Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Graphics and Modelling 22 (2004) 249–262
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Application of the linear interaction energy method (LIE) to estimate the binding free energy values of *Escherichia coli* wild-type and mutant arginine repressor C-terminal domain (ArgRc)–L-arginine and ArgRc–L-citrulline protein–ligand complexes

A.M. Asi [a], N.A. Rahman [b], A.F. Merican [a,*]

[a] *Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia*
[b] *Department of Chemistry, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia*

Received 27 March 2003; received in revised form 21 August 2003; accepted 12 September 2003

Abstract

Protein–ligand binding free energy values of wild-type and mutant C-terminal domain of *Escherichia coli* arginine repressor (ArgRc) protein systems bound to L-arginine or L-citrulline molecules were calculated using the linear interaction energy (LIE) method by molecular dynamics (MD) simulation. The binding behaviour predicted by the dissociation constant (K_d) calculations from the binding free energy values showed preferences for binding of L-arginine to the wild-type ArgRc but not to the mutant ArgRc(D128N). On the other hand, L-citrulline do not favour binding to wild-type ArgRc but prefer binding to mutant ArgRc(D128N). The dissociation constant for the wild-type ArgRc–L-arginine complex obtained in this study is in agreement with reported experimental results [J. Mol. Biol. 235 (1994) 221–230]. Our results also support the experimental data for the binding of L-citrulline to the mutant ArgRc(D128N) [J. Mol. Biol. 279 (1998) 753–760]. These showed that LIE method for protein–ligand binding free energy calculation could be applied to the wild-type and the mutant *E. coli* ArgRc–L-arginine and ArgRc–L-citrulline protein–ligand complexes and possibly to other transcriptional repressor–co-repressor systems as well.

Keywords: Protein–ligand binding free energy; Protein–ligand binding affinity; Force field based simulation; Molecular modelling; Arginine repressor protein; *Escherichia coli* K-12

1. Introduction

The *Escherichia coli* arginine repressor (ArgR) protein is an L-arginine-dependent DNA-binding protein that regulates transcription of genes involved in the biosynthesis of L-arginine. This repressor protein is also required as an obligate accessory protein in Xer site-specific recombination at cer and related recombination sites in natural multicopy plasmids [1–3]. Mutational and crystallographic studies have shown that the N-terminal domain of ArgR (ArgRn) is responsible for DNA binding while the C-terminal domain (ArgRc) is responsible for L-arginine binding and hexamerization [4–7].

ArgR is composed of a dimer of trimers and L-arginine acts as a co-repressor for ArgR. The crystal structure of ArgRc–L-arginine complex showed that six L-arginine molecules are involved in stabilising the homohexameric ArgR [6]. Mutational analysis showed that L-arginine binding is also required for ArgR binding to DNA [4,5]. However, the role of L-arginine molecules in the activation of the repressor to bind to DNA remains unclear. On the other hand, mutants that are defective in trimer–trimer interaction can bind to DNA in an L-arginine independent manner [8].

There are several computational methods that could be used for protein–ligand binding affinity calculations. A popular method for assessing protein–ligand binding affinity is the force field based simulation, either using partitioning method, where the free energy is partitioned into different contributions [9–15] or using non-partitioning method where the free energy of a system is related to the ensemble average of an energy function that described the system [9]. The non-partitioning method such as free energy perturbation (FEP) is the more accurate method to calculate absolute protein–ligand binding free energy. However, it is computationally expensive, time consuming and often subjected to

* Corresponding author. Tel.: +60-3-79674189; fax: +60-3-79674178.
E-mail address: merican@um.edu.my (A.F. Merican).

1093-3263/$ – see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2003.09.003

convergence problems [10,11,13]. Simplified methods of regression approach using predictions based on molecular mechanics [16,17] or empirical scoring functions [18,19] has also been used to calculate protein–ligand binding affinity. This scoring function method calculates the score of a single conformation of protein–ligand complex.

In this study, the linear interaction energy (LIE) method, a linear response semi-empirical approach based on force field simulation developed by Aqvist and co-workers [10,12–15] was used to calculate the binding free energy of the wild-type and mutant ArgRc protein–ligand complexes. The LIE method employs molecular dynamics (MD) simulation for sampling the interaction energies trajectory. In LIE, the binding free energy is estimated using averages of force field energies. LIE method considers the contributions from Lennerd–Jones (LJ) and electrostatic interactions to the total binding energy by the equation given below:

$$\Delta G_{\text{bind}} = \alpha(\langle V_{\text{l–s}}^{\text{LJ}}\rangle_{\text{bound}} - \langle V_{\text{l–s}}^{\text{LJ}}\rangle_{\text{free}}) + \beta(\langle V_{\text{l–s}}^{\text{el}}\rangle_{\text{bound}} - \langle V_{\text{l–s}}^{\text{el}}\rangle_{\text{free}}) \quad (1)$$

