Jurnal Teknologi

Full Paper

IN SILICO STRUCTURAL CHARACTERIZATION OF *L. lactis subsp. cremoris* MG1363 FFH-FTSY COMPLEX IN PROTEIN TARGETING INTERACTION

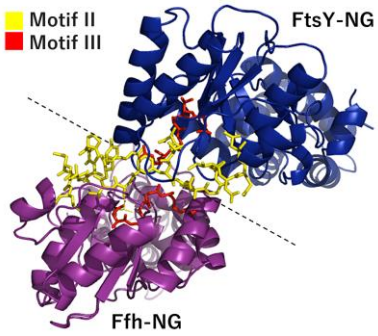
Noor Izawati Alias[a], Abdul Munir Abdul Murad[b], Farah Diba Abu Bakar[b], Rosli Md. Illias[a*]

[a]Department of Bioprocess and Polymer Engineering, School of Chemical and Energy Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
[b]School of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor 43600, Malaysia

Article history
Received
3 January 2019
Received in revised form
30 January 2019
Accepted
6 February 2019
Published online
18 February 2019

*Corresponding author
r-rosli@utm.my

Graphical abstract



Abstract

In bacteria, gene conservation and experimental data show that *Lactococcus lactis* has the simplest version of protein secretion system compared to *Escherichia coli* and *Bacillus subtilis* whose systems are more complex. *L. lactis* only possess the signal recognition particle (SRP) pathway, where the specific interaction of Ffh and FtsY is known to be essential for the efficiency and fidelity of its protein targeting. Therefore, modelling and structural characterization study of Ffh and FtsY will give an idea of its crucial region and amino acids that are critical in Ffh-FtsY interaction during protein targeting. This work is the first attempt to model *L. lactis* Ffh-FtsY complex, which was derived by computational docking, where a blind dock was applied. Results showed that the complex interface was predominantly stabilized by four hydrophobic interactions and 17 hydrogen bonds, where these putative binding interfaces are mostly confined at the motifs II and III in each G domain of Ffh and FtsY. Several residues were expected to play important roles in initiating or regulating guanosine triphosphate hydrolysis, including residue R142. This structural information will allow for the rational design of *L. lactis* Ffh-FtsY association in the future.

Keywords: *Lactococcus lactis*, signal recognition particle, homology modelling, protein docking, molecular dynamics simulation

Abstrak

Dalam bakteria, pemuliharaan gen dan data eksperimen menunjukkan bahawa *Lactococcus lactis* mempunyai versi sistem rembesan protein yang paling mudah berbanding *Escherichia coli* dan *Bacillus subtilis* yang sistemnya lebih rumit. *L. lactis* hanya mempunyai laluan partikel pengecaman signal (SRP), di mana interaksi khusus di antara Ffh dan FtsY diketahui adalah penting bagi kecekapan dan ketepatan penargetan proteinnya. Oleh itu, pemodelan dan pencirian struktur Ffh dan FtsY akan memberi gambaran kawasan genting dan asid amino-asid amino yang kritikal dalam interaksi Ffh-FtsY semasa penargetan protein. Kajian ini merupakan percubaan pertama untuk memodelkan kompleks *L. lactis* Ffh-FtsY, yang diperolehi dengan pengkomputeran dok, di mana dok rambang digunakan. Hasil kajian menunjukkan bahawa antara muka kompleks kebanyakannya distabilkan oleh empat interaksi hidrofobik dan 17 ikatan hidrogen, di mana kebanyakan ikatan antara muka terletak pada motif II dan III dalam setiap domain

G Ffh dan FtsY. Beberapa residu dijangka memainkan peranan penting dalam memulakan atau mengawal hidrolisis trifosfat guanosin, termasuk residu R142. Maklumat struktur ini akan membolehkan reka bentuk rasional bagi penyatuan *L. lactis* Ffh-FtsY pada masa akan datang.

Kata kunci: *Lactococcus lactis*, partikel pengecaman signal, pemodelan homolog, dok protein, simulasi molekul dinamik

1.0 INTRODUCTION

The signal recognition particle (SRP) directs secretory and membrane proteins to the cellular translocon, a complex of protein translocation machinery, during translation [1]. Generally, SRP interacts with the signal sequence of nascent proteins as it appears from the ribosome to form an SRP-ribosome complex. This cytosolic complex is targeted to the translocon embedded in the endoplasmic reticulum (in eukaryote) or the cytoplasmic membrane (in prokaryote) via an interaction with the SRP receptor.

The discovery of the SRP components was first identified in mammalian cells in the early 1980s, later the identification of its homologs in bacteria and genomics analysis of numerous organisms revealed that components of the SRP pathway are universally conserved [2,3]. The bacterial SRP consists of three components, 4.5S RNA, SRP protein (Ffh) and SRP receptor (FtsY) [4]. Both Ffh and FtsY are guanosine triphosphate (GTP) binding proteins. As GTP-binding proteins, Ffh and FtsY function as reciprocal GTP-ase activating proteins for each other. Their GTP activities are critical to the targeting and translocation of proteins but neither has significant GTP binding activity by itself [5].

In the past decades, the possibility to secrete heterologous proteins in a Generally Recognized as Safe Gram-positive, lactic acid bacteria (LAB) have been addressed in several studies [6,7,8]. LAB have been studied extensively and are now among the best-characterized microorganisms with respect to their genetics, physiology, and applications [9]. Many studies have used LAB *Lactococcus lactis* to produce recombinant proteins due to its remarkable advantage; that it does not produce endotoxic lipopolysaccharides (LPS) or any proteases as other well-known protein producers do [10]. The bacterium has a well-established safety profile, thus makes it suitable to be used as delivery vehicles in pharmaceuticals and in fermented food product industry [11]. However, the low secretion level of heterologous proteins by *L. lactis* becomes a bottleneck for its application in industry. Therefore, a variety of strategies have been explored and developed to improve the production yields of *L. lactis* secreted proteins, but most studies focused on developing effective expression systems, strains optimization and modifications [12].

The comparative analysis of the complete genome sequence of *L. lactis* revealed that its secretion machinery comprised fewer components than the well-characterized *Escherichia coli* and *Bacillus subtilis* Sec machinery. Identification of genes specifically involved in protein secretion in *L. lactis* has resulted in detection of genes encoding homologs of Ffh (ffh), HBsu (hslA), FtsY and scRNA from the SRP system. Thus in *L. lactis*, it is believed that secretion is mainly cotranslational when no homologs of *E. coli* SecB or *B. subtilis* CsaA have been identified in the genome of *L. lactis* IL1403 [13]. In cotranslational protein export, gene conservation analysis and experimental data showed that the targeting and/or translocation process via the complex Ffh and its receptor, FtsY play a major role in the protein secretion [12]. Previous studies claimed and proved that during protein targeting to the membrane, Ffh and its specific interaction with FtsY ensure the efficiency and fidelity of protein translocation [12,14,15].

The present work is an *in silico* approach to model structures of *L. lactis* Ffh and FtsY using homology modelling. Structure optimizations and computational docking were then used to predict potential binding conformations of protein complex *L. lactis* Ffh-FtsY. The resulting model was assessed and discussed in the sequel. To the best of our knowledge, to date, no three-dimensional (3D) structures are yet available concerning *L. lactis* Ffh and FtsY and its protein targeting interaction have not been reported. LAB, *L. lactis subsp. cremoris* MG1363 was used as a model organism.

