desired direction. Since the electrons are charged particles, the lenses for an electron microscope are electromagnetic lenses, made of ferromagnetic materials, having the capacity of generating electromagnetic fields by a circulating internal current. The generated fields bend the electron beam path to the optical axis. The condenser lenses, located immediately below the Wehnelt, have the objective to control the electron beam that is coming out from the electron gun, thus achieving a better coherence in the illumination on the sample.

Figure 16 - Electromagnetic lenses from an electron microscope (left) and a cartoon showing how could be the electron beam path due to the electromagnetic field generated byb the lenses Source: Adapted from AMMRF.[55]

1.2.3.4 Objective lens

This set of lenses is the first that participate directly in the image formation process. Due to its small focal length (only some millimeters), the sample to be imaging is immersed inside the bore of this lens. It is important to mention that in the back focal plane of this lens, the scattered parallel beams are focused in a single point, thus, a "diffraction pattern" of the sample is formed on this plane. Using another set of lenses, the diffraction information can be recombined to form the image of the specimen. In addition, the objective aperture -a physical aperture- that is positioned just below this lens, aid to eliminate the widely scattered beams, forming what is called *amplitude contrast*. Moreover, the electron beam suffers a phase shift while passing through the sample, also known as *phase contrast*. This phase contrast is what allows, in practical terms, to visualize the particles of the biomolecules under study.[56–59]

1.2.3.5 Intermediate and projector lens

The set of intermediate lenses are capable of impose different magnifications while varying its internal current. Usually, for screening and data collection, the magnification range is about a few hundreds to tens of thousands, respectively. It changes according to the experimental conditions, microscope and position of the camera. As their name suggests, these set of lenses, project the final image into the screen or the detector.[56]

1.2.3.6 Fluorescent screen and detectors

A fluorescent screen is a metal plate coated with a thin (range 10–20 mm, depending on voltage) layer of a fluorescent material that converts impinging electrons into bursts of light. It is placed into the image plane and serves as a tool for optimizing the operation conditions and screening before starting with data collection. Devices like CCD or CMOS cameras and direct electron detectors process the signals from the electron beam that had interacted with the sample, allowing the digital image recording. It is important to mention that unlike shown in Figure 13, for Single Particle Cryo-EM experiments, the cameras are typically

mounted bellow the fluorescent screen. [56,60]

2 OBJECTIVES

2.1 General objective

The objective of this work is to structurally study the DNA wrapping phenomenon in the transcription initiation complex (RP) to better understand the interactions between the RNAP and the DNA, suggesting a possible conformation of the wrapped DNA. The RP will be assembled between the *E. coli* RNA polymerase-σ^{70} holoenzyme and the λ PR wild-type (-100 to +30) promoter, and the structure will be obtained by transmission electron microscopy (TEM) and single particle analysis (SPA),

2.2 Specific objectives

- Amplify the 186 bp length DNA fragment containing the lambda PR wild-type promoter by polymerase chain reaction (PCR). Optimize the RP complex assembly conditions and the TEM grids preparation using negative stain method;

- Perform the data collection and image processing using the SPA to obtain a 3D map of the RP complex;

- Obtain a model of the RP based on molecular fitting of the available RNAP crystallographic structure and an *in silico* modeled DNA promoter, into the 3D density map previously obtained by the SPA;

- Analyze the orientation of different protein domains that could be interacting with the DNA promoter enabling the wrapping formation;

- Investigate and determine the measure of the wrapping length, as well as the possible bent angle of the promoter, and compare the results with previous reported measures.

3 EXPERIMENTAL PROCEDURE AND METHODOLOGY

3.1 Image Processing

In Single Particle Analysis (SPA), the 3D map of the macromolecule is a result of many iterative rounds of alignment, classification, average and angular assignment of several images of particles in random orientations and possibly in multiple conformations. To carry out this task, different software packages are available: IMAGIC 4D, RELION, EMAN, SPIDER, etc.,[61–63] each one perform the image analysis with different philosophies. In this section, the different steps in image analysis will be discussed, focused on methods implemented in IMAGIC 4D software (Image Science, Germany), used in the “Align by Classification” (ABC) processing strategy.[48,64]

Before starting the image processing description, it is useful to mention some important general concepts and examples of practical applications about the Fourier transform, since it is widely used through the image processing. The Fourier transform is a mathematically complex operation that decomposes a function, in real space, into component frequencies represented in the reciprocal space. For an integral and continuous function *f:* **R** → **C**, the Fourier transform will be defined as in Equation (1):

$$F(\kappa) = \int_{-\infty}^{\infty} f(x) \exp(-2\Pi i \kappa x) dx \qquad (1)$$

So, F(K) is the Fourier transform of *f*(x), where K is a real number that express the frequency and *x* is the variable of the function *f* in the real space. In addition, *i* is the imaginary unit. As it was mentioned before, it is denoted that the function *f* is represented in the **real space** and its Fourier transform (**F**) is depicted in the **reciprocal space**. Therefore, the Fourier transform can be represented as a complex number: **F** = x + y*i* or **F**=|**F**|(cosθ + *i* sinθ), where “x” and “y” are the real and imaginary part of the Fourier transform, respectively, and the angle θ is known as the complex argument or **phase**. In the images below (Figure 17), there are some practical examples about the Fourier transform for 2D-images and how the low and high spatial frequencies are related with different features of the image in the real space.

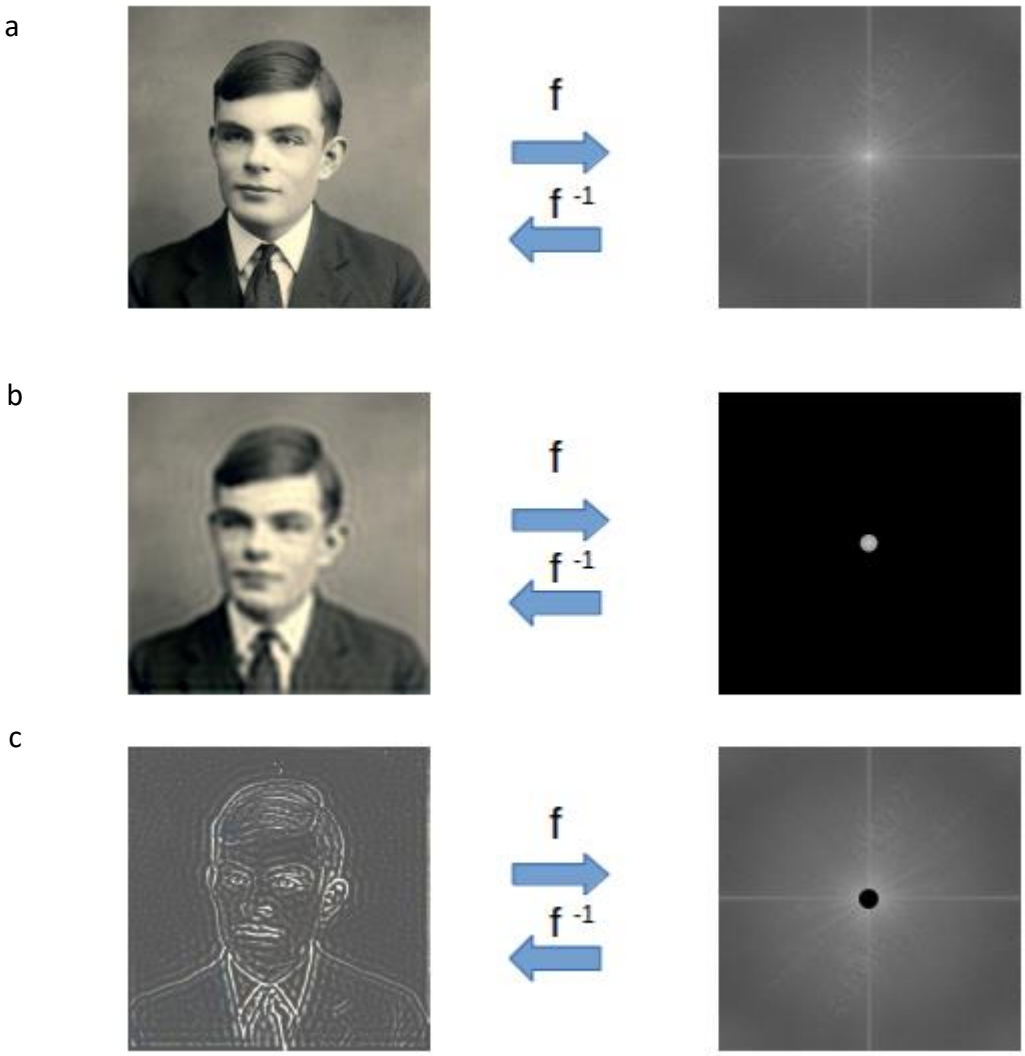


Figure 17 - (a) The 2D image of the Alan Turing's protrait (real space) and its 2D Fourier transform (2DFT) in the reciprocal space. The f indicates the application of the forward 2D Fourier transform. The same portrait after to apply a low-pass filter (b) and its corresponding 2DFT. In (c) it is shown the Touring's portrait after to apply a high-pass filter, and its corresponding 2DFT.
Source: By the author

In Figure 17a, it is shown an Alan Turing's portrait (left) and its respective Fourier transform (right). Considering the Fourier transform of the image, it is important to note that the low spatial frequencies are represented by those intensities closer to the zero-order frequency, which is the center of the reciprocal space. These frequencies preserve the information about the shape, size or in other words, the "larger" features of the image. On the other hand, the high spatial frequencies are represented by the intensity spots in the outer areas of the Fourier transform, and these are related to the "finer" or "smaller" details of the image. In Figure 17b, it is represented the original portrait after applying a low-pass filter, and its

corresponding Fourier transform (right). As it can be observed, the image only preserves its coarse characteristics. Figure17c shows an example of a high-pass filter in which the low spatial frequencies, closer to the center in the Fourier transform, were suppressed. As a result, it is difficult to observe in the recovered image the borders and limits, but it is conserved the fine details of the portrait. The image processing will be detailed systematically in the next sections.

3.1.1 Preparing the micrographs

After collecting the micrographs and converting to the appropriate format, it is necessary to prepare these images before starting the analysis. Usually, the very “large” details (i.e., density ramps) and the extremely “small” details (mostly noise) are hiding the motif -the particles- which the analysis is interesting in. The preparation of the micrographs consists in applying a filter to suppress, at the same time, these extreme frequencies in the reciprocal space, which are associated with unwanted information of the image. This operation is performed through a band-pass filter.

There are some aspects that should be considered when a band-pass filter is applied. The values of the low-pass (LP) and high-pass (HP) filters shall be chosen according to the particle size as well as the resolution that is expected to be achieved in the analysis. Also, the inversion of densities may be necessary depending on how the software understands the white and black levels as signal. In the case of IMAGIC 4D, it is necessary to have white particles on the gray background. For a common camera setup, if the micrographs are a Cryo-EM dataset, an inversion of densities should be performed. In the same conditions, if the micrographs came from negatively staining samples, invert densities will be not necessary.

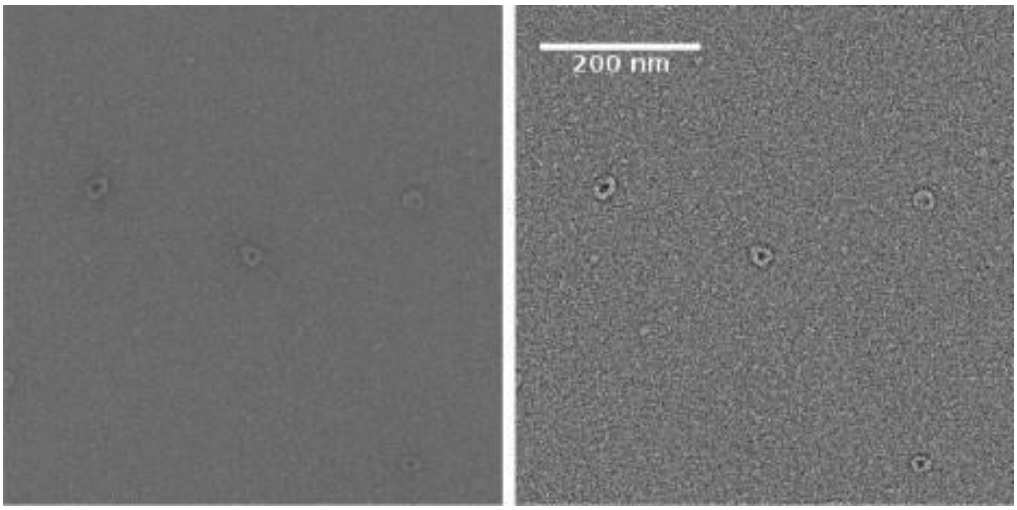


Figure 18 - A square patch of a negatively stained sample, of a protein embedded in a lipid membrane. The figure shows the micrograph before (left) and after (right) applying the band-pass filter. The pixel size is 2.58 Å
Source: By the author.

In Figure 18, it is shown a micrograph before (left) and after applying the corresponding band-pass filter (right). The filtered micrograph shows how the background information is reduced as well as the noise.

3.1.2 CTF correction

The contrast transfer function (CTF) is a mathematical representation that explains how the total phase shift of the electron wave, after interacting with the object, is translated into amplitude modulation in the image. For some spatial frequencies, the phase-shift will be “translated” as positive or negative amplitude modulation in the CTF. So, it is necessary to “recover” the information present in the negative spatial frequencies, inverting the amplitudes. Otherwise, the information at these spatial frequencies will contribute to the image formation in a “negative” way. The correction is performed by flipping the phases of the negative lobe of the CTF.[48,58,59]

The process begins with the calculation of the Fourier transform (its amplitudes) for each one of the micrographs (Figure 19A). Next, these amplitudes are classified following a strategy based on a multivariate statistical analysis (Section 3.1.4), generating class-averages of the amplitude spectra. Each of the generated class-averages, is an average of a certain number of the individual amplitudes of the micrographs. Therefore, the visualization of the Thon rings is

improved in the class-averages when compared to each one of its members (Figures 19A and 20A).

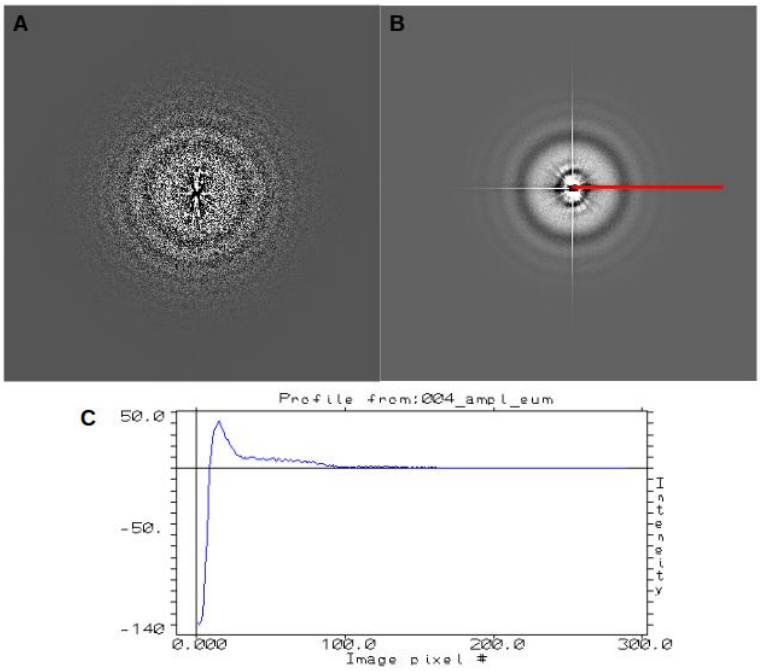


Figure 19 - Amplitude spectrum and CTF profile. (A) It is shown the amplitude spectrum calculated for one micrograph. (B) The average of all the amplitude spectra calculated for each one of the micrographs. The Thon rings are improved. (C) An amplitude variation profile was created along a central line of the average amplitude spectrum (red central line in B). This profile shows how the CTF is transmitting the information along the spatial frequencies.

Source: By the author.

Next, a ring-shaped mask is applied to each of the class-averages (Figure 20B). For this masking process, the internal and external radius of the mask will be selected according to the observed amplitude oscillation profile in a radial direction of the spectrum (Figure 19B, red line, and Figure 19C, profile plot). The selected region, according to the profile plot, is from the first zero-crossing of the contrast transfer function (CTF) until the region where no more oscillations are observed. Inside this region, theoretical CTFs will be generated, based on experimental parameters and estimating the defocus and astigmatism (defocus angle and intensity), for each class-average. Then, the estimated CTFs (Figure 20C, down right-half) are compared to the corresponding input experimental CTF (Figure 20C, up left-half) until the most similar is found. The estimated CTF parameters together with the experimental ones will be used to correct the images for the CTF. In the "half-half" images showed in figure 20a, the equivalence of the Thon rings can be used to check the accuracy of the CTF estimation, where the Thon rings of both

half should fit. Those class-averages showing the better adjustment are selected. Knowing the individual members (amplitudes) of the classes allows to extract the corresponding micrographs from the raw data and undergo a phase-flipping correction process with their own defocus and astigmatism values.