where α is the empirical scaling factor for LJ interaction energy and β the scaling factor for electrostatic interaction energy. $\langle V_{\text{l–s}}^{\text{JL}}\rangle_{\text{bound}}$ denotes the LJ interaction energy contribution between the bound ligand and its surrounding environment (i.e. water molecules and protein atoms) and $\langle V_{\text{l–s}}^{\text{JL}}\rangle_{\text{free}}$ the LJ interaction energy contribution between free-state ligand and its surrounding environment, $\langle V_{\text{l–s}}^{\text{el}}\rangle_{\text{bound}}$ represents the electrostatic contribution between the bound ligand and its surrounding environment while $\langle V_{\text{l–s}}^{\text{el}}\rangle_{\text{free}}$ is the electrostatic interaction energy contribution between the free-state ligand and its surrounding environment.

The binding free energy values of protein–ligand interactions for the wild-type complexes ArgRc–L-arginine and ArgRc–L-citrulline, and the mutant complexes ArgRc-(D128N)–L-arginine and ArgRc(D128N)–L-citrulline were studied. ArgRc(D128N) is a mutant protein obtained by side-directed mutagenesis in which Asp128 was substituted with asparagine [20]. By isothermal titration calorimetry, in contrast to the wild-type ArgR, the mutant protein ArgRc(D128N) was found to bind to L-citrulline more strongly than to L-arginine [20]. The simulations in this study were carried out to investigate the effect of specific mutations on protein–ligand binding affinity using the LIE method, in an attempt to understand the structure–function relationship of ArgR.

2. Methodology

The crystal coordinates of the ArgRc–L-arginine complex (PDB accession number 1xxa; at 2.2 Å resolution) [6] was used as the initial conformation for the construction of the mutant ArgRc(D128N) protein system and ligand substitution with L-citrulline molecule. The ArgRc–L-arginine crystal structure complex was used as template and was subjected to Accelrys InsightII Biopolymer module on a Silicon Graphics (SGI) Octane R12000 to generate ArgRc(D128N)–L-arginine, ArgRc–L-citrulline and ArgRc(D128N)–L-citrulline protein–ligand complex structures. Fig. 1 shows the structures of wild-type and mutant ArgRc–L-arginine complexes, focusing on the location of Asp128 and the Asn128 residues in the wild-type and mutant proteins, respectively, in relation to the L-arginine binding site. Mutation of Asp128 to asparagine residue was carried out for all the six ArgRc protein subunits. The structures of L-citrulline and L-arginine are shown in Fig. 2. Substitution of L-arginine with L-citrulline was carried out for all the six L-arginine molecules G1–L1. In these simulations,

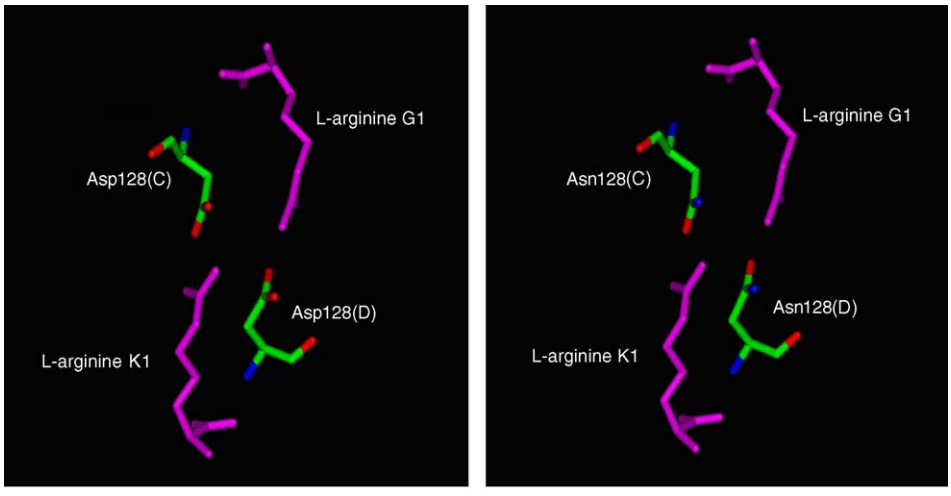


Fig. 1. The wild-type ArgRc–L-arginine and the mutant ArgRc(D128N)–L-arginine complexes. (a) The crystal structure of wild-type ArgRc–L-arginine complex Asp128 of subunits C and D residues, as well as L-arginine K1 which is adjacent to L-arginine G1 molecule, are shown in the figure. (b) ArgRc(D128N) is a mutant protein in which the Asp128 was substituted to asparagine residue. Alphabetical letters in brackets denote the protein subunits.