2.0 METHODOLOGY

2.1 Sequence Retrieval and Analysis

L. lactis subsp. cremoris MG1363 Ffh (accession number A2RJM0) and its docking protein, FtsY (accession number A2RLY9) were retrieved from UniProt/Swiss-prot [16]. Both sequences were retrieved as a query (target) with a total length of 518 and 459 amino acids, respectively. The linear chain of Ffh and FtsY were subjected to sequence analysis using PSI-BLAST [17], HHpred [18] and Phyre [19]. The conserved domain search was determined using available online bioinformatics tool, InterProScan [20,21] available at EBI http://www.ebi.ac.uk/Tools/pfa/iprscan/.

2.2 3D Structure Modeling and Evaluation

There was a reliable percentage of sequence identity and similarity between the sequences of *L. lactis* Ffh and FtsY with experimentally determined structures, 2J37 and 2YHS, respectively. Therefore, homology modelling was used to construct the 3D structure of Ffh and FtsY using MODELLER version 9.9 [22]. 50-full atom models of Ffh and FtsY were constructed by the satisfaction of spatial restraints, using its 'automodel' class. Models with the lowest energy value (DOPE) and objective function profiles were selected and evaluated by PROCHECK [23] for the Ramachandran plot quality evaluation, VERIFY3D [24] for measuring the compatibility of an atomic model (3D) with its own amino acid sequence (1D), Errat [25] for detecting local errors, root mean square deviation (RMSD) and TM-score [26] to measure protein fold and global topology. A model with the most satisfactory quality was chosen for refinement and validation. The steepest descent energy minimization was done to remove steric clashes.

2.3 Molecular Dynamics (MD) Simulations

The resulted models were then subjected to structural stability optimization and energy minimization step using the steepest descent algorithm. These structure models were refined with an MD simulation using GROMACS [27]. In this study, GROMACS 4.5.3 package and the all-hydrogen function, GROMOS96 force field applications run on the operating system, Linux as follows. Energy minimization was first performed with steepest descent method for 2000 steps. Simulations were conducted at 300 K and isotropic pressure coupling was applied. The equilibration dynamics of the entire system was performed for 50 ps. Lastly, 10 ns MD simulation was conducted at 1 atm. A time step of 2 fs was used, where coordinates were collected every 1 ps. The refined structure was taken from the trajectory system for the determination of the protein geometry quality and the structure reliability. When analyzing the resulting trajectories of the final optimized models, GROMACS interact with PyMol [28] and Grace where both applications will support analysis of MD simulation. A Python-enhanced molecular graphics tool, PyMol is an application to visualize molecule structure, and Grace is an application in Linux to display graphs.

2.4 Protein-protein Docking and Analysis

Ffh and FtsY protein docking were performed to probe for possible interaction or binding sites using a fully automated algorithm for a protein-protein docking web server, ClusPro using the improved docking program, PIPER [29]. PIPER utilizes a Fast Fourier Transform (FFT)-based docking method with pairwise interaction potentials during the initial rigid-body docking step [30], as part of its scoring function. A blind dock was performed where the default values for all parameters were used. PIPER ranks their energy models based on a cluster size which is based on the number of complexes that have the largest number of neighbours within a certain fixed cluster radius of ≤10.0 Å C_α RMSD as the distance measure. Subsequent, Protein Interaction Calculator (PIC) [31] was used to identify the key binding characteristics of the docked proteins.

3.0 RESULTS AND DISCUSSION

3.1 Template-based Modeling

The tertiary structure of *L. lactis* Ffh and FtsY was not publicly available in the Protein Data Bank (http://www.rcsb.org/). Hence, homology modelling was used to construct an atomic-resolution model of both Ffh and FtsY from its amino acid sequence and a crystal structure of its homologous protein. The protein sequences of *L. lactis* Ffh (A2RJM0) and FtsY (A2RLY9) have been analyzed by computer programs to find similar sequences in databases and perform structure prediction. Structure similarity searching using PSI-BLAST, HHpred and Phyre proposed the experimentally solved X-ray crystallized 3D structural homologs of the *L. lactis* Ffh and FtsY were retrieved from PDB database to be considered for the homology modelling. Table 1 and Table 2 shows the homologous protein structures resulted from all servers. By considering the sequences producing significant alignments with E-value better than threshold (an E-value better than 10^{-10}), the *L. lactis* Ffh and FtsY sequences are like many solved 3D protein structures in the PDB database, defined as 'signal recognition particle' thus confirming that both proteins are well conserved within different organisms, in both eukaryotic and prokaryotic (archea) groups.

A crystal structure of SRP 54 kDa (SRP54/Ffh) from mammalian *Canis sp*. (PDB ID: 2J37) [32] and a crystal structure of SRP receptor (FtsY) from *E. coli* (PDB ID: 2YHS) [33] were chosen as the most suitable template for the construction of *L. lactis* Ffh and FtsY 3D model, respectively. Both 2J37 and 2YHS appeared as top five ranked in the search results from all servers. 2J37, ranked 1 in the search results from all servers despite it being from a different class of taxonomy (animalia). It is well-known that SRP protein targeting is highly conserved in all three kingdoms of life. Evolutionarily, two proteins with similar amino acid sequences normally possess similar protein structures, resulting in naturally occurring as homologous proteins [34].

Table 1 Templates to generate the *L. lactis* Ffh model. The alignment search results against the PDB database from various servers

Server	PDB ID_chain	Protein	Organism	Percentage (%)			Resolution (Å)
				Coverage	Identity	Similarity	
PSI-BLAST	2J37_W	SRP54	*Canis sp.*	94	32	49	NA
	3DM5_A	SRP54	*Pyrococcus furiosus*	85	33	57	2.51
	2FFH_A	Ffh	*Thermus aquaticus*	82	47	66	3.20
	3NDB_B	SRP54	*Methanocaldococcus jannaschii*	81	37	60	3.00
	2XXA_A	SRP54	*Escherichia coli*	80	50	70	3.94
HHpred	2J37_W	SRP54	*Canis sp.*	95	31	47	NA
	2XXA_A	SRP54	*Escherichia coli*	84	43	67	3.94
	3DM5_A	SRP54	*Pyrococcus furiosus*	85	33	56	2.51
	2FFH_A	Ffh	*Thermus aquaticus*	82	47	74	3.20
	2V3C_C	SRP54	*Methanocaldococcus jannaschii*	82	37	63	2.50
Phyre	2J37_W	SRP54	*Canis sp.*	90	31	ND	NA
	2IY3_A	Ffh	*Thermus aquaticus*	82	43	ND	16.0
	3DM5_A	SRP54	*Pyrococcus furiosus*	80	34	ND	2.51
	2J28_9	SRP54	*Escherichia coli*	83	49	ND	NA
	1QZW_C	SRP54	*Sulfolobus solfataricus*	82	35	ND	4.10

Note: SRP: Signal recognition particle, NA: not available, ND: not determined

Table 2 Templates to generate the *L. lactis* FtsY model. The alignment search results against the PDB database from various servers