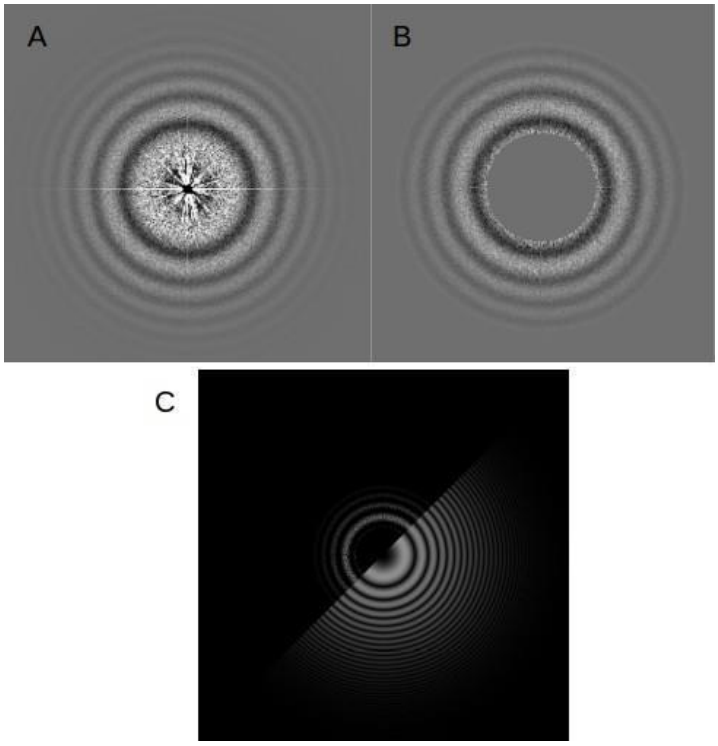


Figure 20 - (A) A class-average of some individual amplitude spectrum (C) A ring masked class-average, which will be then used to estimate the CTF, as well as the defocus and astigmatism direction (defocus angle). (D) The estimated CTF (down right-half) will be compared to the experimental or input CTF (up left-half), to then perform the CTF-flip process.
Source: By the author.

3.1.3 Particle Picking

Once the CTF flip is already done, the particle picking can be performed on the "corrected" micrographs. Particles can be selected in many ways. In the ABC methodology, it is suggested to use the local variance to identify and perform the particle picking, avoiding any bias in the process.[65] Particle picking can also be done by cross-correlation, using filtered projection images from a known model or, as an initial approach, a few rotationally averaged particles.[64] In the beginning, 3 to 6 different particles of different views are selected from some particular micrograph

and extracted in square boxes. Then these particles are centered, masked by a circular mask to reduce the influence of the neighboring particles and background, and finally rotationally averaged. In Figure 21, are shown an example of four selected particles (top row), masked and centered, and their corresponding rotational-averages (bottom row).

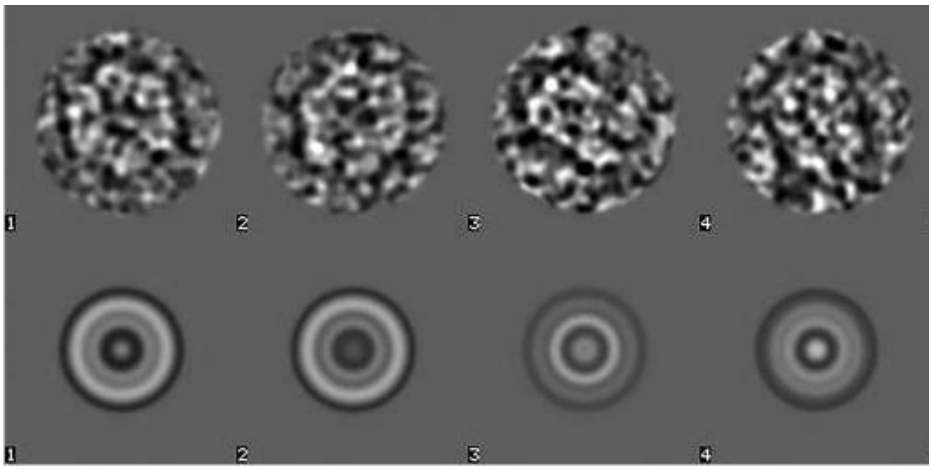


Figure 21 – Some selected particles (top row) and their corresponding rotational-averages image that will be then used as reference to perform a particle picking by cross-correlation
Source: Adapted from AFANASYEV.[64]

These rotational-average images are used as featureless templates to perform a particle picking by cross-correlation, on all the "corrected" micrographs. After this first round of particle picking, a lot of junk (thick carbon, ice, molecules aggregates, etc.) will be present in the particle data set. So, this undesired particles can be excluded based on histograms distribution of different criteria (e.g.: standard deviation, average density, cross-correlation coefficient, etc.).[64]

3.1.4 Multivariate statistical analysis (MSA): Hyperspace data compression and classification

A critical step in image analysis is to organize in the better viable way, all the information in the large particle data set. In this sense, we know that each particle is a 2D image that can be represented by a certain number of pixels being each pixel a specific value of density. Thus, if we consider that each particle is, say an image

of 100 x 100 pixels in size, then we have 10,000 values of density for each of these particles. Also, if our data set, has about 50,000 particles, all with the same size, then, each of these particles can be represented by a vector with 10,000 values in its array. It means that one would be dealing with a 50,000 x 10,000 matrix of pixels.

Alternatively, we can consider each image as a point in a multi-dimensional space, known as hyperspace, where the dimension is the same of the number of pixels in the image.[66] Since, in this hyperspace, each particle is represented by a point, those particles which are similar will be separated by shorter distances than those particles that are quite different. Unfortunately, it does not help much on the complexity of the problem since this hyperspace, following the example, would consist of 50,000 particles or points, a data cloud, represented in a 10,000-dimensional hyperspace. In Figure 22 A is represented the simplest case of a set of images, with two pixels each one. Therefore, the principal idea of the MSA, is to optimize the representation of the data cloud, through a principal component analysis of the data set, in such a way the new principal axis will coincide with the largest elongation of the cloud or the "greatest variance of the dataset". In Figure 22 B, these new axes are represented by green arrows, and the corresponding vectors (called eigenvectors) are indicated as "u", in the direction of the greatest variance of the dataset, and "v", in direction of the remaining greatest variance. This second direction represents only small modulations with respect to the main trend of the data set and may be ignored in this case. Therefore, in Figure 22, the points of the cloud can be represented in function of the main variance of the cloud using the vector “u”. This is the basic idea of the MSA: the hyperspace data compression. Hence, after performing the MSA, the huge number of coordinates is reduced to just few "principal" components, so the particles can now be represented as a linear combination of some "eigenvectors", which represent the most important variations in the data set, weighting them according to different "eigenvalues".[66,67]

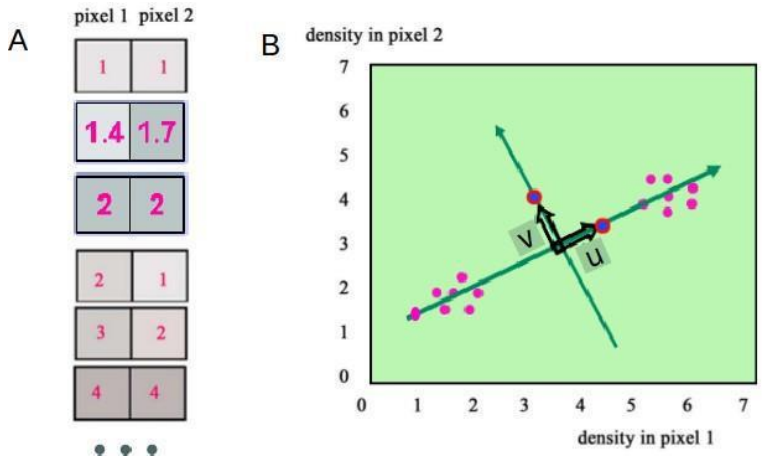


Figure 22 – Multivariate Statistical Analysis (MSA). A) Representation of the simplest case, of images with two pixels each one. Each number inside a pixel, represents a density value. B) Representation of each image in (A), in the hyperspace. Now, each image is represented as a point, according to its density values. The aim of the MSA, is to optimize the representation of the data cloud, through a rotation and translation of the coordinate system. Therefore, the new principal axis will coincide with "greatest variance of all dataset". These new axes are represented by green arrows, and the corresponding eigenvectors are presented as "u", in the direction of the greatest variance of all dataset, and "v", in direction of the remaining greatest variance. This last direction represents only small modulations with respect to the main trend of the data set and may be ignored. Therefore, the points of the cloud can be represented only in function of the main variance of the cloud, thus reducing the number of coordinate necessaries to define and image in the hyperspace.

Source: Adapted from VAN HEEL.[67]

After performing the MSA, the images, or the points in the hyperspace, which are sharing common features (for example, showing the particle in similar orientations) are grouped into "sub-clouds" or "classes" by an intra-class variance minimization, which is equivalent to the maximization of the inter-class variance. Once determined which particles belong to each class, an averaged of those particles will be performed, to result in a class-average, which will have a better signal to noise ratio (SNR) than their individual members. Commonly the number of classes-averages is chosen to have around few tens of particles per class, however, it will vary according to the quantity and quality of the data set.

3.1.5 Euler angular assignment

The class-averages or "*classums*" must be subjected to a process of assigning Euler angles to allow them to be available to the 3D reconstruction process. There are two methods of assigning Euler angles: angular reconstitution and projection

matching. The former is an unsupervised method, since it finds the intrinsic angular orientation between images based on their own information. The latter is a supervised method that uses projections of a given 3D model, generated in different spatial directions. Then, comparing these projections to the class-averages, the Euler angles can be assigned. This method when performed without the necessary caution may result in some bias induced by the model.[54]

In the current work, it will be used the angular reconstitution method, which is based on the “common line projection theorem". The underlying idea of this theorem is that, if there are two 2D-projections from a 3D volume, these share at least one 1D-projection, or a “common line”. Therefore, if one could idetify the common line between each pairwise of classes (from all dataset), then the Euler angles could be assigned for each class-average.[68,69]

In Figure 23, it is shown a graphic representation of how the angular-reconstitution process works. In the beginning, each 2D-class is projected into a 1D-line. Then, the class is rotated one degree (1°) in the plane, and again projected into a 1D-line, and so on over a 180° range. Therefore, all this set of 1D-projections from one specific class can be displayed as a single 2D-image known as a “sinogram” of the corresponding class (Figure 23: S1, S2, and S3). All sinograms are then compared pairwise in the following manner: the cross-correlation coefficient between each line of the first and each line of the second sinogram are computed. All correlation coefficients are then stored into a sinogram correlation function (SCF), that is represented as a 2D-image (CSC12 and CSC32). After to apply a fitting procedure over this SCF, the "common line" or "common tilt axis" between the two projections can be found (blue lines, common lines between projections 2 and 1, while red lines indicate the common lines between projections 2 and 3).[68,69] Also, the angular distance between the common lines (blue and red) gives the angle between projections 1 and 3. Hence, starting with a set of three projection images, it is possible to find the common tilt axis between each given pair of projections.

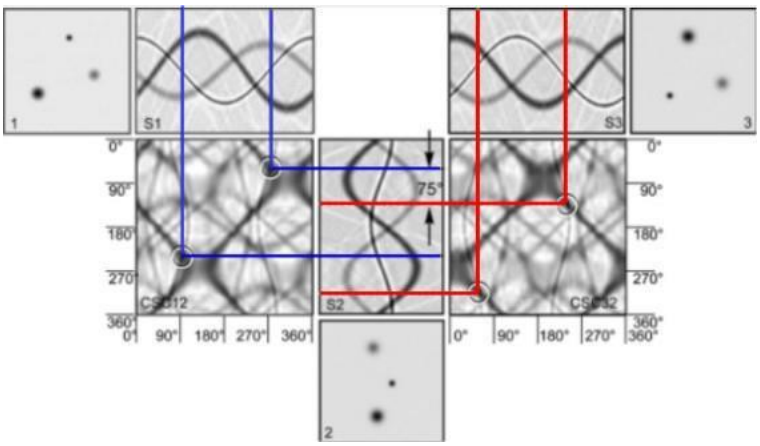


Figure 23 - Sinograms and sinogram correlation function for a model structure. There are shown three projections (#1, #2, #3), from a model composed of three Gaussian dots with different densities. The 1D projections set or sinograms are namely S1, S2 and S3, for each 2D projection respectively. CSC12 and CSC32 represent cross-sinogram correlation functions between projections 1 and 2 and projections 3 and 2, respectively. Each point of the sinogram correlation function contains the correlation coefficient of a pair of lines from the two sinograms. Blue solid lines indicate the common lines between projections 2 and 1, while the red solid lines indicate the common lines between projections 2 and 3. Each CSC has two peaks because projections from 180° to 360° mirror those from 0° to 180°. The angular distance between the common lines (blue and red) gives the angle between projections 1 and 3.

Source: Adapted from ORLOVA.[70]

3.1.6 Three-dimensional reconstruction

Once the Euler angles were already assigned to the different class-averages or "*classums*", they can be used for the three-dimensional reconstruction process. To carry out this task, the class-averages are back-projected into a 3D-space.[48,68] In Figure 24, it is shown a schematic representation of how the three-dimensional reconstruction is performed from the 2D-classes. After performing a 3D-reconstruction, this 3D map is re-projected into the same directions as the class-averages in a way that each class can be compared with its respective re-projection, and the similarity degree would give a measure of how good the angular assignment was performed.

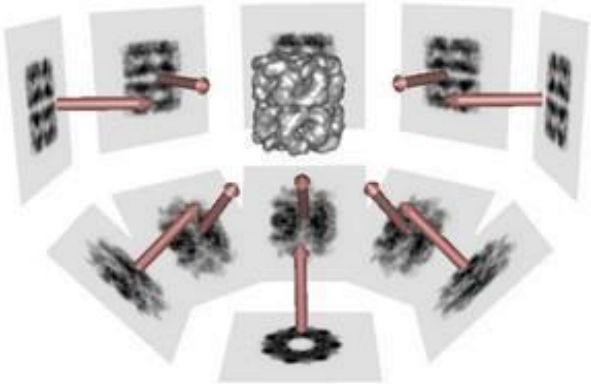


Figure 24 - The back-projection of the 2D class-average to produce a 3D density map. This 3D density map captures the electron density throughout the macromolecular complex.
Source: Adapted from MIT.[71]

The initial 3D map can be projected into different spatial directions to obtain a new set of projections, also known as "anchor-set", to be used as references, since they already have known angles, to refine the Euler angles assignment of the classes-averages. This is an example of 2D-projections set (“anchor-set”), that could be used to refine the previous angles of the classes-average. This process is iteratively performed until achieving a stable 3D map. All these steps will be detailed in section 5.

Another important aspect to mention is that the one can create a separate set of 2D-projections and use them as references to align the original particles. This process known as *multi-reference-alignment* (MRA),[48] works as follows: firstly, each particle will be compared by cross-correlation with each one of all projections. Then, the particles will be aligned (by displacements and rotation in the image plane) with respect to those projections they had shown to have the highest cross-correlation coefficient (CCC). After these process, the particles can be subjected for a new MSA process and re-classification.[64]

3.2 DNA production

A puC-lambda plasmid was used to produce, by Polymerase Chain reaction (PCR), the DNA fragment of 186 bp length, containing the wild-type (wt) λ PR promoter (-100 to +30). Below is the Promoter sequence as well as a representation of the consensus regions.

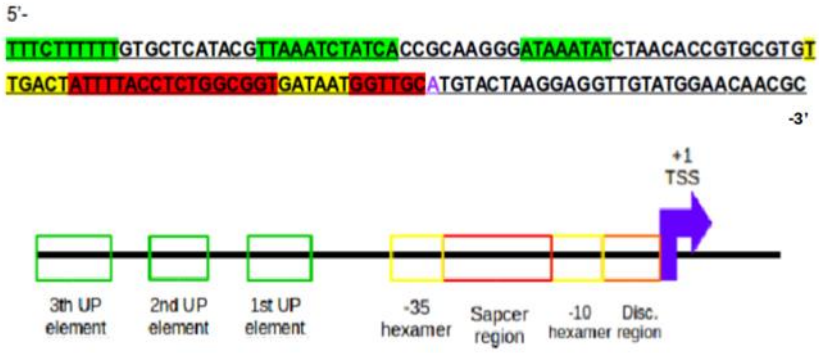


Figure 25 - Complete sequence of the DNA wild type Lambda PR promoter (in black letters and underlined). Sketch of the DNA promoter in which are remarked the principal regions of the DNA that interacts with the RNAP during the RP complex formation.
Source: By the author.