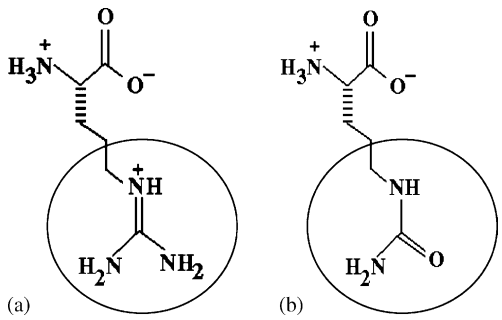


Fig. 2. Structures of L-arginine (a) and L-citrulline (b). Circles highlight the differences between the L-arginine guanidino group and L-citrulline urea moiety.

we chose L-arginine G1 (as defined in the crystal structure; Fig. 3) or L-citrulline G1 (in the case of L-arginine substitution with L-citrulline) as the ligand molecule for the LIE protein–ligand binding free energy calculations since trimer 1 (comprising subunits A–C) has a lower crystal structure B-factor average value as compared to trimer 2 (subunits D–F) (data not shown). L-arginine G1 is adjacent to subunits A and C (Fig. 3).

The L-arginine/L-citrulline G1 alpha carbon was chosen as the centre atom from which an 18 Å spherical boundary was defined for dynamic simulation. The 18 Å spherical boundary was chosen through trial and error modelling of charges on polar amino acid residues in order to obtain an electro-neutral state of the wild-type ArgRc protein atoms. The modelling of charges was also applied to L-arginine molecules which were not defined as the ligand in the simulations, i.e. H, I, J, K and L. The amino acid residues charges of the wild-type ArgRc protein atoms within the 18 Å simulation spherical boundary were modelled to electro-neutral condition by turning off the distant amino acid residues charges within the outermost 3 Å of the simulation spherical boundary and turning on the amino acid residues charges within 15 Å from the centre atom. For the amino acid residues that were switched off during the modelling of the amino acid charges, a correction value of +1.0 kcal mol^{-1} had been included [10] (the value was obtained by Coulomb's law calculation using a uniform dielectric constant of 80, which would be a typical value long-range charge–charge interactions in solvated system).

To take into account for the loss of negative charges by the substitution of Asp128 with asparagine in ArgRc(D128N) mutant protein, Ala147 of subunit A, Ala105 of subunit B and Ala126 of subunit D, located within the outermost 3 Å of the simulation spherical boundary were substituted with aspartic acid residue (Table 1). To counteract for the loss of positive charges caused by L-arginine substitution with L-citrulline (not including the defined ligand atoms of the G1 molecule), Ala94, Pro135 and Ala136 of subunit A as well as Ala109 of subunit E, which are located within the outermost 3 Å of the simulation spherical boundary were substituted with lysine residue (Table 1). The alanine and proline residues were chosen for substitution to minimise major structural changes in the protein conformations because these residues are rarely involved in secondary alpha-helix or beta-strands protein conformation. The outermost 3 Å spherical solute boundary was subjected to a 20 kcal mol^{-1} $Å^{-2}$ polarisational restraint during MD simulation. With these mutations, the overall charges of ArgRc–L-arginine and ArgRc(D128N)–L-arginine protein–ligand complexes were