Server	PDB ID_chain	Protein	Organism	Percentage (%)			Resolution (Å)
				Coverage	Identity	Similarity	
PSI-BLAST	2QY9_A	FtsY	*Escherichia coli*	65	45	67	1.90
	2XXA_B	FtsY	*Escherichia coli*	65	45	67	3.94
	2YHS_A	FtsY	*Escherichia coli*	65	45	67	1.60
	1FTS_A	FtsY	*Escherichia coli*	64	45	67	2.20
	2OG2_A	FtsY	*Arabidopsis thaliana*	62	43	60	2.00
HHpred	2YHS_A	FtsY	*Escherichia coli*	97	38	58	1.60
	2OG2_A	FtsY	*Arabidopsis thaliana*	69	40	65	2.00
	1ZU4_A	FtsY	*Mycoplasma mycoides*	65	45	76	1.95
	3B9Q_A	FtsY	*Arabidopsis thaliana*	65	42	67	1.75
	5L3R_A	SRP54	*Arabidopsis thaliana*	63	32	49	2.50
Phyre	2YHS_A	FtsY	*Escherichia coli*	66	45	ND	1.60
	3B9Q_A	FtsY	*Arabidopsis thaliana*	65	42	ND	1.75
	3DM5_A	SRP54	*Pyrococcus furiosus*	62	35	ND	2.51
	2IY3_A	Ffh	*Thermus aquaticus*	63	32	ND	16.0
	2QY9_A	FtsY	*Escherichia coli*	65	45	ND	1.90

Note: SRP: Signal recognition particle, NA: not available, ND: not determined

Table 3 Results of the stereochemical validation and quality assessment of *L. lactis* Ffh-NG and FtsY-NG model before and after energy minimization

Model/Template	Model quality validation							
	PROCHECK/Ramachandran plot (%)				VERIFY3D	ERRAT	TM-align	
	Core	Allow	Gener	Disall	(%)	(%)	RMSD, Å	TM-score
Model Ffh:								
Before energy minimization	76.9	20.0	1.2	1.9	50.9	90.4	0.79	0.98
After energy minimization	85.8	11.6	1.7	0.9	81.0	90.5	0.82	0.95
Model FtsY:								
Before energy minimization	76.1	22.4	1.1	0.4	92.6	90.5	0.50	0.96
After energy minimization	92.6	6.6	0.0	0.7	99.7	90.6	0.68	0.99

A 3D structure of Ffh and FtsY were built using MODELLER that produced 50 models comprising all non-hydrogen atoms with different conformations and were optimized using conjugate gradients. The model was generated in such a way that a set of spatial and empirically determined restraints were optimally satisfied [22]. Models that produced high violations of the restraints were considered as poor, which in turn lead to higher DOPE and objective function, calculated by a CHARMM-22 forcefield [35]. Thus the best model was determined to be the one with the lowest DOPE and objective function score.

Therefore, model ffh9 (Ffh) and model ftsy34 (FtsY) with the lowest value of the DOPE score, -41783.51 and -37749.71 kJ/mol, respectively, were selected as final models and subjected into MD simulation using GROMACS for extensive energy minimization (steepest descent) algorithms. Table 3 shows the selected model and evaluation by PROCHECK, VERIFY3D, Errat, RMSD, and TM-score before energy minimization.

3.2 Refinement of Template-based Structures by MD Simulation

Template-based models usually have high-energy levels due to unfavourable bond lengths, bond angles, torsion angles and contacts [36]. Thus energy minimization was carried out for the initial NG domain of *L. lactis* Ffh (termed Ffh-NG) and FtsY (termed FtsY-NG) and resulted in a model with reduced energy, without significantly altering its overall structure. The process involved geometry optimization to regularize local bond and angle geometry and to relax close contacts in the geometric chain [36]. Two independent MD simulations using the steepest descent algorithm were performed for 10 ns in an attempt to reach a more preferential minimum level of total energy, as protein conformations are more stable at low energy levels [37].

It is noteworthy that, after 2 ns, the deviation (RMSD) of the resulting energy-minimized Ffh-NG and FtsY-NG relative to their original starting structures were levelled off to ± 0.5 nm (5.0 Å) and ± 0.18 nm (1.8 Å), respectively. The RMSD values did not change significantly after 3 ns of simulations (Figure 1A). These RMSD values indicate that the employed simulation time was long enough to obtain an equilibrium and stable structure of Ffh-NG and FtsY-NG. However, Ffh-NG stabilizing at a slightly high 5.0 Å RMSD, indicates that the structure may undergo structural rearrangement and conformational changes after solvation of the protein. Therefore, the compactness of the structure throughout the simulation was monitored by measuring the radius of gyration (R_g). R_g measures the relative mass of the atom(s) and the center of mass of the molecule. Ideally, stably folded proteins are likely to maintain a relatively steady value of R_g, whereas unfolded or collapsed proteins will be indicated by R_g change over time. Analysis of the R_g for Ffh-NG plotted a reasonably constant R_g value (2.0 nm) and this indicates that the protein remains very stable in its compact (folded) form throughout the simulation at 300 K (data not shown).

The stability was further evaluated by computing residues fluctuation (RMSF) in order to probe the dynamics of the Ca atoms at each time point of the trajectories of the Ffh-NG and FtsY-NG (Figure 1B). Higher RMSF values in a certain area indicate greater flexibility during the simulation. The RMSF profiles showed higher RMSF values around residues 1-80 of N-terminal Ffh-NG, indicating that these regions have a larger atomic motion and are more flexible. However, the overall ratio of RMSF calculated has lower flexibility, thus confirming its higher stability.

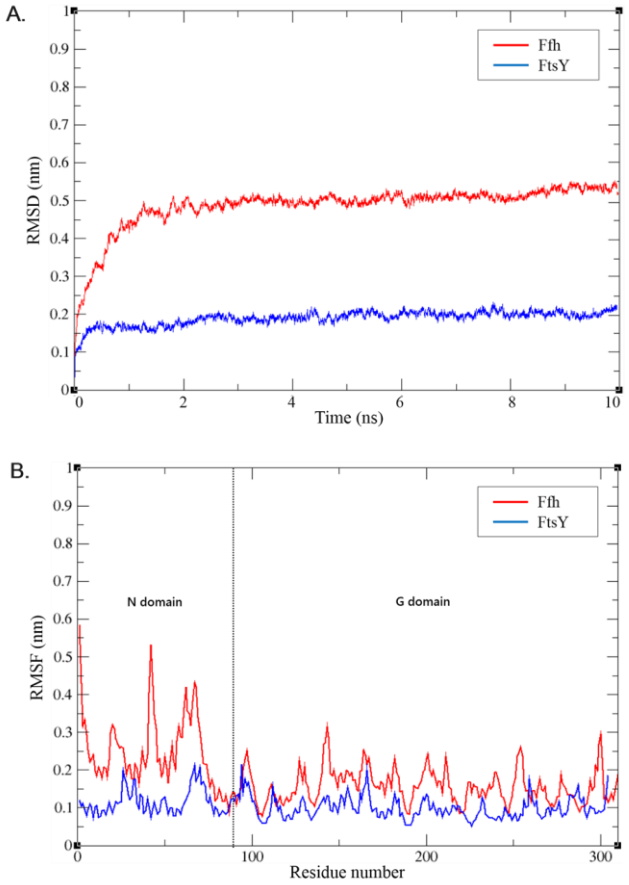


Figure 1 Analysis of MD trajectories of *L. lactis* Ffh-NG and FtsY-NG at 10 ns. (A) Plot of backbone RMSD as a function of time at 300 K displayed a consistent RMSD value. (B) Plot of RMSF of the Ca atoms from the initial structure as a function of residue number at 300 K