Figure 25 shows the complete sequence of DNA wt λPR promoter (-100 to +30), in black letters underlined, and with the transcription start site in purple letter. The colored shadows are explained in the promoter sketch (bottom). In this sketch, a purple arrow represent the transcription start site (TSS) and also by a yellow rectangle are represented the -10 hexamer region (from -7 to -12) and the -35 hexamer region (from -30 to -35). The discriminator region (from -1 to -6) is represented by an orange rectangle and the spacer zone (from -13 to -29) is represented by a red rectangle. In addition, the AT-rich regions, known as UP-elements, are represented by green rectangles: first UP-element (from -45 to -58), the second UP-element (from -68 to – 79) and the third UP-element (from -90 to -100). Once the DNA was amplified by PCR, the purification procedure started with Polyacrylamide Gel Electrophoresis (PAGE), which separates DNA by their base pairs length. After identifying the desired DNA band, it was purified by GeneJET PCR purification Kit (Thermo Fisher Scientific, USA), and its concentration was quantified.

3.3 Transcription Initiation complex assembly

To assembly the RP complex, the amplified DNA and the *E. coli* RNAP (New England Biolabs, USA) were used. The complex was prepared in transcription buffer TB5X (Tris-HCl 100 mM, pH 8.0, $MgCl_2$ 25 mM, KCl 250 mM), maintaining the RNAP:

DNA in a 1:2 molar ratio, and according to the protocol described in table IV. The protocol described here, which includes the corresponding binding test (Electrophoretic Mobility Shifting Assays), was widely reported in different previous studies of the same RNAP-DNA complex.[22,23,72,73]

Table 1 - Protocol for Transcription initiation complex assembly

Reagent	Ci	Cf	Volume (µL)
DNA	0.3 µM	60 nM	4
RNAP	1.0 µM	30 nM	0.6
TB 5X	5X	X	4
DTT [(a)]	10 mM	1 mM	2
H_2O			9.4
	Final Volume		20 [(b)]

Source: By the author.

(a) Dithiothreitol
(b) Once mixed, the mixture is incubated at 37°C for 20 minutes. After this, 2 µL of sodium heparin (2mg/mL) is added, and the mixture is incubated for 20 minutes at 25°C.

3.4 Specimen preparation and TEM imaging

An ultra-thin carbon coated grid (PELCO, USA) was used for this work. This grid was negatively charged (at 15 mA and 25 seconds) using an easiGlow™ Glow Discharge Cleaning System (PELCO, USA), for a better deposition of the sample on the grid. After that, 3 uL of the RP complex sample (with a concentration around ~ 0.02 mg/mL) were deposited on the grid, and after one minute, the excess of solution was removed using a filter paper. Immediately after, 3uL of uranyl acetate (2% m/v) was applied on the grid for the negatively stain the sample, and after 30 seconds the excess of solution was removed by a filter paper. The uranyl acetate solution was applied twice. Later, the micrographs were collected using the JEM 2100 microscope, (JEOL, Japan) equipped with an F-416 CMOS camera (TVIPS, Germany). Finally, 500 micrographs were collected using 200 kV acceleration

voltage, at 60,000x of magnification with a pixel size of 1.78 Å and in a constant defocus level of about -1.7 µm.

3.5 Single particle analysis

All the image processing was carried out using the Imagic4D software (Image Science, Germany) according to the most recent methodology reported in the literature for this package, the Align by Classification process.[64] It includes the various stages described in Figure 26, that were previously detailed in section 3.1. This began with the pre-treatment of the raw micrographs, to remove the extreme low and high spatial frequencies and it was followed by the CTF correction by phase-flipping. After that, a first particle picking was performed. These particles were boxed in 180x180 pixels images, then centered and circularly masked. These particles were then normalized and classified into ~2,500 class-averages using a *multivariate statistical analysis* (MSA) and a *hierarchical ascendant classification* (HAC).[67] The class-averages with extreme lower members per class or with low quality due to thick carbon, grid borders or overlapped particles, were removed and a better particle stack was obtained and re-classified. Once a good set of class-averages was obtained, the particles were centered and circular masked for the angular assignment and posteriorly three-dimensional reconstruction. This first 3D-map was used to refine the previous steps in the process until obtaining a consistent 3D final map. The resolution of the final 3D-map was calculated using Fourier-Shell-Correlation and the ½-bit criteria.[74,75]

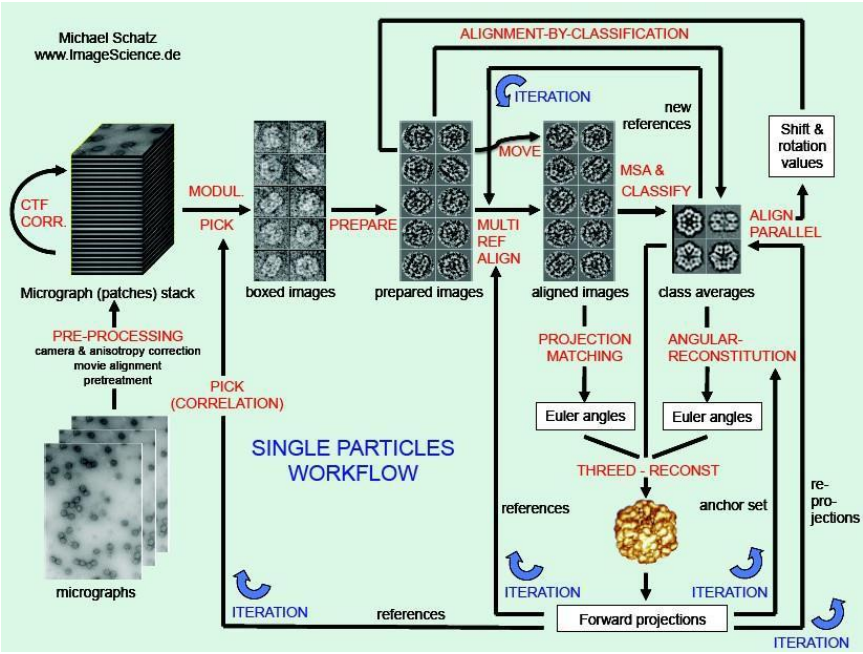


Figure 26 - Single particle analysis Cryo-EM workflow using IMAGIC 4D software Source: Adapted from IMAGIC-MANUALS.[76]

3.6 Modelling

For the molecular fitting of the RNAP into the 3D density map, different crystallographic structures were used. The RNAP holoenzyme and the DNA transcription bubble were fitted using the available crystal structure of the *E. coli* Transcription Initiation Complexes with a complete bubble, considering a promoter with 17 bp spacer.[6] The DNA sequence of the complete transcription bubble, until the -35 hexamer region, was replaced by the sequence of the promoter used in this study. The DNA region upstream the -35 hexamer toward -90, was modeled by segments using 3D-DART.[77] Then the DNA segments were manually fitted into the 3D map, following the strong observable density in it, and using Chimera package. The interaction of the αCTDs with the UP-element regions of the promoter was fitted according to the available structures of this domain in complex with CAP and a DNA region.[18] Finally, the N-terminal σ70 subunit, the domain 1.1, was taken from the structure reported by Bae *et al.* and manually fitted.[13] For all the rigid-body and manual fitting processes, it was used the UCSF Chimera 1.12 software.[78]

4 RESULTS AND DISCUSSIONS

4.1 Transcription Initiation Complex Assembly and TEM imaging

The DNA fragment, obtained by the process detailed in section 3.1, was used here to prepare the RP complex. In Figure 27, it can be observed an image of the Polyacrylamide Gel Electrophoresis (PAGE) of the DNA PCR.

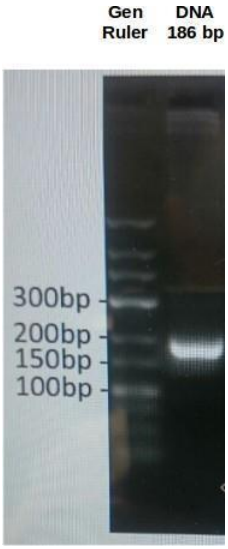


Figure 27 - Polyacrylamide Gel Electrophoresis (PAGE) of the amplified DNA fragment. In the image it can be observed the DNA fragment amplified by PCR, after to perform a PAGE assay. The first column of the image points the gen ruler, and at the right it is shown the DNA fragment of 186 bp length containing the λ PR wild-type promoter (-100 to +30).
Source: By the author.

The RP complex sample preparation, as well as the negatively stained specimens, were successfully produced according to the protocols previously explained in sections 3.2 and 3.3. The Figure 28 partially shows one of the micrographs at 60,000x magnification, with a pixel size of 1.78 Å. In this image, both, the protein and the DNA, have a high contrast. The RP complex formation (in red circles) and the free DNA (yellow rectangles) can be observed. The presence of the free DNA in the background is due to the 2:1 DNA:RNAP molar ratio used for the complex preparation.

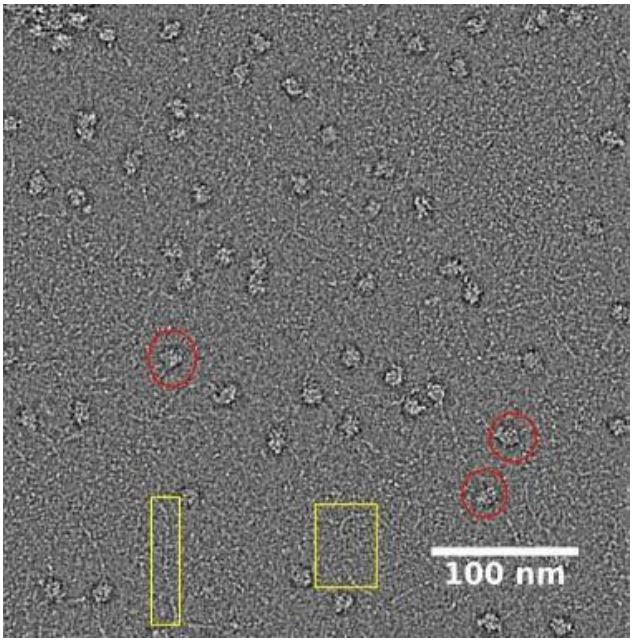


Figure 28 - A square patch of a micrograph obtained by TEM of the RP complex sample, negatively stained. Examples of RP complex particles are marked by a red circle and the free DNA by a yellow rectangle.
Source: By the author.

4.2 Image processing – Single particle analysis

4.2.1 Micrographs preparation and contrast transfer function correction

All the images collected were initially pre-treated applying a band-pass filter and a normalization of its densities values. The band-pass filter is a double Gaussian filter applied in Fourier space, as represented in Figure 29.

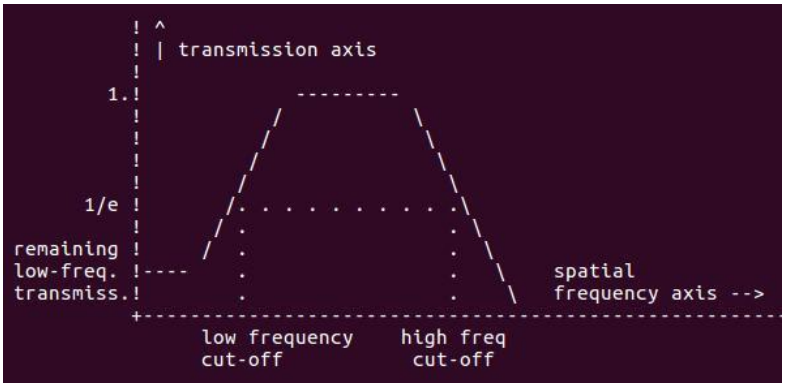


Figure 29 - Representation of the band-pass filter frequencies cut-off. The band-pass filter includes two cut- off values in the reciprocal space. The high frequency cut-off (HFC), associated with a LOW-PASS filter, and the low frequency cut-off (LFC), associated with a HIGH-PASS filter.
Source: By the author

The high frequency cut-off (HFC) is associated with a LOW-PASS Gaussian filter which will gradually cut off spatial frequencies higher than this value, meanwhile the low frequency cut-off (LFC) is associated with a HIGH-PASS Gaussian filter which will gradually cut off spatial the frequencies lower than this value. The remaining transmission allows us to leave a fraction of the low-frequency components intact. All of them are measured as a fraction of the Fourier space (from 0 to 1). As explained in section 3.1, to apply the cut-off parameters it is necessary to consider the particle size, pixel size and the expected resolution (all of them in angstroms). The pixel size is 1.78 Å and the longer axis of the particle was measured to be around 102 pixels (180 Å). Considering an expected final resolution of 15 Å, it was used a LFC of 0.02 and a HFC of 0.24. It is important to mention that, in this step, the density values in each one of all the 4,096 x 4,096 pixels micrographs were normalized to a given standard deviation and zero mean density. In Figure 30, it is shown two micrographs, before (left) and after (right) the pre-treatment process. It is observed that in the band-pass filtered image (right), the background signal and noise were damped, and the particles are easier to distinguish.

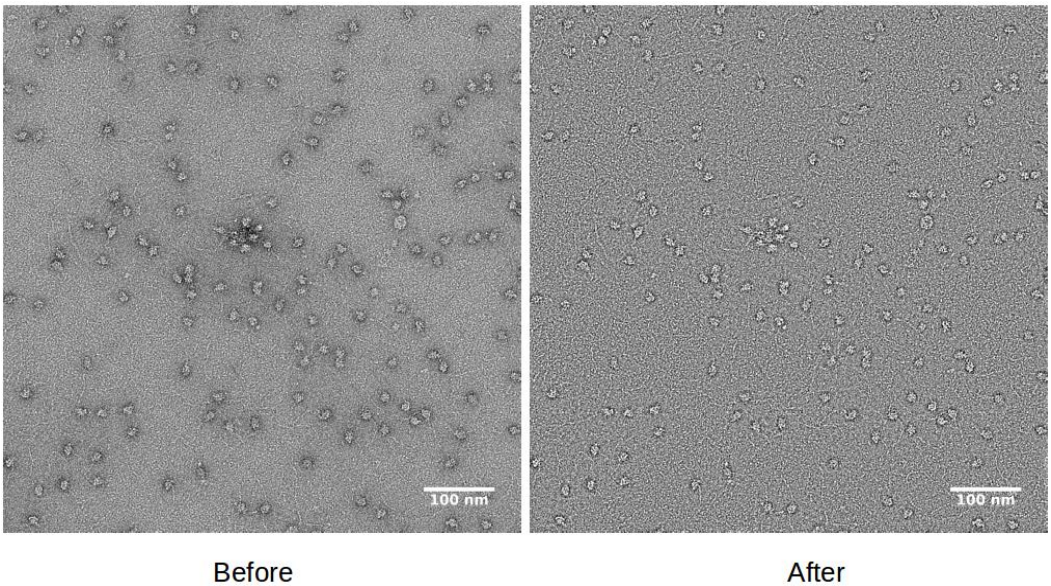


Figure 30 - A square patch of the same micrograph (one of the 500 collected for this work) before and after to perform the pre-treatment process.
Source: By the author

After the band-pass filtering and normalization of the micrographs, the CTF correction was performed to correct the transmitted information. Initially, the

amplitude spectrum was calculated for each micrograph (Figure 31A), followed by a band-pass filtering, to eliminate the background. After that, all of the amplitude spectra were averaged in a single spectrum (Figure 31B). An amplitude variation profile was created along a central line of the average spectrum to validate the filter parameters. This profile obtained shows how the CTF is transmitting the information along the spatial frequencies. As the CTF curve approach zero for the high frequencies, the filter parameters were chosen correctly (Figure 31C).

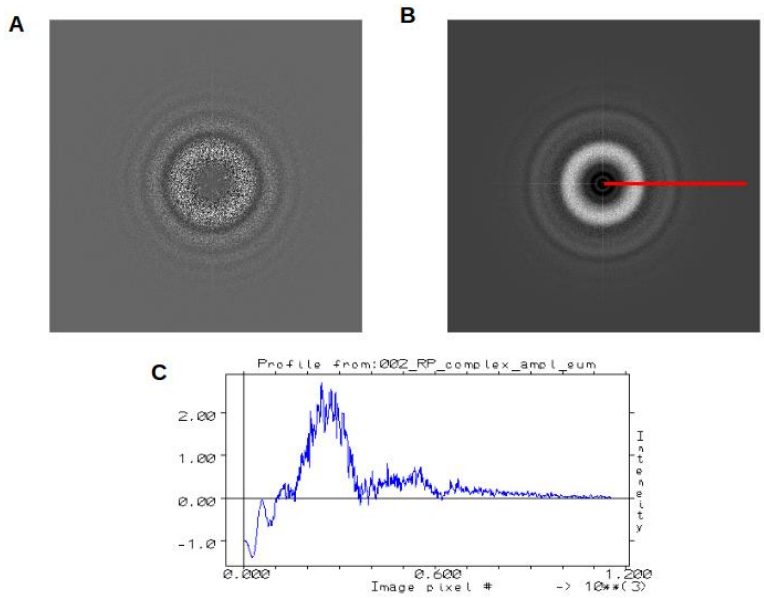


Figure 31 - Amplitude spectrum and CTF profile. (A) It is shown the amplitude spectrum calculated for one micrograph. (B) After to perform an average of all the amplitude spectrum calculated for each one of the micrographs the Thon-rings are easier to observe. (C) An amplitude variation profile was created along a central line of the average amplitude spectrum (red central line in B). This profile shows how the CTF is transmitting the information along the spatial frequencies.
Source: By the author.