For model validation and assessment, results of Ramachandran plot (PROCHECK) was significantly changed after energy minimization (Table 3). Basically, PROCHECK measures residue-by-residue stereochemical quality and overall structure geometry of the model. The Ramachandran plot generated from the PROCHECK analysis signified that 85.8% (Ffh-NG) and 92.6% (FtsY-NG) of the residues

were in the most favoured regions (core) of the plot. Only a relatively low percentage of Ffh-NG (0.9%) and FtsY-NG (0.7%) contain error in these refined models (residues that were in the disallowed region). From the VERIFY3D analysis, the calculated 3D profile score was 81.0% and 99.7% for refined model Ffh-NG and FtsY-NG, respectively. This means that more than 80% of Ffh-NG and FtsY-NG amino acids were correctly positioned and complemented with its 1D-3D profile when the program determines its compatibility. When VERIFY3D score gives a percentage of more than 80%, the predicted model was of satisfactory quality [24].

The Errat scoring for both models were unchanged after energy minimization steps, with an overall quality factor of up to 90%, indicating that atom distribution of the refined models closely resembles atom distribution of available crystal protein structures. A model with Errat score higher than 50% is considered as a high-quality model [38]. Protein structure alignment between Ffh-NG and FtsY-NG models and their corresponding reference structures were found to be very significant with a TM-score value of 0.98 and 0.99, respectively, and RMSD scores below 1.0Å each suggested that the refined model and crystal structure were classified in the same fold [39]. Based on the scoring from these assessment analyses, a reasonable predicted model of *L. lactis* Ffh-NG and FtsY-NG were constructed. The scoring obtained fell in a thermodynamically stable zone and significant stereochemical parameters have been achieved for reliable structures.

3.3 Structural Analysis of 3D Models, Its Conserved Residues and Ffh-FtsY NG Domain Complex

The conserved domains identified in the EBI InterProScan server revealed that three domains can be distinguished in full-length *L. lactis* Ffh (Figure 2), N domain (residues 1-95), G domain (residues 93-299) and M domain (residues 299-431). The N domain is an N-terminal domain which is tightly packed against the G-domain. The G domain (or GTPase domain) is composed of putative GTP-binding sites. The M domain is a C-terminal methionine-rich domain which contains the signal sequence binding site. Both full-length gene of *L. lactis* Ffh and FtsY share a highly homologous N and G domain regions and constitute SRP GTPase, a unique subfamily of G proteins [40,41,42]. An N domain together with G domain forms a structural and functional unit, collectively called the "NG domain". In Ffh (and its SRP54 homologs), the NG domain occurs N-terminally to the M-domain.

In comparison, in FtsY, the NG domain occurs C-terminally to the A-domain. FtsY has an additional N-terminal acidic of A domain, that is thought to be responsible for membrane association of the SRP receptor [43,44]. This NG domain which is conserved between Ffh and FtsY, is a specialized domain that enables it to mediate protein targeting. In this study, 153 residues of *L. lactis* FtsY A domain (M1-E153) were not modelled due to lack of suitable template, and conflicting reports existed for the arrangement of Ffh M domain. Thus, only the NG domains of Ffh and FtsY were used in the subsequent discussion.

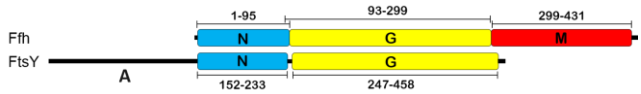


Figure 2 Linear outline of the full-length *L. lactis* Ffh and its receptor protein FtsY. Three domains can be distinguished in Ffh, N: N-terminal domain, that is tightly packed against the G-domain (residues 1-95), G: putative GTP-binding, GTPase domain, indicative of guanine-nucleotide-binding sites (residues 93-299) and M: C-terminal methionine-rich domain, which mediates signal sequence recognition (residues 299-431). In Ffh, the NG domain occurs N-terminally to the M-domain. FtsY has an additional N-terminal membrane-integrated β subunit, acidic "A domain" that is thought to be responsible for membrane targeting of the receptor. In comparison, in FtsY, the NG domain occurs C-terminally to the A domain

L. lactis Ffh-NG and FtsY-NG 3D models adopted all structural features from its reference structure, indicating that both models are an accurate representation of the experimental structures like those of other SRP members (Figure 3).

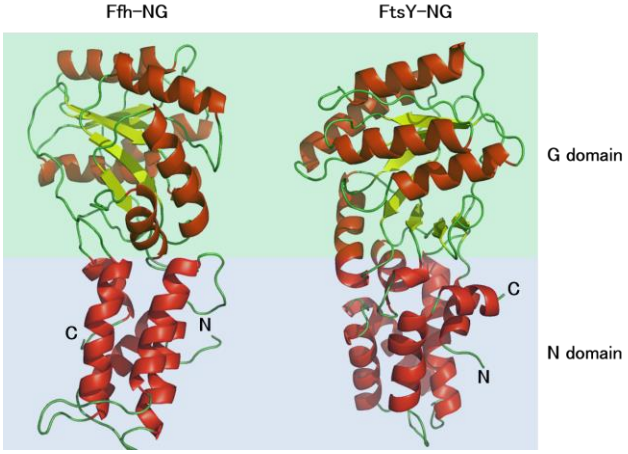


Figure 3 Molecular representation of the refined and energy-minimized model of *L. lactis* Ffh-NG and FtsY-NG final structures after 10 ns of MD simulation at 300 K. The α-helices are coloured in red, the β-strands are coloured in yellow, while the loops/coils are coloured in green. N and C represent N- and C-terminal, respectively. Images were generated using PyMol

The final model exists as a single polypeptide chain with the major secondary structural elements being α-helices. Although the composition of Ffh/SRP54 and FtsY varies in different organisms, the results of motif scanning showed a striking conservation in all reported Ffh/SRP54 and FtsY proteins, in both eukaryotic and prokaryotic (archea) groups. Sequence alignment of the putative

conserved GTPase NG domain and/or motif of *L. lactis* Ffh and FtsY with some experimentally determined homologs from *E. coli* [33], *T. aquaticus* [45], archaea *A. ambivalens* [46], and mammalian *Canis sp.* [32] were found to be highly conserved throughout all GTPases of SRP subfamily, i.e. the ALLEADV motif, motif I (P-loop), motif II (IBD-loop), motif III-IV, DARGG, and closing loop motif (Figure 4).

ClusPro was used to assemble these two unbound Ffh-NG and FtsY-NG models of *L. lactis* into a biologically relevant complex and ideally close to the native structure, Ffh-FtsY NG domain complex (hereafter referred to as Ffh-FtsY NG). These NG domains of Ffh and FtsY were docked against each other as rigid bodies. To ensure exhaustive sampling, 1000 models with varying potential binding conformations of Ffh-NG to FtsY-NG were generated and selected for clustering using the pairwise RMSD as the distance measure and a fixed clustering radius of 10.0 Å was used. Top 30 putative complexes were ranked accordingly to their clustering properties (data not shown). The best near-native conformations calculated by FFT-based docking program with pairwise potentials (the top-ranked model) was taken as the putative binding mode (Figure 5). A ternary complex model of the Ffh-FtsY NG with two molecules of non-hydrolyzable GTP analogue GMPPNP was docked into the active site of Ffh-FtsY NG by referencing the *L. lactis* Ffh-FtsY NG onto the well-studied crystal structure of PDB ID 2J7P [45], 1OKK [47] and 1RJ9 [48].