The individual filtered amplitudes were then treated by multivariate statistical analysis (MSA), classified and masked by a cross-ring mask. In Figure 32A it is shown one of the class-averages obtained in this stage. Also, in Figure 32B, it can be observed the cross-ring mask applied on the class-averages. The mask showed in Figure 32B is a binary mask, in which the region in black has density values equal to zero, while the region in white indicates values of density equal to 1. Therefore, when

the software applies this mask over the images of a previously calculated amplitudes, all those density values that are in the same region as the "black" part of the mask will be zeroed, and will not be taken into account for the subsequent operations. Thus, an example of one masked class-averages is shown in Figure 32C. After that, as described in section 3.2, the theoretical CTF was estimated, as well as the defocus and astigmatism direction, for each class-average. This estimated CTF (Fig. 32B, down-right half) was compared with the corresponding input -experimental- CTF (Fig. 32B, up-left half). The classes with the better fitting were selected and the micrographs belonging to those selected classes were extracted from the dataset and subjected to a CTF-flip process, as described in section 3.2.

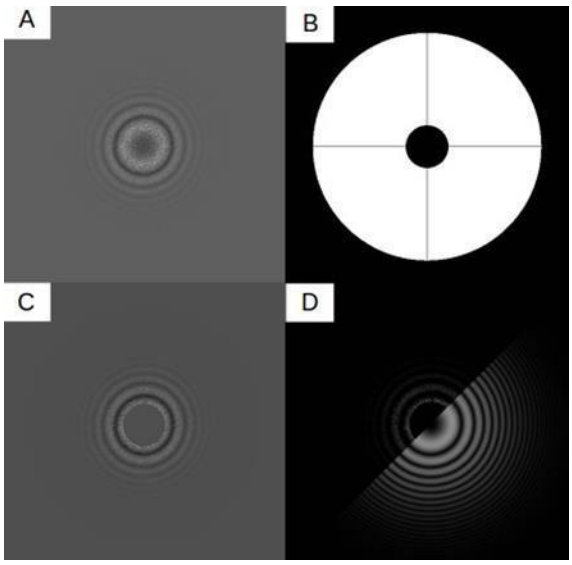


Figure 32 - (A) A class-average of some individual amplitude spectrums (B). A binary cross-ring-mask applied on each. (C) The masked class-average, which will be then used to estimate the CTF, as well as the the defocus and astigmatism direction (defocus angle).
(D) The estimated CTF (down right-half) will be compared to the experimental or input CTF (up left-half), to then perform the CTF- flip process.
Source: By the author.

4.2.2 Particle picking

Using the CTF-flipped or CTF-corrected micrographs, the particle picking by cross-correlation was performed. Around 100,000 particles were picked, using a few rotationally averaged particles as initial references, and boxed out in 180 x 180 pixels square images. Almost 40% of the total particles were artifacts (thick carbon, empty regions or overlapped particles).

So, after discarding these undesired elements, a final set of 60,393 particles was obtained. Figure 33A shows some of the particles obtained in this stage. These particles were normalized to a given standard deviation and zero mean density, taking into account all the density values inside a circular mask with a diameter of 70% of the box size. All density values outside this circular mask were zeroed. In Figure 33B is shown the same particle set at figure 42A, but now with a circular mask applied.

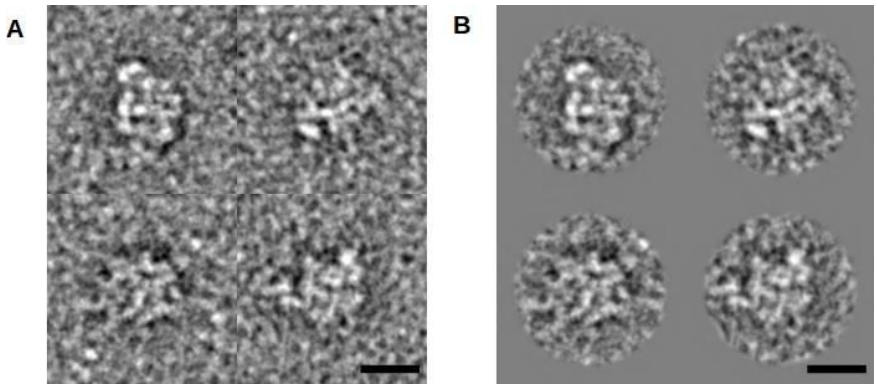


Figure 33 - A) Some of the 60, 393 particles selected from the micrographs, each one with a size of 180 x 180 pixels, with 1.78Å per pixel. B) These same particles, after to be circular masked and normalized. All the density values outside the circular mask were zeroed. The scale bar is 10 nm in size,
Source: By the author.

4.2.3 Multivariate Statistical Analysis and classification

The previous normalized and masked particles were centered. A hyperspace representation of the data, based on Principal Component Analysis (PCA), was performed (“MSA-RUN” in IMAGIC package). As a result, each one of the particles can be described as a linear combination of its principal components. Since it is an *eigenvector-eigenvalue* analysis, the principal components are also referred as *eingenvectors* or *eigenimages*. Figure 34 shows some of the 69 eigenimages generated by the MSA-RUN. The first image represents the “center” or “average” of all the *eigenimages* in the compressed hyperspace. The second *eigenimage* represents an axis which describes most of the variance of the data set. The third image represents the axis which describes the remaining variance, perpendicular to the first one, and so on.

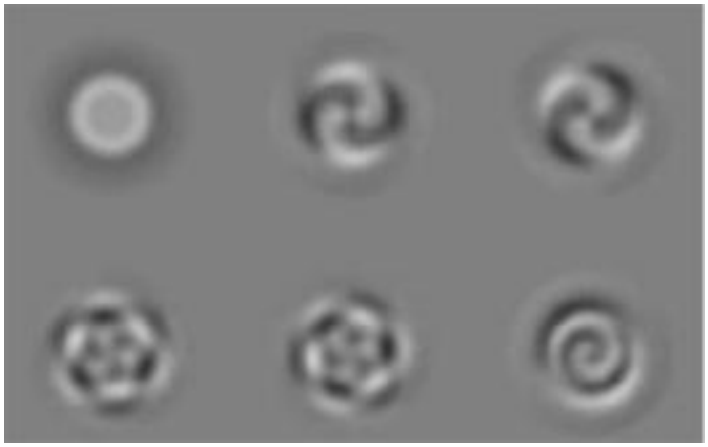


Figure 34 - Some eigen-images obtained in the first round of the SPA. The first image (top left corner) represents the “center” or “average” of all the *eigenimages* in the compressed hyperspace. The second *eigenimage* represents an axis which describes the greatest variance of the all data set. The third image represents the axis which describes the remaining greatest variance, perpendicular to the first one, and so on.
Source: By the author.

After representing the images in the hyperspace, the dataset was subjected to a classification (“MSA-CLASSIFY” in IMAGIC package) by the hierarchical ascendant classification method (HAC) combined with k-means. HAC seeks to minimize the intra-class variance while maximizing inter-class variance. In other words, the similar images are grouped into clusters. Finally, the particles belonging to each one of these clusters were averaged (“MSA-SUM” in IMAGIC package), resulting in the formation of class-averages, also known as *classums*. Therefore, the original 60,393 particles were classified into 2,500 classes, resulting in about 24 particles per class. In the resulted *classums*, the signal/noise ratio is clearly improved in comparison to the individual members. Therefore, it is easier to observe different orientations of the complex, and to identify those bad classes with a few particles or containing some sort of artifacts (thick carbon, protein aggregates or empty grid regions). Bad classes were discarded at the same time as its members were excluded from the dataset. Iterative cycles of MSA- RUN, MSA-CLASSIFY, and MSA-SUM were performed to obtain better class-averages together with eliminating bad particles from the dataset. Figure 35 shows some of the first class-average obtained.

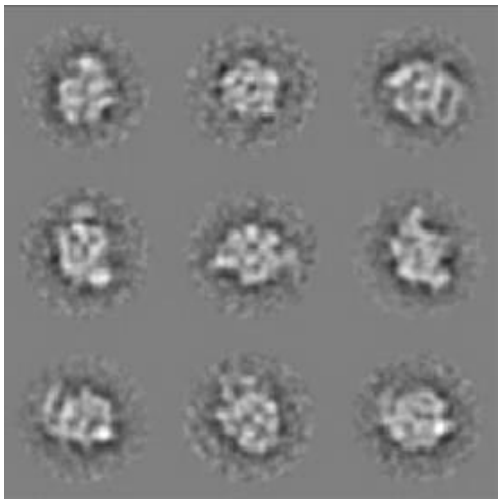


Figure 35 - Some of the first classes-averages, or *classums*, obtained from the image analysis. Each *classum* is obtained as an average of those particles which were classified inside a common “cluster”, during the classification in the hyperspace. The better Signal to Noise Ratio (SNR) of the class-averages when compared to individual particles, make it easier to observe the different orientations of the object under study.

Source: By the author.

4.2.4 Angular reconstitution

For the next step in the image analysis, 50 class-averages were selected from the data set to perform the first angular reconstitution (“ANG-REC” in IMAGIC package). This selection was based in their quality and the different orientations displayed. To find their spatial orientation, or the **Euler angles**, these *classums* were initially centered and circularly masked to obtain more accurate results.

The angular reconstitution was carried out in a “random start-up” approach, assuming it as a non-symmetrical object (C1 symmetry). This process works as follows: given a set of *classums* with unknown Euler angles, these angles will be assigned randomly and then refined for each *classums*, comparing one-by-one with the reprojections of the reconstructed map. This process is performed iteratively during an arbitrary number of cycles, until achieving convergence in which the Euler angles do not change any more. For this analysis, angular reconstitution / random start-up was made for 25 cycles of refinement using a mask of 0.65 (a bit lower than the circular mask on the *classums*). The mask is used to define the area within the Euler angles will be calculated. In Figure 36 it is shown some of the class-averages and below each one, their respective

Euler angles (α/β/γ). It is important to mention that an error associated with the angular reconstitution process ("AR-error"), is also calculated for each class. This AR-error is useful for discarding some class-averages during the refinement cycles

.

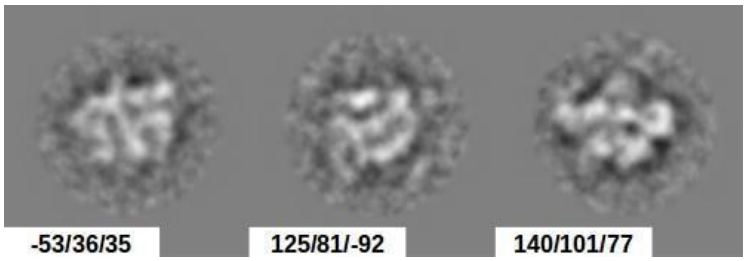


Figure 36 - Angular reconstitution: here it is shown some of the first class-averages obtained, and their respective Euler angles (white panel below each image, α/β/γ) after performing the angular reconstitution process. The Euler angles indicate the spatial orientation of these class-averages that will be used to perform the three-dimensional reconstruction.
Source: By the author.

4.2.5 Three-dimensional reconstruction

The next step was to carry out the first three-dimensional reconstruction ("THREED-REC" in IMAGIC package) using the class-average whose Euler angles have already been assigned. For the 3D reconstruction, some parameters like molecule symmetry and circular mask radius of the 3D volume must be set-up. For this case, the C1 symmetry and a 60% circular mask were used.

Once reconstructed, the 3D volume is automatically re-projected in the same directions as the input *classums*, creating an output file of "re-projections". The average similarity between the input classes and their corresponding re-projection images is defined as the "3D-error" and it is a measure of the quality of the reconstruction. Also, this 3d-error is a useful information to improve the analysis of next steps. In Figure 37A it is shown at the top row, three of the class-average (their locations numbers are below in white squares: 35, 36 and 37 from left to right) used for this first 3D reconstruction. In the bottom row, there are their respective re-projections. The 3D-error for each class-average (marked by a yellow rectangle) is shown in Figure 37B. In addition, in figure 46B it can be observed

the three Euler angles assigned for each class-average as well as the angular reconstitution error (AR-error).

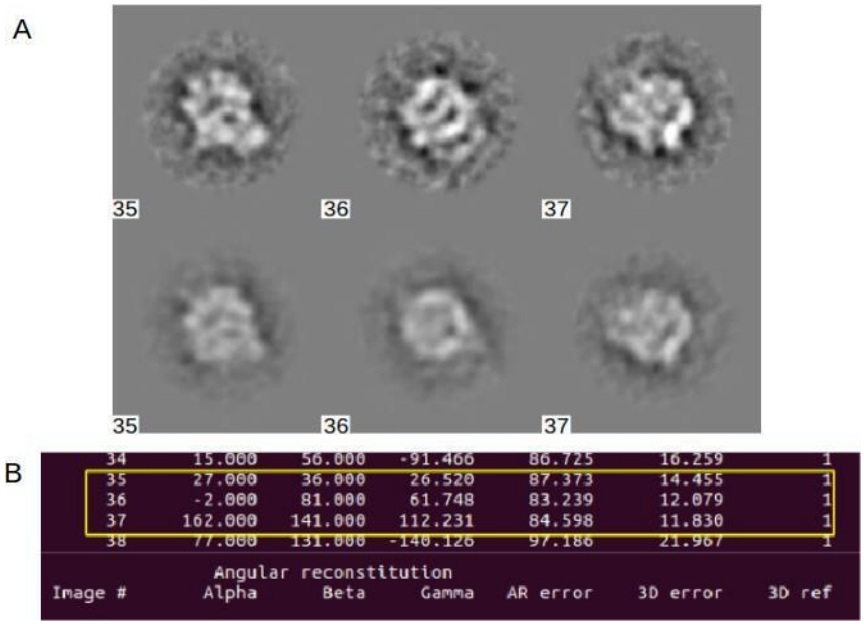


Figure 37 – A) Some class-averages (top row) and their respective reprojections (bottom row), these last obtained by project the 3D-model in the same direction than the input class-average. B) A part of a summarize report with the results of this three-dimensional reconstruction round. In a yellow rectangle, is remarked the parameters corresponding to the class-averages in (A). Here it is shown the three Euler angles assigned to each class-averages, as well as the AR-error and the corresponding 3D-error.

Source: By the author.

In Figure 38, it is shown the very first 3D reconstruction obtained in this image processing (left). The resolution of this 3D model was assessed using a Fourier-Shell-Correlation (FSC) and ½-bit criterium (Figure 38, right). The FSC plot shows that the obtained resolution is ~34 Å, which is also an indicator of the low overall quality for this first reconstruction. Thus, further refinement steps were carried out to obtain a better preliminary 3D-model and these will be explained in the next sections.

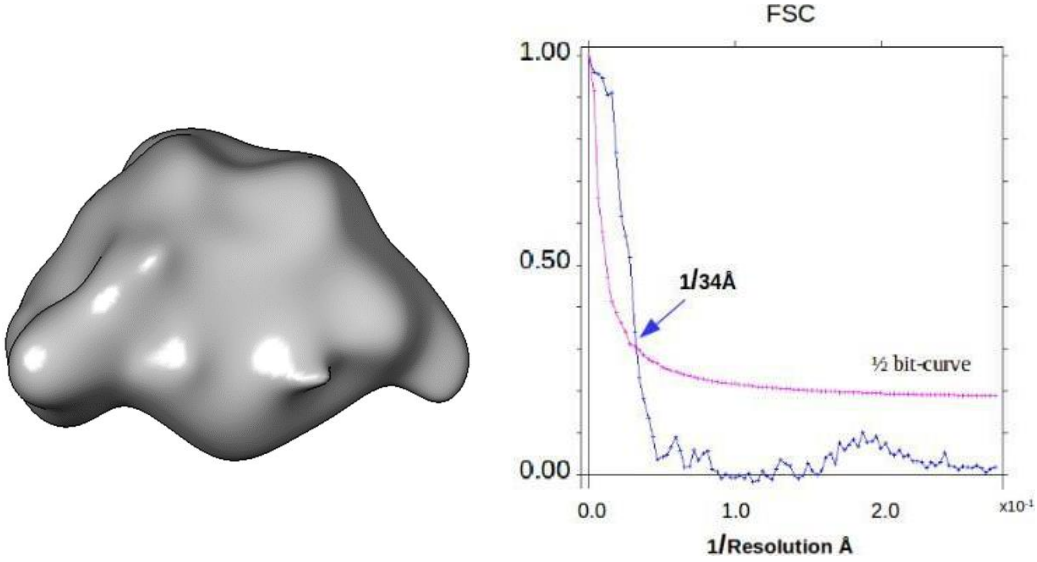


Figure 38 – The first 3D model obtained in this round (left), and the corresponding FSC vs ½-bit curve plot (right), with which the resolution of this model was estimated at 34 Å.
Source: By the author.

4.2.6 Iterative refinement loop

At this stage, to continue improving the quality of the 3D-model, each class-average was aligned with its respective re-projection. In this process, the class-average is translationally and rotationally aligned in the image plane relative to its respective 3D-reprojection, which is called “parallel alignment” or ALI-PARALLEL according to the IMAGIC package. Therefore, after performing another reconstruction using the aligned classes the 3D model is improved.

However, in these firsts processing cycles, some classes maintain a high 3D-error and must be temporarily excluded. After a visual inspection, those classes that were too dissimilar in respect to their respective re-projections were excluded from the dataset. After that, new cycles of reconstruction and parallel alignment to re-projections were carried out, until achieving stability in the 3D reconstruction.

Once a consistent, although preliminary, 3D model was obtained, it was projected in different directions, obtaining a set of 2D-images (“forward-projections”) that were used as references (“anchor-set”) to perform a new angular assignment cycle on those classes-averages previously aligned to its re-projections. One important aspect after to perform the iterative parallel alignment and angular assignment cycles on the class-averages is that these can be also used as references to align each one of its member particles, using the same parameters

assigned to the class-average. Therefore, all the translational and rotational displacements that a class-average has suffered will be "transferred" to each of its members. Then, those “aligned” particles can be again classified and averaged to obtain new and better classes until achieving stability in their quality. This process is known as Alignment-by-Classification (ABC).[64]

It is important to mention that the number of forward-projections created, as well as the angular projection range and the angular distance between projections, are parameters to be chosen according to the symmetry of the complex and the expected resolution of the analysis. For example, for highly symmetric objects, like a viral capsid, a few (5-10) forward-projections would be enough. Also, if a high-resolution study analysis is being carried out, there will be necessary to keep a minimal angular distance between projections of a few degrees (~3–5 degrees). A structural study performed on negatively stained samples of non-symmetric structures, as the RP complex, will require projections with a higher angular distance between them, since the expected resolution would be limited to ~20 Å.

Another important aspect in this stage is that the forward projections were always treated by a high-cut-off filter before being used as references for a new angular reconstitution cycle. In this context, a high-cut-off filter in the reciprocal space consists in completely removing all the spatial frequency higher than a certain value. For example, if the resolution of the first 3D model, was calculated to be ~30 Å, then the forward-projections were high-cutoff filtered at 35 Å. Thus, it means none of the projections, will have any information higher than ~35 Å avoiding any bias in the process. After performing new iterative cycles, if the 3D model quality and its resolution show an improvement, it is clearly due to the information present in the data, and not from any miss-correlation of high frequencies. Throughout the refinement cycles, the 3D model resolution is calculated, so the high-cut-off frequency is adjusted for each new stage. The Figure 39 shows a schematic representation of the refinement cycle loop that was carried out until achieving convergence in the quality of this "first" three-dimensional reconstruction.

.

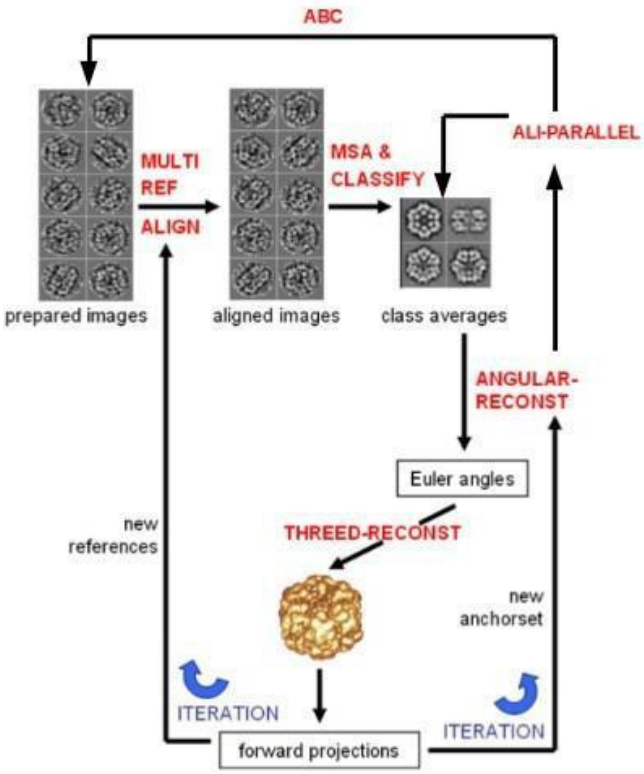


Figure 39 - In this diagram, it is shown the different steps in the refinement cycle. It begins with the MSA and classification step. After that, the angular reconstitution is performed on the class-averages, follow by the three-dimensional reconstruction. From this 3D model, a set of reprojections is created, in the same directions than the input class-averages that can be used to perform the parallel- alignment (ALI-PARALLEL) of these class-averages. Also, another set of projections (forward- projections) can be created to refine the Euler angles of the class-averages. Moreover, those class- averaged that have been subjected to ALI-PARALLEL and angular refinement, can be used to align the particles (ABC) and carry out a new MSA and classification stage.
Source: By the author.

Once the convergence in this first three-dimensional reconstruction is achieved, this 3D model was projected to generate an "anchor-set" to assign new Euler angles for all the class-average obtained at the beginning of the image processing. After that, another cycle of 3D-reconstruction and all the steps of the loop refinement were performed again, until getting a convergence of the quality for this new 3D model. An important aspect to point out is that the steps of the refinement loop were executed the necessary number of times needed to achieve a stable result. In the case of the alignment process, the cycles were performed three times, since the last two cycles did not show a significant improvement in the quality of the class-averages. On the other hand, the refinement loop, that includes alignments, angular reconstitution and 3D reconstruction, was performed several times.

In Figure 40 it is shown some of the class-averages, and their re-projections, that were used in the final cycle for this stage. It is observed the improvement in the quality of each class-average and the better matching with its re-projections, when compared to the first stages of the image processing.

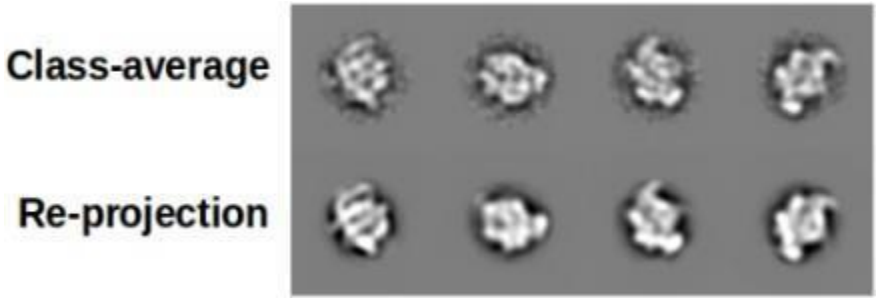


Figure 40 - Some class-averages (top row), used for the final three-dimensional reconstruction of this stage, and their respective reprojections generated (bottom row). It is clearly observed the improvement in the quality of each class-average and the better matching with its reprojections, than in the first stages of the image processing.
Source: By the author

In Figure 41 is shown one view of the final 3D model obtained (left) in this stage. The FSC plot (right) shows how the resolution achieved has a significant improvement, raising from ~34 Å for the first three-dimensional reconstruction to ~20 Å for this model. From this 3D model, a set of 20 reprojections was created, and its mirror images, and were filtered at 30 Å. This set was used to perform a new particle picking in the CTF-corrected micrographs. Also, a set of 100 2D-projections, and its mirror images, was created, with the aim to serve as references to perform an MRA process on the new picked particles. Finally, also using this 3D model, an anchor-set of 30 projections was created, to serve as references to assign Euler angles in the later classes.

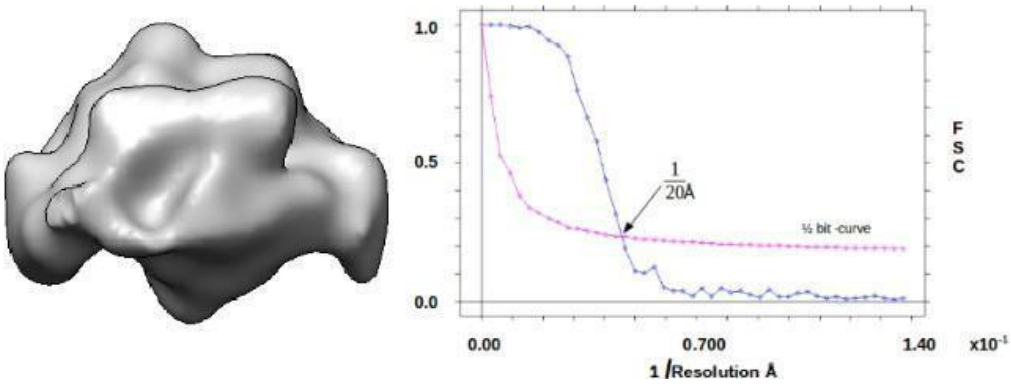


Figure 41 – A view of the final 3D model in this stage of the image processing (left). The FSC vs ½-bit curve to estimate the resolution of the 3D-model (right).
Source: By the author

4.2.7 Final 3D reconstruction

By cross-correlation and using the filtered reprojections as references, about 57,000 particles were picked from the micrographs. According to the SPA workflow explained before, new classes were generated and all the steps, already discussed, were performed iteratively.

Finally, the last 3D-model was generated from 16,015 particles, grouped into 666 class-averages. An average 3D-error of ~ 11% was obtained in this last round, and the resolution of the model was estimated in 20 Å, by FSC with 1/2-bit criterium. In Figure 42, it is shown the obtained results. In Figure 42A, is shown the 3D-model and in Figure 42B are shown some of the class-averages (top row) used in the three-dimensional reconstruction, as well as their respective re-projections. Finally, in Figure 42C, it is shown the FSC vs ½-bit curve plot, in which the resolution was estimated at 20 Å.

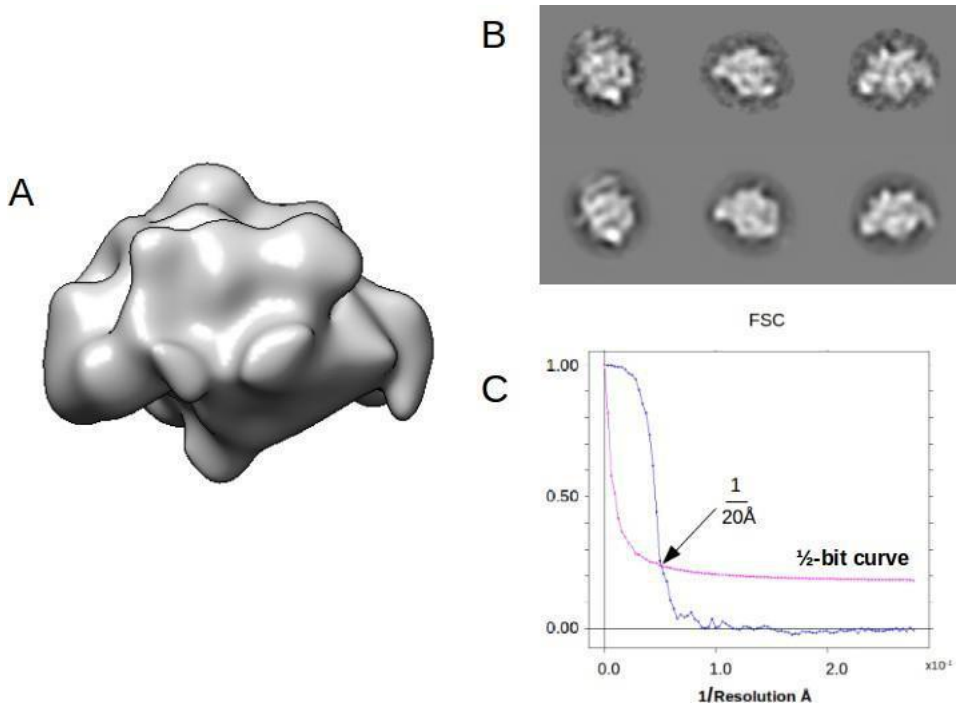


Figure 42 - (A) The last 3D-model obtained in this work. (B) Some class-averages used for this final three- dimensional reconstruction, and their respective reprojections obtained. (C) The FSC vs ½-bit curve plot, in which is pointed that the resolution of this model was estimated at 20 Å

Source: By the author.

4.3 Model fitting

4.3.1 RNA Polymerase rigid body fitting and DNA modeling

To obtain a coordinate model of the RP complex, different rigid-body and manual fitting were performed. In the beginning, the coordinates of the crystallographic structure from the RNAP holoenzyme and the DNA transcription bubble (PDB 4YLN) were fitted into the 3D-density map. This rigid-body fitting was performed using the Fit-in-map tool of the UCSF Chimera software which was also used for the further manually fitting steps. Also, some domains of the RNAP, like: the β insert 4 (βi4, residues 225-343), also known as β dispensable region 1 (βDR1); insert 9 (βi9, residues 938-1042) and the β' insert 6 (β'i6, residues 942-1129), were manually adjusted into the density map after the initial rigid-body fitting. The DNA region upstream of the -35 hexamer until the 2nd UP-element (-80 bp), was modeled by segments using the 3D-DART on-line server and then manually

fitted into the 3D map following the strong observable DNA density. The interaction of the αCTDs with the UP-element regions of the promoter was manually fitted, as well as the $\sigma_{1.1}$ domain. All this process was performed as was explained in section 3.6.

4.3.2 Structure of the Transcription Initiation complex with wrapped promoter

In figure 43A, it is shown one view of the 3D density map obtained in this study. In 43B, it is shown the PDB 4YLN coordinate-model of the RNAP, fitted in the map. The general shape of the RNAP structure can be observed, with densities associated to some β and β' domains, σNCR, αNTD-dimer (N-terminal domain-dimmer) and ω subunit. The color code is as following: β subunit in purple, β' in green, ω in orange, α^{I} in blue, α^{II} in cyan, σ in yellow, all of them in ribbon view. In Figure 43C, there are shown new coordinates fitted, which positions were located according to the 3D density map solved here. It can be observed the full upstream promoter region (from -36 to -79, in pink spheres). Also, the promoter region which is downstream the transcription bubble (from +12 to +18) is shown inside the main channel of the RNAP. Moreover, the densities corresponding to the αCTDs domains (in blue and cyan ribbon) can be observed. In addition, our model suggests that the β'i6 domain would appear folded in a closed conformation, performing a movement to be oriented toward the βi4 domain, which results in closing the gap between the β'jaw and β lobe/i4 domains. Therefore, this observation could indicate that these DME and βlobe/Bi4 would be regulating the grip of the grab craw on the downstream DNA promoter, as suggested before.[79,80,81] This observation is similar to those described in the crystallographic structure of the RP complex reported by Narayanan *et al*.[9] In Figure 43D it can be observed the full pseudo-atomic coordinate model, fitted in the 3D-reconstruction. The domains previously known (PDB 4YLN) are indicated with black text, and the new elements, domains and DNA, by green and red text.