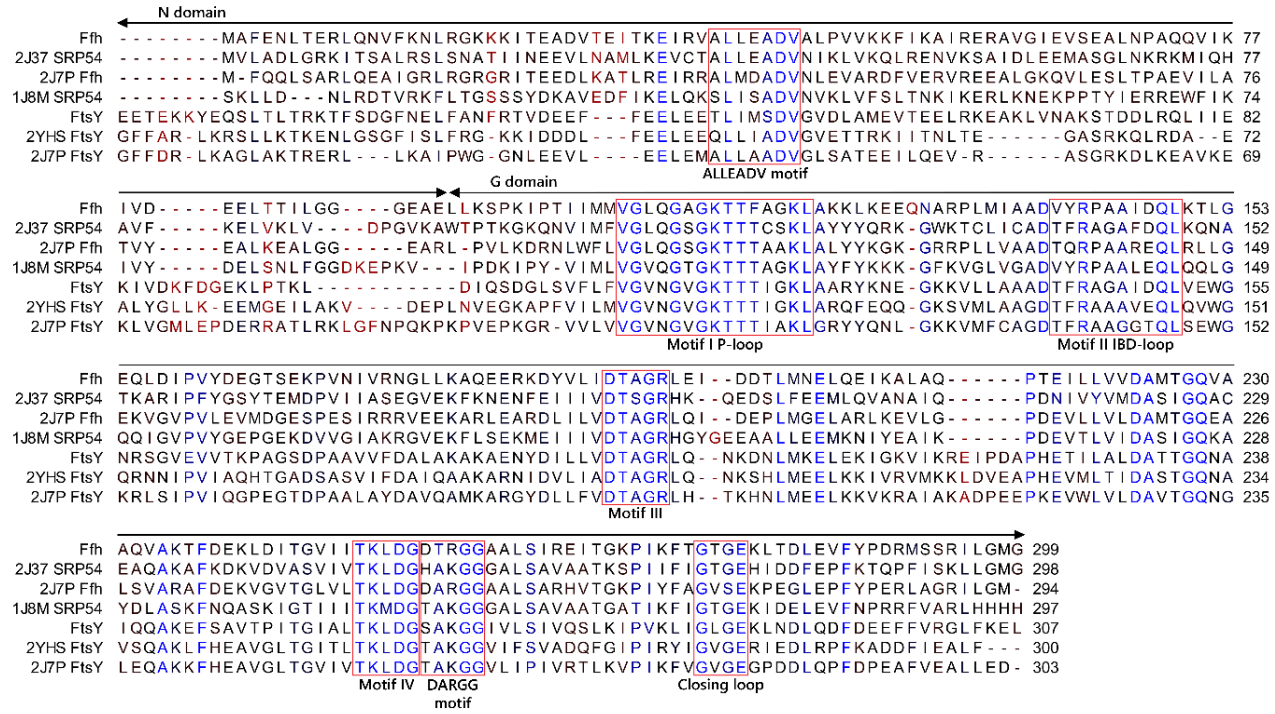


Figure 4 Motif analysis and sequence alignment of *L. lactis* Ffh-NG and FtsY-NG with known structures. Sequences of Ffh/FtsY, from *Canis sp.* (PDB ID 2J37), *T. aquaticus* (PDB ID 2J7P), *A. ambivalens* (PDB ID 1J8M), and *E. coli* (PDB ID 2YHS). Conserved regions (blue), include motifs I to IV (are indicated in Roman numerals), the ALLEADV and DARGG motifs, and the closing loop are shown

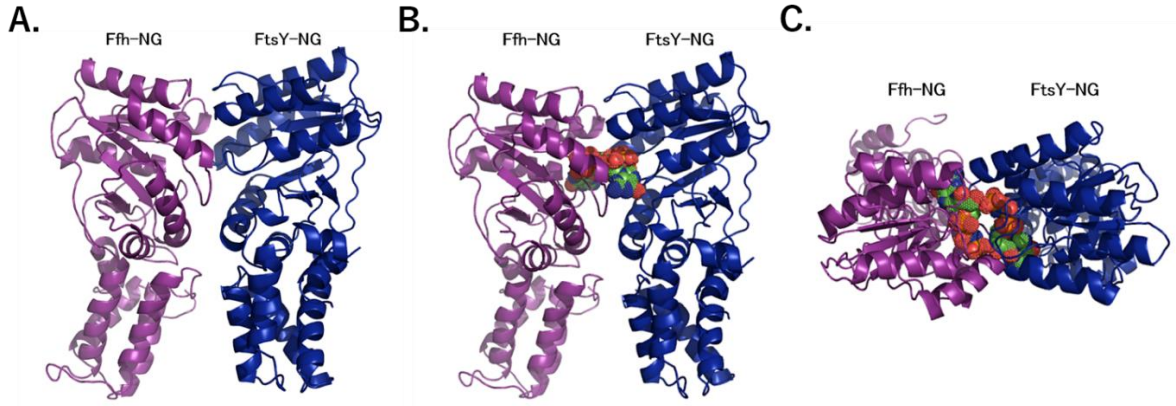


Figure 5 (A) The best-scoring model of nucleotide-free, *L. lactis* Ffh-FtsY NG. Model of GMPPNP-bound, *L. lactis* Ffh-FtsY NG shown in two orientations, front view orientation (B) and top view orientation (C). Two nucleotides (GMPPNP) are shown as space-filled models. The Ffh-NG structure is coloured in the purple and FtsY-NG structure is coloured in blue. Images were generated using PyMol

3.4 Interacting Residues

Identification and analysis of the putative key binding interactions suggested that the interface is stabilized primarily by the extensive pairing of hydrophobic and hydrogen bond interactions, with no disulfide bridges found between *L. lactis* Ffh-NG and FtsY-NG. Recognition of these interactions from the atomic coordinates revealed that the predicted contact residues were mainly confined at the highly conserved motifs II (5' TFRAGAIDQL 3') and III (5' DTAGR 3') in each G domain of *L. lactis* Ffh and FtsY, confirming that complex association is driven by the pairing of G domains. Three amino acid residues of Ffh-NG (Y141, L196 and I198) interact with four amino acid residues of FtsY-NG (V113, F141, L196 and A236) to form four hydrophobic interactions, $V113^{FtsY}$-$I198^{Ffh}$, $F141^{FtsY}$-$Y141^{Ffh}$, $L196^{FtsY}$-$L196^{Ffh}$ and $A236^{FtsY}$-$I198^{Ffh}$. Computationally docked *L. lactis* Ffh-FtsY NG also showed that 11 amino acid residues of Ffh-NG (D139, Y141, R142, A144, K168, G194, E197, I198, D199, E205 and Q232) interact with 12 amino acid residues of FtsY-NG (N111, R142, A143, G144, A165, D168, R195, Q197, N201, E205, N235 and Q239), forming 17 hydrogen bonds (Table 4, Figure 6). A neighbouring residue of motif III (DTAGRLE**I**DDTL), residue Ffh-I198 plays a dual role. It forms two hydrophobic interactions with residues FtsY-V113 and FtsY-A236 and a hydrogen bond with residue FtsY-Q239.