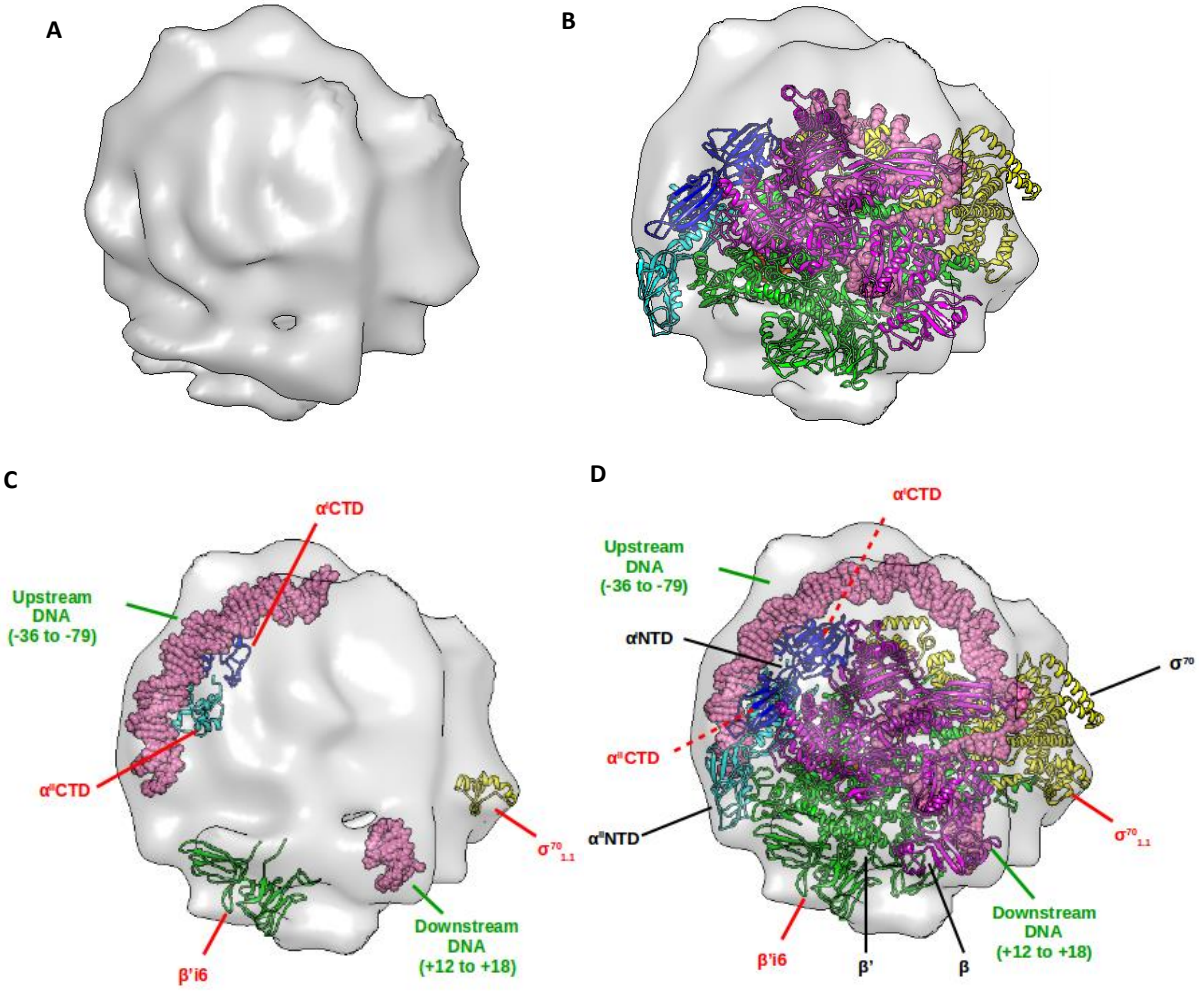


Figure 43 - (A) 3D-reconstruction obtained in this study. (B) RNAP PDB (4YLN), fitted in the density map. It can be observed the general shape of the RNAP structure with densities that belong to some β and β' domains, σNCR, αNTD-dimer (N-terminal domain-dimmer) and ω subunit. The color code is as following: β subunit in purple, β' in green, ω in orange, α^{I} in blue, α^{II} in cyan, σ in yellow and the DNA promoter in pink spheres. (C) New domains and DNA fitted in the 3D density map. It can be observed the full upstream promoter region (from -36 to -79, in pink spheres), as well as the promoter region which is downstream the transcription bubble (from +12 to +18). Moreover, it can be observed the densities, which correspond to the αCTDs domains (in blue and cyan ribbon). In addition, our model suggests that the β'i6 domain would be folded (green ribbon) and the $\sigma_{1.1}$ domain (yellow ribbon) would be outside of the main channel. (D) It is shown the full pseudo-atomic coordinate model fitted in the 3D-reconstruction. The domains previously known (PDB 4YLN) are indicated with black text, and the new elements, domains and DNA, by green and red text. The color code is the same used in (B). It is observed the overall extension of the wrapped promoter, and the domains of the RNAP that could be interacting with the promoter.

Source: By the author.

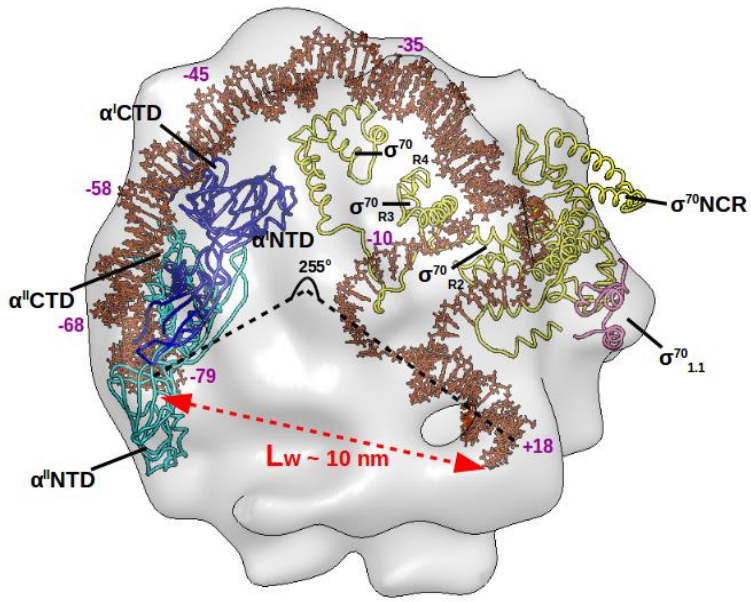


Figure 44 –Overall DNA wrapping extension around the RNAP. The β, β' and ω subunits were hidden. Here it can be observed the domains of the σ subunit that would be interacting with the promoter to form the DNA wrapping. The σ_{R4}, σ_{R3}, σ_{R2}, would be oriented toward interacting with the -35 hexamer, -10 hexamer and the transcription bubble, respectively. Also, the αCTD of each α subunit (I and II), would be oriented to interact with the first UP-element (-45 to -58) and the second UP-element (-68 to -79) in the promoter, therefore making possible the formation of the DNA wrapping. DNA wrapping would span from +18 to -79, thus involving approximately 100 bp or ~32 nm in length. Purple numbers mark the positions in the promoter. The DNA is in ball and sticks representation, in red color, except for the UP-elements, which are colored in green.

Source: By the author.

As it can be seen in Figure 44, the model suggests that the αCTD of each α subunit (I and II), would be independently oriented to interact with the first and second UP-elements in the promoter (colored in green ball and sticks), enabling the formation of the DNA wrapping. Also, according to our model, the DNA wrapping spans from +18 to -79 positions of the promoter, therefore the total wrapping length is around 32 nm. Moreover, it can be observed that the overall bent angle of the wrapped DNA is ~ 255°.

In Figure 45, it is shown a close-up view of the interaction between the α^{I}CTD and the first UP-element of the promoter. This interaction would be through the R265 and K298 (in spheres) present in the Helix-hairpin-Helix (HhH) motif of the αCTD. These residues are oriented in a way to interact with the phosphates in the minor-groove of the DNA, as previously reported by Benoff *et al*.[18] An interesting feature of this model is that the first UP-element seems to exhibit an

abrupt bend (~ 125°) of DNA, possible after its interaction with the α^{I}CTD. Thus, this interaction with the first UP- element (-45 to -58), would be indicating the beginning of promoter wrapping.

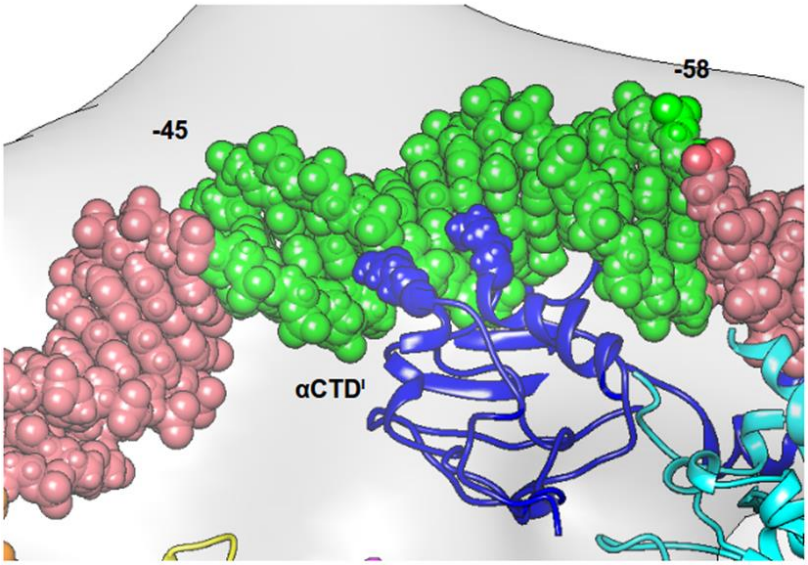


Figure 45 - A close-up view of the interaction between the α^{I}CTD and the first UP-element (-45 to -58) of the promoter. This interaction would be through the R265 and K298 (in spheres) present in the Helix-hairpin-Helix (HhH) motif of the α^{I}CTD.
Source: By the author.

In an analogous way, in Figure 46, it is shown a close-up view of the α^{II}CTD oriented toward the 2° UP-element of the promoter (-70 to -79). Like the previous case, this α^{II}CTD would be interacting with the DNA minor groove, through its R265 and L298 residues of the HhH motif, according to structure previously reported by Benoff *et al.*[18] It can be also observed that the DNA bending generated due to the interaction of this second UP-element with the α^{II}CTD, is smoother (~ 60°) than the one occurred in the first UP-element. An important aspect to mention is that the DNA was modelled as an overall B-DNA, since it is the most probably DNA structure of the system studied here. However, due to the strong bent of the DNA due to the wrapping, it is possible that local variations of the DNA structure would be generated, like A-DNA or even Z-DNA, as a local conformation of an overall behavior of the promotor. Nonetheless, in the current study, we do not have enough resolution to ensure or discard these possibilities.

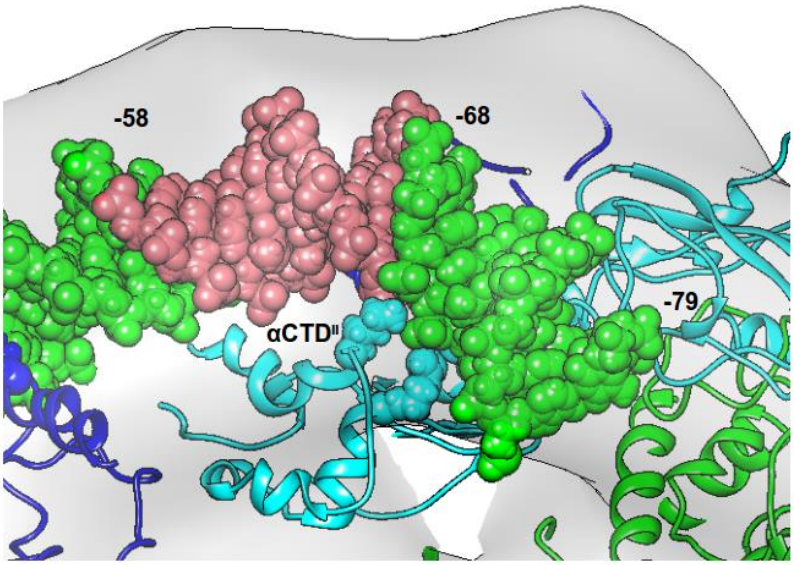


Figure 46 - A close-up view of the α^{II}CTD oriented toward the 2° UP-element of the promoter (-68 to -79). Similar to the previous case, this α^{II}CTD would be interacting with the DNA minor groove, through its R265 and L298 residues of the HhH motif (in cyan spheres), according to a structure previously reported.
Source: By the author

It was not observed a traceable density that could be associated to the third UP-element (- 90 to -100). This is possibly because this region is not interacting with the RNAP in a way to affect the wrapping formation. This hypothesis is supported by optical tweezers studies performed on this same RP complex, which reported that when using mutant λ PR promoter, in which the region from -80 to -100 was replaced by heterologous DNA, the wrapping length was exactly the same as the wild-type promoter.[72] Also, when the region from -60 to -100 was replaced by heterologous DNA, the wrapping length was reduced in ~ 8 nm comparing to the wt promoter (Sosa, personal communication). Thus, those observations reinforce that the α^{II}CTD is possibly interacting with the promoter in a region around the second UP-element (-70 to -79). Another hypothesis would be that the enzyme could interact with this third UP-element, in a transient conformational state during the formation of the RPo.

4.4 Promoter wrapping formation and RPo assembly

The structural model presented here corresponds to a low-resolution, single conformational of the transcription initiation complex. Thus, it is not possible to

establish a “mechanism” for the wrapping formation during the different steps of the RP complex assembly. However, based in all the previous studies that support this model, we could suggest a possible mechanism. A set of steps for the wrapping promoter formation are summarized in Figure 47, in which we indicate five main steps (each one identified by a roman number in bold). According with all the previous kinetic studies, [73,80,82] the only stages known in this “mechanism” are stages I and V. They correspond to the formation of the very first RPc complex and the final RPo complex, the one studied in this work. For the steps II, III and IV there is no structural information. Based on this EM model and the available kinetic studies, it has been possible to speculate what would be the interaction between the promoter and some RNAP domains during the wrapping formation.

In the beginning (Stage I), the RNAP (gray surface) recognize the promoter (in green ribbon, from -79 to +18) through the σ^{70} subunit (in the light-blue ribbon), binding the -35 and -10 hexamers. This state, in which the promoter is still as a double strand DNA, [26] is known as the RP close complex (RPc). Next, the αCTD^{I} (in the blue ribbon) binds to the 1st UP-element (-45 to -58), possibly interacting with the phosphates of the DNA backbone minor groove as was reported by Benoff *et al*.[18] Also, this αCTD^{I} would perhaps be interacting with the σ^{70}R4, as has been shown for other promoters containing UP- elements.[19] All these set of interactions mentioned above will produce the first strong bent of the promoter, about ~ 130° of bending (60° + 75°, red and yellow promoter segments, respectively). The observation of this strong bending is in accordance to previous studies, which reported enhanced activity of DNase foot-printing, around the -50 position, and periodic protection in the upstream promoter region.[25,26]

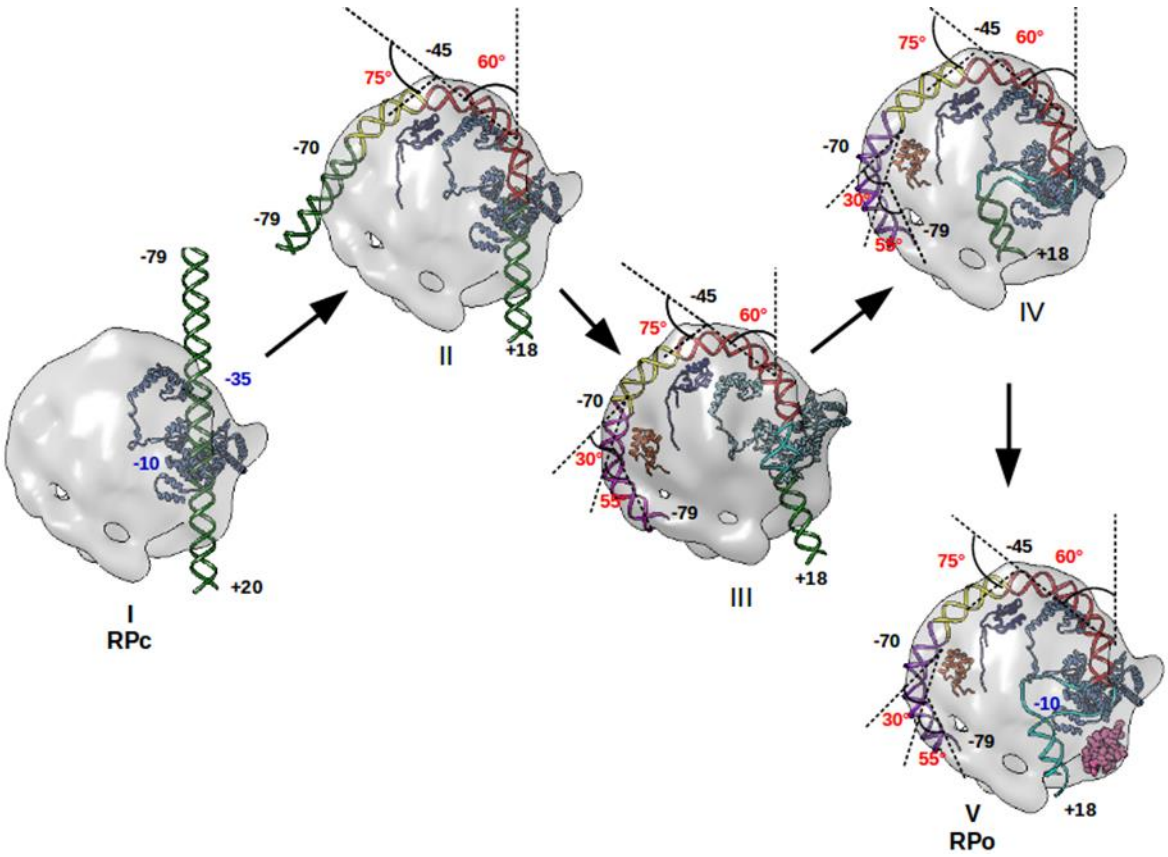


Figure 47 - Proposed promoter wrapping “mechanism”. It was proposed five different stages (identified by roman bold number) for the wrapping formation during the RP complex assembly. The first stage (I), the closed complex (RPc), is the RNAP binding to the promoter (DNA in the green ribbon), which is still completely as double strand; and the last (V), the open complex (RPo), in which the upstream promoter is wrapped around the RNAP surface, the transcription bubble is already formed, and the downstream promoter is inside the cleft. The RNAP is in grey surface and the consensus regions in the promoter are remarked by blue numbers (-10 and -35). Also, the downstream and upstream ends of the promoter are pointed in black numbers (+18 and -79). The DNA segment in red and yellow ribbon (in stage II), remark the first and second bending of the promoter (60° and 75°, respectively), after the binding of the αCTD^{I} (in the blue ribbon) to the first UP-element, and its possible interaction with the σ4. The DNA segment in purple, points out the DNA bent ~ 90° after the binding of the αCTD^{II} (in the orange ribbon, stage III), to the second UP- element. The DNA in the cyan ribbon, point out the different steps of the base pairs unwinding until the transcription bubble formation in stage V. The σ is in the light-blue ribbon, and specifically the $\sigma_{1.1}$ domain is in pink surface, pointing out the it was expelled from the RNAP main channel to allow the stable formation of the RPo.

Source: By the author

Continuing to the stage III, the αCTD^{II} (in the orange ribbon) would binds to the promoter in the 2nd UP-element region (-68 to -79), in a similar conformation as the αCTD^{I} does to the first UP-element. Therefore, a second bent of the promoter of about 90° is performed (purple region of DNA). Previous foot-printing assays support these observations in this model. Those studies have shown that there is a periodic protection of the promoter up to -90 position, with possible pronounced bending

around the -75 position on the promoter.[80]

It is important to mention that in the stage II a considerable force would be imposed on the DNA, that is bent around the 1st UP-element. Thus, a mechanical tension (in stage II represented by pink segmented curve), is generated and transmitted toward the downstream region until achieving the -10-hexamer zone. This -10 hexamer is interacting with the 1.2–3.0 domains of the sigma factor. Those 1.2-3.0 sigma domains have quite less flexibility or mobility than the σ^{70}R4 or αCTDs while interacting with the promoter. Hence, this region will serve as a "fulcrum" or a "pivot", supporting the tension exerted by the upstream region. This stress may induce an unwinding of some base pairs around the -10 hexamer (stage III, cyan DNA fragment). Then, the downstream region is bent to be partially set into the cleft of the RNAP, followed by other conformational changes in different subunits of the enzyme, which are not described here. This stage would be possibly related to a transient state, identified in previous studies, were it was reported a partial unwinding of the base pairs in this region, indicating the beginning of the transcription bubble formation.[73,82,83]

Continuing with the proposed mechanism, this "mechanical stress" persist in the stage III (Figure 47, pink segmented curve). At this stage, different β' jaw and β clamp would occur to squeeze the downstream promoter inside the cleft of the enzyme.[84] Moreover, specifically in this step, it has been proved that the interactions between the upstream promoter regions and the RNAP, the wrapping, are necessaries to overcome the kinetic barrier (rate limiting step) of the DNA isomerization, forming the transcription bubble[73] represented in the stage IV as a DNA fragment in the cyan ribbon.

After the stage IV, the $\sigma^{70}_{1.1}$ domain must be expelled out to allow the complete insertion of the downstream promoter inside the cleft (Figure 47, $\sigma^{70}_{1.1}$ is in the pink surface in RPo), and the RNAP jaws and clamp could be adjusted to stabilize this promoter region and form the RPo stable complex.

The density corresponding to the $\sigma^{70}_{1.1}$ domain, emerge out of the enzyme main channel, near to a positive boundary of the σ^{70}NCR. This $\sigma^{70}_{1.1}$ domain has been widely studied and was shown be quite important in the regulation of the promoter recognition and in an efficient formation of the RPo complex.[12,13,85,86]

4.5 DNA wrapping: agreement of the EM structure with other proposed models

All the previous studies discussed here support how important is the wrapping for several aspects of the RP complex formation. The DNA wrapping seems to be a phenomenon dependent on the promoter sequence as well as of specific RNAP domains, also being affected by some transcriptional regulators as showed before.[28] The 3D model presented in this study is the first structural confirmation of the promoter wrapping.

Recently, Sosa *et al*. reported a mechano-chemical study using optical tweezers on the same RP complex,[72] showing results that seems to be in contradiction with all the previously reported studies performed by AFM, in which the wrapping length was estimated at ~30 nm.[22,23] By optical tweezers, the DNA "wrapping length" for wt λ PR promoter (-100 to +30), was measured to be ~17 nm (~50 bp). A question that arises is how these two independent techniques, AFM and optical tweezers, may report quite different results for the same biological system, and how could be the model presented in this study related to them? To answer that we have to consider the geometry in which the optical tweezers experiments were carried out. As it can be seen in Figure 48, the RPo complex is subjected to an external force (Figure 48A: red arrow, applied from the downstream end) to change from the wrapped (W) to the unwrapped (U) state. According to this geometry, what is really recorded as a measure of "wrapping" in optical tweezers, is a partial unwrapping (Figure 48B, ΔL) of the RP complex. Thus, this extension change would only consider the unwrapping of promoter region from -35 to -90.

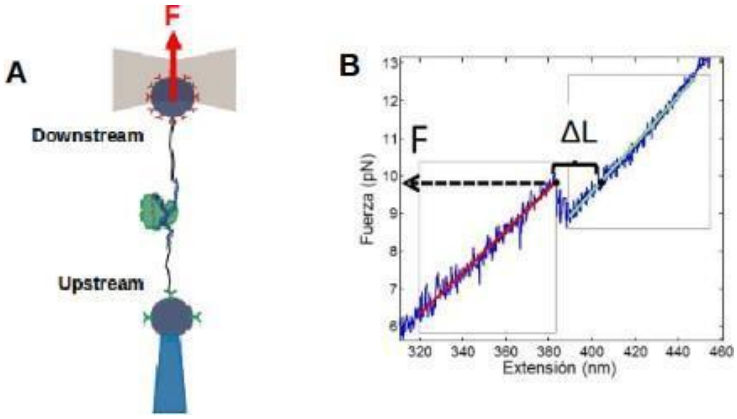


Figure 48- Optical tweezers experiments showing in (A) the geometry to perform the experiments, in which the downstream end of the DNA is fixed to a polystyrene bead -through an antibody-antigen interaction- and pulled with an external force while the upstream end is fixed. (B) Here is shown the Force vs extension points, recorded by the experiment. Around the Force ~ 10 pN, there is an abrupt change in the extension of about 20 nm, which would be related to the unwrapping of the promoter.
Source: Adapted from SOSA.[72]

According to our model, after subjecting the complex to the mechanical stress, the upstream region from -35 to -79 will be unwrapped (Figure 49, red DNA segment), being this extension change (ΔL) of approximately 56 bp or ~ 15 nm, very similar to the ~17 nm reported by optical tweezers studies.

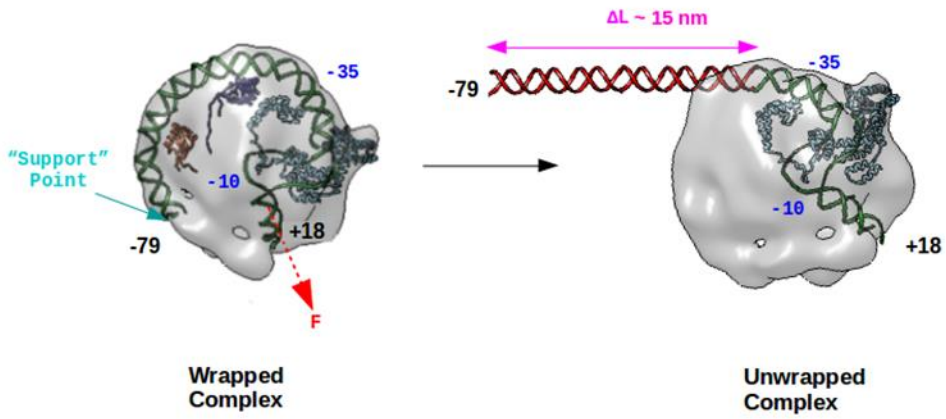


Figure 49 - Representation of EM model and its relationship with the unwrapping mechanism occurring during optical tweezers experiment. In the beginning, in the wrapped complex, the wrapping of the promoters spans from +18 to -79 positions (DNA in green ribbon). The optical tweezers pull the promoter from the downstream end (red arrow in +18), while the upstream end is fixed. After that, the upstream region of the wrapping, from -35 to -79 (DNA in red ribbon), is unwrapped, losing interactions with the RNAP. This extension change is about ~ 15 nm.
Source: By the author.

This hypothesis is supported by the experiment, since the authors reported that

even when the RP is subjected to a mechanical force, it is kept as an open complex. Thus, the transcription bubble cannot fall apart, and the downstream promoter region would not be expelled from the RNAP cleft. As it can be observed, the RP complex model and the proposed hypothesis strongly support the results obtained by optical tweezers.

In the other hand, in the reported AFM experiments, the measurement process is quite different from optical tweezers. In AFM studies, the wrapping extension was calculated as is in Figure 50, being a measure of the difference between the mean of the free-DNA contour length distribution and the mean of the DNA length distribution of the promoter in RPo complex. As it was reported, this compaction length was estimated in ~30 nm, thus would be ~90 base pairs of the promoter forming the overall wrapping around the RNAP. [22–24,28]

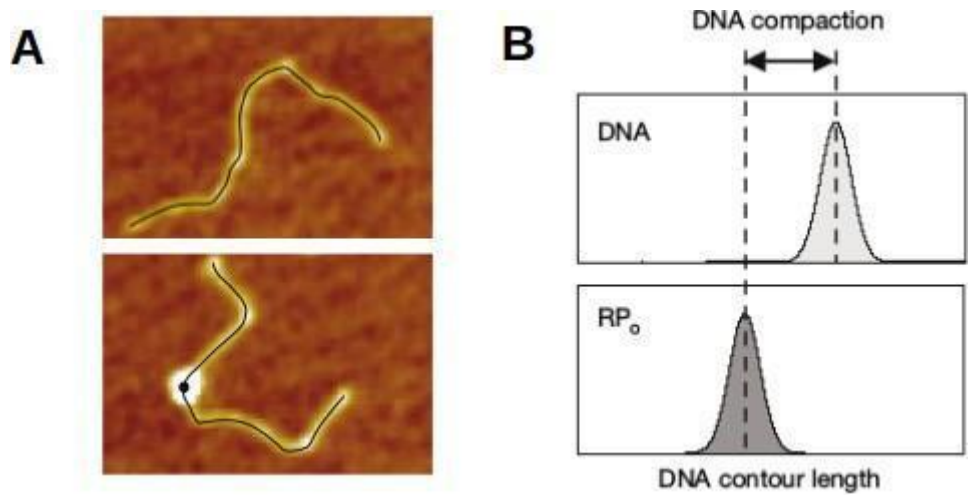


Figure 50 - A) Representative images of how is performed the measure of the free-DNA contour length (top image), and the DNA length in RP complex (bottom image, the RNAP is the brilliant dot). B) After to perform the measures on hundreds or thousands DNA fragments and in complex, there is estimated the mean free-DNA contour length as well as the DNA length in RP complex. The differences between those lengths is the estimated wrapping length.
Source: Adapted from CELLAI.[24]

According to the model presented here, as it was shown in Figure 44, the overall extension of the promoter interacting with the enzyme, spans from about -79 to +18 position, which is around 32 nm of wrapping length.

Therefore, the proposed model can reconcile two independent results obtained for the promoter wrapping length, which at first seemed to be quite discordant.

In another study using Fluorescence Resonance Energy Transfer (FRET) on the RP complex (RNAP - λ PR) analysis,[27] it was reported a reduction of the distance between the -100 and +14 positions on the promoter. It is bigger than 400 Å in the closed complex (RPc), being reduced to less than 100 Å, after the RPo complex formation. In order to verify our proposed model against those findings, we directly measured these two conformations, as show in Figure 51, which represents a direct transition between the RPc and the RPo. The principal locations in the promoter are marked: the -10 and -35 hexamers (in blue), the -100 and +18 sites (in black) and the +14 bp position of the promoter (in purple). Since we are assuming that the region corresponding to the third UP-element of the promoter (-90 to -100) was not interacting with the RNAP, a straight DNA segment was added after the -90 position. According to the suggested model, the distances between the -100 and +14 bp on the promoter would change from ~390 Å in the RPc to ~70 Å in the RPo. Hence, those measure are fairly in agreement with those reported in FRET studies.

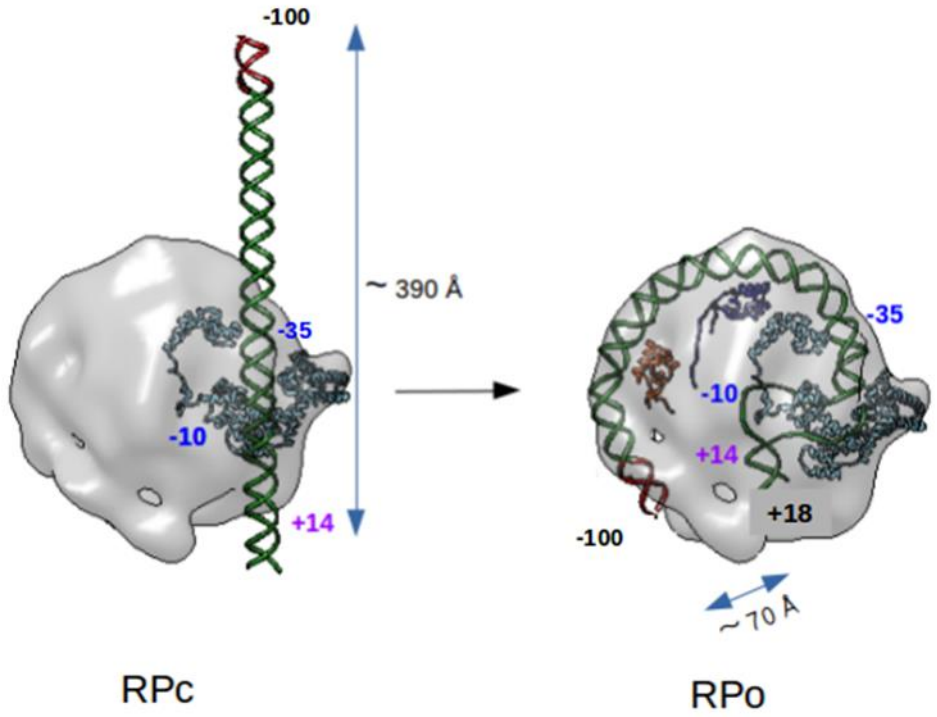


Figure 51 - In the beginning, in the RPc state, the promoter is totally straight and the distance between the+14 and -100 positions is about 390 Å. After the formation of wrapping in the RPo state, the distances between those positions is shortened to ~ 70Å. These observations suggested by the EM model are according with the measures reported FRET experiments. Also, according to the EM model, the third UP-element of the promoter (-90 to -100, in the red ribbon), possibly is not interacting with the RNAP during the RPo state.
Source: By the author

As it has been demonstrated, this model allows corroborating and reconciling the different observations made based on previous independent studies, for this same biological complex, which reinforces the reliability of the proposed structure. Furthermore, different promoters have been reported to form a DNA wrapping, and all of them share similar features in their upstream regions. [29-31] It would be important in future studies to evaluate this interesting phenomenon, with particular interest in its genetic sequence.

5 CONCLUSIONS AND FUTURE PERSPECTIVES

In this work, we successfully assembled the transcription initiation complex between the λ PR wild-type promoter (-100 to +30) and the *E. coli* RNA polymerase-σ70 holoenzyme (RP complex). Using Transmission Electron Microscopy and Single Particle Analysis, it was possible to determine the 3D-structure for this complex at 20 Å resolution. This 3D-model is the very first direct structural confirmation of the DNA promoter wrapping in a bacterial transcription initiation complex.

For this study, the RP complex preparation was performed according to previous biochemical and structural studies, primarily obtaining the RP open complex as the final product. Thus, even being the first direct structural confirmation of the wrapping existence, the suggested model only represents a conformation, the final stage, among all the possible ones, from the closed complex formation to the stable wrapped RP open complex.

The rigid-body fitting of RNA polymerase and the modeled DNA promoter into the 3D-map allowed proposing possible DNA-RNAP interactions that would enable the wrapping formation. According to this model, the αCTDs, probably due to the flexibility of the α-linker, would be the domains independently interacting with the UP-elements of the promoter. Specifically, the αCTD^{I} would be oriented to interact with the first UP element of the promoter (-45 to -58) and the αCTD^{II} would be oriented to interact with the second UP-element (-70 to -79). It was not possible to observe a clear density assigned to the third UP-element of the promoter (-90 to -100), which could suggest that this region is not interacting with the RNAP, during the final stage of wrapping formation. All these observations are strongly supported by different biochemical studies and other indirect structural analysis performed by independent techniques.

According to the proposed structural model, the promoter spans an interaction region with the enzyme from -79 to +18. Moreover, it was observed that the promoter wrapping implies an overall bending of ~255°, as well as a total length of ~32 nm. Those observations have shown to be fairly in agreement with previous independent studies performed by optical tweezers, AFM and FRET.