Table 4 FtsY amino acid residues that may interact with Ffh in *L. lactis subsp. cremoris* MG1363 as reported by PIC server. The respective residues involved in interaction type of interest are indicated in bold and underlined. The numbering of amino acid residues starts from the N-terminal NG domain

Residues	Peptide sequences (5' → 3')	Functional region/motif	Type of interaction
$V113^{FtsY}$	VGVNG**V**GKTTTIGKL	Motif I P-loop	
$I198^{Ffh}$	DTAGRLE**I**DDTL	Motif III	
$F141^{FtsY}$	AADT**F**RAGAIDQL	Motif II IBD-loop	
$Y141^{Ffh}$	AADV**Y**RPAAIDQL	Motif II IBD-loop	Hydrophobic
$L196^{FtsY}$	DTAGR**L**QNKD	Motif III	
$L196^{Ffh}$	DTAGR**L**EIDD	Motif III	
$A236^{FtsY}$	TGQN**A**IQQA	(G-domain)	
$I198^{Ffh}$	DTAGRLE**I**DDTL	Motif III	
$R142^{FtsY}$	ADTF**R**AGAIDQL	Motif II IBD-loop	
$D199^{Ffh}$	DTAGRLEI**D**DTLM	Motif III	
$A143^{FtsY}$	DTFR**A**GAIDQL	Motif II IBD-loop	
$E205^{Ffh}$	DTAGRLEIDDTLMN**E**LQEI	Motif III	
$G144^{FtsY}$	TFRA**G**AIDQL	Motif II IBD-loop	
$E205^{Ffh}$	DTAGRLEIDDTLMN**E**LQEI	Motif III	
$Q197^{FtsY}$	DTAGRL**Q**NKDN	Motif III	
$G194^{Ffh}$	IDTA**G**RLEI	Motif III	
$R142^{Ffh}$	ADVY**R**PAAIDQL	Motif II IBD-loop	
$R195^{FtsY}$	DTAG**R**LQNK	Motif III	
$A144^{Ffh}$	VYRP**A**AIDQL	Motif II IBD-loop	
$N201^{FtsY}$	DTAGRLQNKD**N**LMKE	Motif III	
$K168^{Ffh}$	GTSE**K**PVNI	(G-domain)	
$A165^{FtsY}$	VTKP**A**GSDP	(G-domain)	Hydrogen bond
$I198^{Ffh}$	DTAGRLE**I**DDTL	Motif III	
$Q239^{FtsY}$	NAIQ**Q**AKEF	(G-domain)	
$D199^{Ffh}$	DTAGRLEI**D**DTLM	Motif III	
$N111^{FtsY}$	VGV**N**GVGKTTTIGKL	Motif I P-loop	
$R195^{FtsY}$	DTAG**R**LQNK	Motif III	
$D139^{Ffh}$	MIAA**D**VYRPAAIDQL	Motif II IBD-loop	
$N235^{FtsY}$	TTGQ**N**AIQQ	(G-domain)	
$E197^{Ffh}$	DTAGRL**E**IDDT	Motif III	
$N235^{FtsY}$	TTGQ**N**AIQQ	(G-domain)	
$Q232^{Ffh}$	QVAA**Q**VAKT	(G-domain)	
$Y141^{Ffh}$	AADV**Y**RPAAIDQL	Motif II IBD-loop	
$D168^{FtsY}$	PAGS**D**PAAV	(G-domain)	
$R142^{Ffh}$	ADVY**R**PAAIDQL	Motif II IBD-loop	
$E205^{FtsY}$	DTAGRLQNKDNLMK**E**LEKI	Motif III	

Note: Residues are considered to participate in the hydrophobic interactions when they fall within 5.0 Å range of a distance cut-off between apolar groups in the apolar side chains

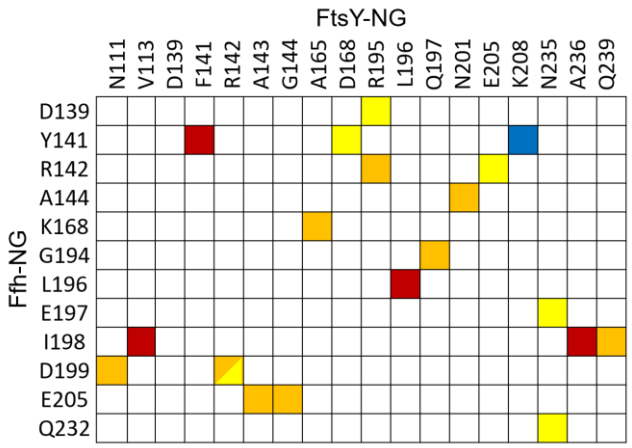


Figure 6 Possible inter-residue contacts presented in a simple matrix diagram. Hydrophobic contacts (dark red), hydrogen bonds (orange for the backbone to side chain and yellow for side chain to side chain), and cation-pi interactions (blue) are shown

This finding, however, when mapped onto the highly homologous structure of *T. aquaticus* complex (PDB ID 1RJ9) revealed fewer number of hydrogen bonds formed between *L. lactis* Ffh-NG and FtsY-NG compared to hydrogen bonds formed in *T. aquaticus* complex [48]. The difference between the *L. lactis* complex model and the *T. aquaticus* crystal structure is at the interface region. In *L. lactis* complex model, two major areas of the interface which contributed to hydrogen bondings lie within the motifs II and III of Ffh-FtsY G domain interface. In contrast, besides motifs II and III of Ffh-FtsY G domain interface, in the *T. aquaticus* crystal structure, interface residues pairing interactions are also present in the highly conserved interface of the ALLEADV motif of its N domain. One possible explanation for the discrepancies between the *L. lactis* complex model and the *T. aquaticus* crystal structure is in *L. lactis* Ffh-FtsY NG model, structure displacements were not allowed during the modelling, treating both unbound Ffh-NG and FtsY-NG as rigid bodies.

Disruption of these interactions may lead to an alteration and conformational rearrangements in *L. lactis* Ffh-FtsY NG model. Although atomic-level precision is hard to achieve in protein docking, by focusing on the G domains of Ffh and FtsY (and its known structural elements important for GTPase activity), Ffh-FtsY NG model suggested several interactions that may be actively involved in the protein-targeting-cycle coupled with GTP hydrolysis. In a stable SRP-SRP receptor complex, biochemical and structural analyses of the bacterial FtsY G domain have proven that the conserved motifs I-V play a critical role in the activation of the GTP hydrolysis because these regions are directly involved in GTP binding and formation of interaction interface [48,49]. Mutagenesis studies on conserved surface residues in *E. coli* FtsY, revealed several residues played a significant role in regulating GTP hydrolysis [47]. As shown in Figure 6, D199^{Ffh-NG}, R195$^{FtsY-NG}$, and E205$^{FtsY-NG}$ interacted with R142$^{FtsY-NG}$ and R142^{Ffh-NG}, both highly conserved residue in motif II (TF**R**AGAIDQL), which was expected to stabilize the gamma-phosphate leaving group before GTP hydrolysis. It was reported that disruption of these interactions was able to block SRP mediated protein targeting and translocation. For example, the hydrogen bond formed between R334$_{FtsY}$ and gamma-phosphate (in *T. aquaticus* complex, PDB ID 1RJ9) could be destroyed by electrostatic repulsion, which may be part of the initiation sequence for hydrolysis of GTP [48, 50].