As a future perspective, we expect that changing the incubation time and letting the RP complex to be assembled at lower temperatures would allow different transient states to be isolated. Thus, using Single Particles Cryo-EM, it would be possible to elucidate a “structural mechanism”, at atomic level, of the DNA wrapping during the transcription initiation. It would be also important to evaluate this biological system using mutant promoters and mutant proteins, and therefore getting new insights about the implication of promoter wrapping in the RP complex formation. Finally, other promoters could be tested for a structural study, to observe how the DNA-protein contacts would change according to the promoter nature, regarding different spacer or discriminator lengths as well as distinct UP-elements configuration.

.

REFERENCES

1 WERNER, F.; GROHMANN, D. Evolution of multi subunit RNA polymerases in the three domains of life. **Nature Reviews. Microbiology**, v. 9, n. 2, p. 85–98, 2011.

2 MURAKAMI, K. S.; DARST, S. A. Bacterial RNA polymerases: the wholo story. **Current Opinion in Structural Biology**, v. 13, n. 1, p. 31–39, 2003.

3 ISHIHAMA, A. Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. **FEMS Microbiology Reviews**, v. 34, n. 5, p. 628– 645, 2010.

4 TRIPATHI, L.; ZHANG, Y.; LIN, Z. Bacterial sigma factors as targets for engineered or synthetic transcriptional control. **Frontiers in Bioengineering and Biotechnology**, v. 2, n.33, 2014. doi:10.3389/fbioe.2014.00033.

5 FEKLÍSTOV, A. et al. Bacterial sigma factors: a historical, structural, and genomic perspective. **Annual Review of Microbiology**, v. 68, p. 357–376, 2014. doi: 10.1146/annurev-micro-092412-155737.

6 MURAKAMI, K. S. X-ray crystal structure of Escherichia coli RNA polymerase σ70 holoenzyme. **Journal of Biological Chemistry**, v. 288, n. 13, p. 9126–9134, 2013.

7 SAECKER, R. M.; RECORD, M. T.; DEHASETH, P. L. Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. **Journal of Molecular Biology**, v. 412, n. 5, p. 754–771, 2011.

8 HOOK-BARNARD, I. G.; HINTON, D. M. Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. **Gene Regulation and Systems Biology**, v. 1, p. 275–293, 2007. doi:10.1177/117762500700100020.

9 NARAYANAN, A. et al. Cryo-EM structure of Escherichia coli σ70 RNAP and promoter DNA complex revealed a role of σ non-conserved region during the open complex formation. **Journal of Biological Chemistry**, v. 293, n.19, p. 7367-7375, 2018.

10 HOOK-BARNARD, I.; JOHNSON, X. B.; HINTON, D. M. Escherichia coli RNA Polymerase Recognition of a σ70-Dependent Promoter Requiring a −35 DNA Element and an Extended −10 TGn Motif. **Journal of Bacteriology**, v. 188, n. 24, p. 8352– 8359, 2006.

11 MURAKAMI, K. S. et al. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. **Science**, v. 296, n. 5571, p. 1285–1290, 2002.

12 CAMARERO, J. A. et al. Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 13, p. 8536–8541, 2002.

13 BAE, B. et al. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. **eLife**, v. 4, 8 set. 2015. doi: 10.7554/eLife.08504.

14 ZHANG, G.; DARST, S. A. Structure of the Escherichia coli RNA polymerase alpha subunit amino-terminal domain. **Science**, v. 281, n. 5374, p. 262–266, 1998.

15 HARAN, T. E.; MOHANTY, U. The unique structure of A-tracts and intrinsic DNA bending. **Quarterly Reviews of Biophysics**, v. 42, n. 1, p. 41–81, 2009.

16 ESTREM, S. T. et al. Identification of an UP element consensus sequence for bacterial promoters. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 17, p. 9761–9766,1998.

17 SAVERY, N. J. et al. Determinants of the C-terminal domain of the Escherichia coli RNA polymerase alpha subunit important for transcription at class I cyclic AMP receptor protein-dependent promoters. **Journal of Bacteriology**, v. 184, n. 8, p. 2273– 2280, 2002.

18 BENOFF, B. et al. Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. **Science**, v. 297, n. 5586, p. 1562–1566, 2002.

19 ROSS, W. et al. An intersubunit contact stimulating transcription initiation by E coli RNA polymerase: interaction of the alpha C-terminal domain and sigma region 4. **Genes & Development**, v. 17, n. 10, p. 1293–1307, 2003.

20 ROSS, W. et al. Escherichia coli promoters with UP elements of different strengths: modular structure of bacterial promoters. **Journal of Bacteriology**, v. 180, n. 20, p. 5375– 5383,1998.

21 BORIN, B. N.; TANG, W.; KREZEL, A. M. Helicobacter pylori RNA polymerase α-subunit C-terminal domain shows features unique to ε-proteobacteria and binds NikR/DNA complexes. **Protein Science: A Publication of the Protein Society**, v. 23, n. 4, p. 454– 463, 2014.

22 RIVETTI, C.; GUTHOLD, M.; BUSTAMANTE, C. Wrapping of DNA around the E.coli RNA polymerase open promoter complex. **EMBO Journal**, v. 18, n. 16, p. 4464– 4475,1999.

23 MANGIAROTTI, L. et al. Sequence-dependent upstream DNA-RNA polymerase interactions in the open complex with λPR λPRM promoters and implications for the mechanism of promoter interference. **Journal of Molecular Biology**, v. 385, n. 3, p. 748– 760, 2009.

24 CELLAI, S. et al. Upstream promoter sequences and αCTD mediate stable DNA wrapping within the RNA polymerase–promoter open complex. **EMBO Reports**, v. 8, n. 3, p. 271–278, 2007.

25 CRAIG, M. L.; SUH, W. C.; RECORD, M. T. HO. and DNase I probing of E sigma 70 RNA polymerase--lambda PR promoter open complexes: Mg2+ binding and its

structural consequences at the transcription start site. **Biochemistry**, v. 34, n. 48, p. 15624–15632, 1995.

26 CRAIG, M. L. et al. DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. **Journal of Molecular Biology**, v. 283, n. 4, p. 741–756,1998.

27 SREENIVASAN, R. et al. Fluorescence resonance energy transfer Characterization of DNA wrapping in closed and open Escherichia coli RNA polymerase- λP(R) promoter complexes. **Biochemistry**, v. 55, n. 14, p. 2174–2186, 2016.

28 DONISELLI, N. et al. New insights into the regulatory mechanisms of ppGpp and DksA on Escherichia coli RNA polymerase-promoter complex. **Nucleic Acids Research**, v. 43, n. 10, p. 5249–5262, 2015.

29 DAME, R. T. et al. Structural basis for H-NS-mediated Trapping of RNA polymerase in the open initiation complex at the rrnB P1. **Journal of Biological Chemistry**, v. 277, n. 3, p. 2146–2150, 2002.

30 MAURER, S. et al. RNA polymerase and an activator form discrete subcomplexes in a transcription initiation complex. **EMBO Journal**, v. 25, n. 16, p. 3784–3790, 2006.

31 SHIN, M. et al. DNA looping-mediated repression by histone-like protein H-NS: specific requirement of Esigma70 as a cofactor for looping. **Genes & Development**, v. 19, n. 19, p. 2388–2398, 2005.

32 KIM, T.-K. et al. Trajectory of DNA in the RNA polymerase II transcription preinitiation complex. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 23, p. 12268–12273,1997.

33 FORGET, D. et al. RAP74 induces promoter contacts by RNA polymerase II upstream and downstream of a DNA bend centered on the TATA box. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 14, p. 7150– 7155,. 1997.

34 KLEYWEGT, G. J. Validation of protein crystal structures. **Acta Crystallographica D:** biological crystallography, v. 56, n. Pt 3, p. 249–265, 2000.

35 YU, H. Extending the size limit of protein nuclear magnetic resonance. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 2, p. 332–334,1999.

36 CRESSEY, D.; CALLAWAY, E. Cryo-electron microscopy wins chemistry Nobel**.** **Nature News**, v. 550, n. 7675, p. 167, 2017.

37 DUBOCHET, J. et al. Emerging techniques: Cryo-electron microscopy of vitrified biological specimens. **Trends in Biochemical Sciences**, v. 10, n. 4, p. 143–146,1985.

38 ADRIAN, M. et al. Cryo-electron microscopy of viruses. **Nature**, v. 308, n. 5954, p. 32,1984.

39 KÜHLBRANDT, W. Biochemistry. The resolution revolution. **Science**, v. 343, n. 6178, p. 1443–1444, 2014.

40 BAI, X.; MCMULLAN, G.; SCHERES, S. H. W. How cryo-EM is revolutionizing structural biology. **Trends in Biochemical Sciences**, v. 40, n. 1, p. 49–57, 2015.

41 KÜHLBRANDT, W. Cryo-EM enters a new era. **eLife**, v.3, p.e03678, 2014. doi: 10.7554/eLife.03678.

42 CAMPBELL, M. G. et al. Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. **Structure**, v. 20, n. 11, p. 1823– 1828, 2012.

43 HE, Y. et al. Near-atomic resolution visualization of human transcription promoter opening. **Nature**, v. 533, n. 7603, p. 359–365, 2016.

44 GUO, X. et al. Structural basis for NusA stabilized transcriptional pausing. **Molecular Cell**, v. 69, n. 5, p. 816- 827, 2018.

45 EMStats: EMDB statistics. Available from: <https://www.ebi.ac.uk/pdbe/emdb/statistics_main.html/>. Accessible at: 11 June 2018.

46 BRENNER, S.; HORNE, R. W. A negative staining method for high resolution electron microscopy of viruses. **Biochimica Et Biophysica Acta**, v. 34, p. 103–110, 1959.

47 BOOTH, D. S.; AVILA-SAKAR, A.; CHENG, Y. Visualizing proteins and macromolecular complexes by negative stain EM: from grid preparation to image acquisition. **Journal of Visualized Experiments**: JoVE, v. 58, pii: 3227, 2011. doi: 10.3791/3227.

48 VAN HEEL, M. et al. Single-particle electron cryo-microscopy: towards atomic resolution. **Quarterly Reviews of Biophysics**, v. 33, n. 4, p. 307–369, 2000.

49 NOGALES, E. The development of cryo-EM into a mainstream structural biology technique. **Nature Methods**, v. 13, n. 1, p. 24–27, 2016.

50 MERK, A. et al. Breaking Cryo-EM resolution barriers to facilitate drug discovery. **Cell**, v. 165, n. 7, p. 1698–1707, 2016.

51NOBELPRIZE-ORG.Available from: <https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2017/advanced.html>. Accessible at: 11 June 2018.

52 ELECTRON Microscopy. Available from: <https://advanced microscopy.utah.edu/education/electron-micro/>. Accessible at: 11 June 2018.

53 GLAESER, R. M.; HALL, R. J. Reaching the information limit in Cryo-EM of biological macromolecules: experimental aspects. **Biophysical Journal**, v. 100, n. 10, p. 2331–2337, 2011.

54 SIGWORTH, F. J. Principles of cryo-EM single-particle image processing. **Microscopy**, v. 65, n. 1, p. 57–67, 2016.

55 MYSCOPE: introduction - aims and learning outcomes. Available from: <http://www.ammrf.org.au/myscope/tem/introduction/>. Accessible at: 11 June 2018.

56 FRANK, J. **Three-dimensional electron microscopy of macromolecular assemblies**. Oxford: Oxford University Press, 2006.

57 DOWNING, K. H.; GLAESER, R. M. Restoration of weak phase-contrast images recorded with a high degree of defocus: the “twin image” problem associated with CTF correction. **Ultramicroscopy**, v. 108, n. 9, p. 921–928, 2008.

58 VAN HEEL, M. Unveiling ribosomal structures: the final phases. **Current Opinion in Structural Biology**, v. 10, n. 2, p. 259–264, 2000.

59 ZERNIKE, F. How I discovered phase contrast. **Science**, v. 121, n. 3141, p. 345–349,1955.

60 MILAZZO, A.-C. et al. Active pixel sensor array as a detector for electron microscopy. **Ultramicroscopy**, v. 104, n. 2, p. 152–159, 2005.

61 LUDTKE, S. J.; BALDWIN, P. R.; CHIU, W. EMAN: semi automated software for high-resolution single-particle reconstructions. **Journal of Structural Biology**, v. 128, n. 1, p. 82–97,1999.

62 FRANK, J.; SHIMKIN, B.; DOWSE, H. Spider—a modular software system for electron image processing. **Ultramicroscopy**, v. 6, n. 4, p. 343–357,1981.

63 VAN HEEL, M. et al. A New generation of the IMAGIC image processing system. **Journal of Structural Biology**, v. 116, n. 1, p. 17–24,1996.

64 AFANASYEV, P. et al. Single-particle cryo-EM using alignment by classification (ABC): the structure of Lumbricus terrestris haemoglobin. **IUCrJ**, v. 4, n. 5, p. 678–694, 2017.

65 VAN HEEL, M. Detection of objects in quantum-noise-limited images. **Ultramicroscopy**, v. 7, n. 4, p. 331–341,1982.

66 VAN HEEL, M. Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). **Ultramicroscopy**, v. 13, n. 1–2, p. 165–183, 1984.

67 VAN HEEL, M.; PORTUGAL, R. V.; SCHATZ, M. Multivariate statistical analysis of large datasets: single particle electron microscopy. **Open Journal of Statistics**, v. 06, n. 04, p. 701, 2016.

68 VAN HEEL, M et al. Angular reconstitution in three-dimensional electron microscopy: historical and theoretical aspects. **Scanning Microscopy**, p. 195–210, 1997.

69 SCHATZ, M. Angular reconstitution in three-dimensional electron microscopy: practical and technical aspects. **Scanning Microscopy**, v.11, p.179-193,1997.

70 ORLOVA, E. V.; SAIBIL, H. R. Structural analysis of macromolecular assemblies by electron microscopy. **Chemical Reviews**, v. 111, n. 12, p. 7710–7748, 2011.

71 MIT-CryoEM. Available from: <https://people.csail.mit.edu/gdp/cryoem.html>. Accessible at: 11June. 2018.

72 SOSA, R. P. **Dinámica del complejo enrollado de la iniciación de la transcripción bacteriana**. 2017. 110p. Tesis (Maestría em Bioquímica y Biología Molecular) – Escuela de Posgrado, Universidad Peruana Cayetano Heredia, Lima, Perú, 2017.

73 DAVIS, C. A. et al. The effects of upstream DNA on open complex formation by Escherichia coli RNA polymerase. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 2, p. 285–290, 2005.

74 VAN HEEL, M.; SCHATZ, M. Fourier shell correlation threshold criteria. **Journal of Structural Biology**, v. 151, n. 3, p. 250–262, 2005.

75 VAN HEEL, M.; SCHATZ, M. Reassessing the Revolution´s Resolutions. **bioRxiv**, p. 224402, 2017.doi.org/10.1101/224402.

76 IMAGE SCIENCE - IMAGIC manuals. Available from: <http://www.imagescience.de/manuals.html>. Accessible at: 14 March 2018.

77 VAN DIJK, M.; BONVIN, A. M. J. J. 3D-DART: a DNA structure modelling server. **Nucleic Acids Research**, v. 37, suppl_2, p. W235–W239, 2009.

78 PETTERSEN, E. F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605–1612, 2004.

79 ARTSIMOVITCH, I. et al. Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. **Journal of Biological Chemistry**, v. 278, n. 14, p. 12344–12355, 2003.

80 RUFF, E. F.; RECORD, M. T.; ARTSIMOVITCH, I. Initial events in bacterial transcription initiation. **Biomolecules**, v. 5, n. 2, p. 1035–1062, 2015.

81 RUFF, E. F. et al. E. coli RNA polymerase determinants of open complex lifetime and structure. **Journal of Molecular Biology**, v. 427, n. 15, p. 2435–2450, 2015.

82 SCLAVI, B. et al. Real-time characterization of intermediates in the pathway to open

complex formation by Escherichia coli RNA polymerase at the T7A1 promoter. **Proceedings of the National Academy of Sciences**, v. 102, n. 13, p. 4706–4711, 2005.

83 BURNS, H. D.; ISHIHAMA, A.; MINCHIN, S. D. Open complex formation during transcription initiation at the Escherichia coli galP1 promoter: the role of the RNA polymerase alpha subunit at promoters lacking an UP-element. **Nucleic Acids Research**, v. 27, n. 9, p. 2051–2056,1999.

84 KONTUR, W. S. et al. Probing DNA binding, DNA opening and assembly of a downstream clamp/jaw in E. coli RNA polymerase – λPR promoter complexes using salt and the physiological anion glutamate. **Biochemistry**, v. 49, n. 20, p. 4361–4373, 2010.

85 VUTHOORI, S. et al. Domain 1.1 of the sigma (70) subunit of Escherichia coli RNA polymerase modulates the formation of stable polymerase/promoter complexes. **Journal of Molecular Biology**, v. 309, n. 3, p. 561–572, 2001.

86 WILSON, C.; DOMBROSKI, A. J. Region 1 of σ70 is required for efficient isomerization and initiation of transcription by Escherichia coli RNA polymerase 11Edited by M. Gottesman. **Journal of Molecular Biology**, v. 267, n. 1, p. 60–74, 1997.