Apart from G domains of Ffh and FtsY, a similar observation was reported when site-directed mutagenesis was introduced to the conserved Motif I P-loop (GXXGXGK) of N domain of *Streptomyces coelicolor* [51]. Each Lys residue in GXXGXGK of *Sc*-Ffh and *Sc*-FtsY structural model provides the predicted hydrogen bond required for GTP binding. It was reported that mutation of the *Sc*-Ffh (K147G) and *Sc*-FtsY (K228G) significantly decreased the GTPase activity and GTP binding affinity of the proteins [51]. These data suggested *Sc*-Ffh-K147 and *Sc*-FtsY-K228 were important for GTP binding. It was highly likely that each Lys residue provided the predicted hydrogen bond that was required for GTP binding. However, because of the absence of crystal structures in complex with GTP analogues for *Sc*-Ffh and/or *Sc*-FtsY, no strong evidence of this interaction has been observed to date.

4.0 CONCLUSION

This is the first model of Ffh and FtsY NG domain from *L. lactis subsp. cremoris* MG1363 and its Ffh-FtsY NG complex model are similar to that of *T. aquaticus*, suggesting that the regulation of the Ffh-FtsY NG association may follow the same general principle for both. Based on the above findings, the modelled complex forms a composite active site at the interface, primarily on conserved, surface residues of motifs II and III of Ffh and FtsY G domain. These extensive surface interactions are functionally crucial for the initiation or regulation of the SRP protein-targeting-cycle coupled with GTP hydrolysis and stability and/or activity of the complex. All this structural information gained can contribute to unravel protein surface interactions through cross-linking studies and future rational design.

Acknowledgement

This project was supported by the Genomics and Molecular Biology Initiatives Programme of the Malaysia Genome Institute (MGI), Ministry of Science, Technology, and Innovation (MOSTI), Malaysia (Project No. 09-05-MGI-GMB003).

References

[1] Doudna, J. A., Batey, R. T. 2004. Structural Insights into the Signal Recognition Particle. *Annual Review of Biochemistry*. 73: 539-557. DOI:10.1146/annurev.biochem.73.011303.074048.

[2] Matlin, K. S. 2002. The Strange Case of the Signal Recognition Particle. *Nature Reviews Molecular Cell Biology*. 3(7): 538-542. DOI:10.1038/nrm857.

[3] Keenan, R. J., *et al.* 2001. The Signal Recognition Particle. *Annual Review of Biochemistry*. 70: 755-775. DOI:10.1146/annurev.biochem.70.1.755.

[4] Zelazny, A., *et al.* 1997. The NG Domain of the Prokaryotic Signal Recognition Particle Receptor, FtsY, is Fully Functional when Fused to an Unrelated Integral Membrane Polypeptide. *PNAS*. 94(12): 6025-6029. DOI:10.1073/pnas.94.12.6025.

[5] Shepotinovskaya, I. V., Focia, P. J., Freymann, D. M. 2003. Crystallization of the GMPPCP Complex of the NG Domains of *Thermus aquaticus* Ffh and FtsY. *Acta Crystallographica Section D-Biological Crystallography*. 59(10): 1834-1837. DOI:10.1107/S0907444903016573.

[6] Cano-Garrido, O., Seras-Franzoso, J., Garcia-Fruitós, E. 2015. Lactic Acid Bacteria: Reviewing the Potential of a Promising Delivery Live Vector for Biomedical Purposes. *Microbial Cell Factories*. 14(1): 137-148. DOI:10.1186/s12934-015-0313-6.

[7] Ferrer-Miralles, N., Villaverde, A. 2013. Bacterial Cell Factories for Recombinant Protein Production; Expanding the Catalogue. *Microbial Cell Factories*. 12(1): 113-116. DOI:10.1186/1475-2859-12-113.

[8] García-Fruitós, E. 2012. Lactic Acid Bacteria: A Promising Alternative for Recombinant Protein Production. *Microbial Cell Factories*. 11(1): 157-159. DOI:10.1186/1475-2859-11-157.

[9] Konings, W. N., *et al*. 2000. Lactic Acid Bacteria: The Bugs of the New Millennium. *Current Opinion in Microbiology*. 3(3): 276-282. DOI:10.1016/S1369-5274(00)00089-8.

[10] Le Loir, Y., *et al.* 2005. Protein Secretion in *Lactococcus lactis*: An Efficient Way to Increase the Overall Heterologous Protein Production. *Microbial Cell Factories*. 4(1): 2. DOI:10.1186/1475-2859-4-2.

[11] Liang, X., *et al*. 2007. Secretory Expression of a Heterologous Nattokinase in *Lactococcus lactis*. *Applied Microbiology and Biotechnology*. 75(1): 95-101. DOI:10.1007/s00253-006-0809-4.

[12] Morello, E., Poquet, I., Langella, P. 2010. Secretion of Heterologous Proteins, Gram-positive Bacteria, *Lactococcus lactis*. *Encyclopedia of Industrial Biotechnology, Bioprocess, Bioseparation and Cell Technology*. M.C. Flickinger. USA, John Wiley & Sons, Inc.

[13] Bolotin, A., *et al.* 2001. The Complete Genome Sequence of the Lactic Acid Bacterium *Lactococcus lactis ssp. lactis* IL1403. *Genome Research*. 11(5): 731-753. DOI:10.1101/gr.GR-1697R.

[14] Chu, F., *et al.* 2004. Unraveling the Interface of Signal Recognition Particle and Its Receptor by Using Chemical Cross-linking and Tandem Mass Spectrometry. *PNAS*. 101(47): 16454-16459. DOI:10.1073/pnas.0407456101.

[15] Tian, H., Beckwith, J. 2002. Genetic Screen Yields Mutations in Genes Encoding All Known Components of the *Escherichia Coli* Signal Recognition Particle Pathway. *Journal of Bacteriology*. 184(1): 111-118. DOI:10.1128/JB.184.1.111-118.2002.

[16] The UniProt Consortium. 2017. UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Research*. 45(D1): D158-D169. DOI:10.1093/nar/gkw1099.

[17] Altschul, S. F., *et al.* 1997. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research*. 25(17): 3389-3402. DOI:10.1093/nar/25.17.3389.

[18] Söding, J., Biegert, A., Lupas, A. N. 2005. The HHpred Interactive Server for Protein Homology Detection and Structure Prediction. *Nucleic Acids Research*. 33(SUPPL. 2): W244-W248. DOI:10.1093/nar/gki408.

[19] Kelley, L. A., *et al.* 2015. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nature Protocols*. 10(6): 845-858. DOI:10.1038/nprot.2015.053.

[20] Hunter, S., *et al*. 2012. InterPro in 2011: New Developments in the Family and Domain Prediction Database. *Nucleic Acids Research*. 40(Database Issue): D306-D312. DOI:10.1093/nar/gkr948.

[21] Quevillon, E., *et al*. 2005. InterProScan: Protein Domains Identifier. *Nucleic Acids Research*. 33(SUPPL. 2): W116-W120. DOI:10.1093/nar/gki442.

[22] Šali, A., Blundell, T. L. 1993. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*. 234(3): 779-815. DOI:10.1006/jmbi.1993.1626.

[23] Laskowski, R. A., *et al.* 1993. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography*. 26(2): 283-291. DOI:10.1107/S0021889892009944.

[24] Eisenberg, D., Lüthy, R., Bowie, J. U. 1997. VERIFY3D: Assessment of Protein Models with Three-dimensional Profiles. *Methods in Enzymology*. 277: 396-406. DOI:10.1016/S0076-6879(97)77022-8.

[25] Colovos, C., Yeates, T.O. 1993. Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Science*. 2(9): 1511-1519. DOI:10.1002/pro.5560020916.

[26] Zhang, Y., Skolnick, J. 2005. TM-align: A Protein Structure Alignment Algorithm based on the TM-score. *Nucleic Acids Research*. 33(7): 2302-2309. DOI:10.1093/nar/gki524.

[27] Van Der Spoel, D., *et al.* 2005. GROMACS: Fast, Flexible, and Free. *Journal of Computational Chemistry*. 26(16): 1701-1718. DOI:10.1002/jcc.20291.

[28] Schrodinger, L. L. C. 2010. The PyMOL Molecular Graphics System, Version 1.3.

[29] Kozakov, D., *et al*. 2017. The ClusPro Web Server for Protein-protein Docking. *Nature Protocols*. 12(2): 255-278. DOI:10.1038/nprot.2016.169.

[30] Kozakov, D., *et al.* 2006. PIPER: An FFT-based Protein Docking Program with Pairwise Potentials. Proteins. 65(2): 392-406. DOI:10.1002/prot.21117.

[31] Tina, K. G., Bhadra, R., Srinivasan, N. 2007. PIC: Protein Interactions Calculator. *Nucleic Acids Research*. 35(Web Server Issue): W473-476. DOI:10.1093/nar/gkm423.

[32] Halic, M., *et al.* 2006. Following the Signal Sequence from Ribosomal Tunnel Exit to Signal Recognition Particle. *Nature*. 444(7118): 507-511. DOI:10.1038/nature05326.

[33] Stjepanovic, G., *et al*. 2011. Lipids Trigger a Conformational Switch that Regulates Signal Recognition Particle (SRP)-mediated Protein Targeting. *Journal of Biological Chemistry*. 286(26): 23489-23497. DOI:10.1074/jbc.M110.212340.

[34] Kaczanowski, S., Zielenkiewicz, P. 2010. Why Similar Protein Sequences Encode Similar Three-dimensional Structures? *Theoretical Chemistry Accounts*. 125(3-6): 643-650. DOI:10.1007/s00214-009-0656-3.

[35] MacKerell, A. D.Jr., *et al.* 1998. All-atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *Journal of Physical Chemistry B*. 102(18): 3586-3616. DOI:10.1021/jp973084f.

[36] Messaoudi, A., Belguith, H., Ben Hamida, J. 2013. Homology Modeling and Virtual Screening Approaches to Identify Potent Inhibitors of VEB-1 β-lactamase. *Theoretical Biology and Medical Modelling*. 10(1): 22. DOI:10.1186/1742-4682-10-22.

[37] Bruce, A., *et al*. 2015. *The Shape and Structure of Proteins, Molecular Biology of the Cell*. Sixth Edition. Garland Science, Taylor & Francis Group, LLC.

[38] Chaitanya, M., *et al.* 2010. Exploring the Molecular Basis for Selective Binding of *Mycobacterium tuberculosis* Asp Kinase Toward its Natural Substrates and Feedback Inhibitors: A Docking and Molecular Dynamics Study. *Journal of Molecular Modeling*. 16(8): 1357-1367. DOI:10.1007/s00894-010-0653-4.

[39] Xu, J., Zhang, Y. 2010. How Significant is a Protein Structure Similarity with TM-score = 0.5? *Bioinformatics*. 26(7): 889-895. DOI:10.1093/bioinformatics/btq066.
[40] Freymann, D. M., *et al.* 1997. Structure of the Conserved GTPase Domain of the Signal Recognition Particle. *Nature*. 385(6614): 361-364. DOI:10.1038/385361a0.
[41] Montoya, G., *et al*. 1997. Crystal Structure of the NG Domain from the Signal-recognition Particle Receptor FtsY. *Nature*. 385(6614): 365-368. DOI:10.1038/385365a0.
[42] Bourne, H. R., Sanders, D. A., McCormick, F. 1991. The GTPase Superfamily: Conserved Structure and Molecular Mechanism. *Nature*. 349(6305): 117-127. DOI:10.1038/349117a0.
[43] Shan, S. O., Walter, P. 2005. Co-translational Protein Targeting by the Signal Recognition Particle. *FEBS Letters*. 579(4 SPEC. ISS.): 921-926. DOI:10.1016/j.febslet.2004.11.049.
[44] Herskovits, A. A., *et al*. 2001. Evidence for Coupling of Membrane Targeting and Function of the Signal Recognition Particle (SRP) receptor FtsY. *EMBO Reports*. 2(11): 1040-1046. DOI:10.1093/embo-reports/kve226.
[45] Gawronski-Salerno, J., Freymann, D. M. 2007. Structure of the GMPPNP-stabilized NG Domain Complex of the SRP GTPases Ffh and FtsY. *Journal of Structural Biology*. 158(1): 122-128. DOI:10.1016/j.jsb.2006.10.025.
[46] Montoya, G., *et al.* 2000. The Crystal Structure of the Conserved GTPase of SRP54 from the Archaeon *Acidianus ambivalens* and its Comparison with Related Structures Suggests a Model for the SRP-SRP Receptor Complex. *Structure*. 8(5): 515-525. DOI:10.1016/S0969-2126(00)00131-3.
[47] Focia, P. J., *et al*. 2004. Heterodimeric GTPase Core of the SRP Targeting Complex. *Science*. 303(5656): 373-377. DOI:10.1126/science.1090827.
[48] Egea, P. F., *et al.* 2004. Substrate Twinning Activates the Signal Recognition Particle and its Receptor. *Nature*. 427(6971): 215-221. DOI:10.1038/nature02250.
[49] Shan, S. O., Stroud, R. M., Walter, P. 2004. Mechanism of Association and Reciprocal Activation of two GTPases. *PLoS Biology*. 2(10): 1572-1581. DOI:10.1371/journal.pbio.0020320.
[50] Chen, S., *et al*. 2008. A Molecular Modeling Study of the Interaction between SRP-receptor Complex and Peptide Translocon. *Biochemical and Biophysical Research Communications*. 377(2): 346-350. DOI:10.1016/j.bbrc.2008.09.119.
[51] Dong, H. J., *et al*. 2006. Analysis of the GTPase Activity and Active Sites of the NG Domains of FtsY and Ffh from *Streptomyces coelicolor*. *Acta Biochimica et Biophysica Sinica*. 38(7): 467-476. DOI:10.1111/j.1745-7270.2006.00186.x.