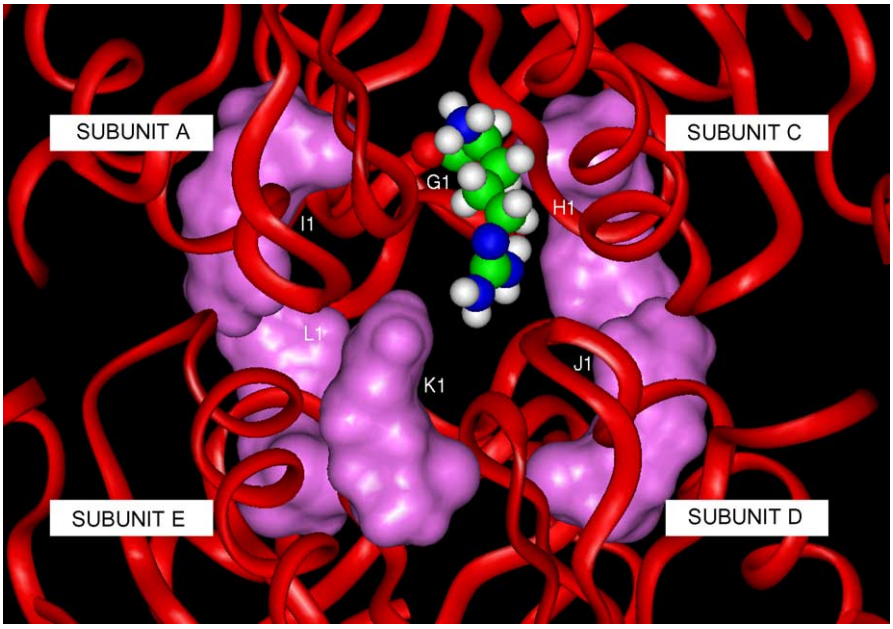


Fig. 3. ArgRc with six L-arginine molecules. L-arginine G1 (shown by CPK representation) was chosen as ligand molecule for LIE protein–ligand binding free energy calculations. L-arginine molecules, H1, I1, J1, K1 and L1 (shown by Connolly surface representation) were included as part of surrounding atoms in MD simulation. L-arginine G1 is adjacent to ArgRc protein subunits A and C and makes protein–ligand trimer–trimer interactions with subunit D.

Table 1
Modelling of amino acid residues charges for molecular dynamics simulation

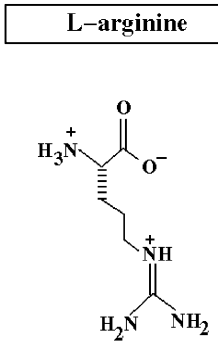
Protein–ligand system	Positive charges within the outermost 3 Å simulation spherical boundary	Negative charges within the outermost 3 Å simulation spherical boundary
ArgRc–L-arginine	Arg 110 (A)	Asp 113 (A)
	Lys 117 (A)	Asp 128 (A)
	Arg 110 (E)	Asp 129 (A)
	L-Arginine (G)	Asp 128 (C)
	L-Arginine (H)	Asp 129 (C)
	L-Arginine (I)	Asp 128 (D)
	L-Arginine (J)	Asp 129 (D)
	L-Arginine (K)	
ArgRc(D128N)–L-arginine	Arg 110 (A)	Asp 113 (A)
	Lys 117 (A)	Asp 128 (A)
	Arg 110 (E)	Asp 147 (A)[b]
	L-Arginine (G)	Asp 105 (B)[b]
	L-Arginine (H)	Asp 129 (C)
	L-Arginine (I)	Asp 126 (D)[b]
	L-Arginine (J)	Asp 129 (D)
	L-Arginine (K)	
ArgRc–L-citrulline	Arg 110 (A)	Asp 113 (A)
	Lys 117 (A)	Asp 128 (A)
	Lys 94 (A)[a]	Asp 129 (A)
	Lys 135 (A)[a]	Asp 128 (C)
	Lys 136 (A)[a]	Asp 129 (C)
	Lys 109 (E)[a]	Asp 128 (D)
	Arg 110 (E)	Asp 129 (D)
ArgRc(D128N)–L-citrulline	Lys 94 (A)[a]	Asp 113 (A)
	Arg 110 (A)	Asp 129 (A)
	Lys 117 (A)	Asp 147(A)[b]
	Lys 135 (A)[a]	Asp 105 (B)[b]
	Lys 136 (A)[a]	Asp 129 (C)
	Lys 109 (E)[a]	Asp 126 (D)[b]
	Arg 110 (E)	Asp 129 (D)

[a]Lys 94, Lys 135, Lys 136 and Lys 109 were originally Ala 94, Pro 135, Ala 136 and Ala 109, respectively.

[b]Asp 147, Asp 105 and Asp 126 were originally Ala 147, Ala 105 and Ala 126, respectively. Overall protein–ligand charges obtained for ArgRc–L-arginine and ArgRc(D128N)–L-arginine were +1, while that of ArgRc–L-citrulline and ArgRc(D128N)–L-citrulline were neutral. Alphabetical letters in brackets denote the protein subunit.

positive one (+1) while the overall charges for ArgRc–L-citrulline and ArgRc(D128N)–L-citrulline protein–ligand complexes were maintained at electro-neutral condition.

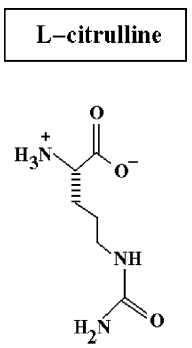
Dynamic simulations were run for L-arginine G1 molecule bound to ArgRc and ArgRc(D128N), L-citrulline G1 molecule bound to ArgRc and ArgRc(D128N) protein systems, and free-states L-arginine/L-citrulline G1 molecule using the Q software package [15]. All simulation systems were solvated with water molecules represented by a rigid single point charge (SPC) model which is subjected to polarisation restraints according to the surface constrained all-atom solvent (SCAAS) model [21]. The wild-type ArgRc–L-arginine simulation included 220 SPC model water molecules while mutant ArgRc(D128N)–L-arginine



Atom no.	name	type	charge
1	N	N.3	0.256
2	HN1	H	0.248
3	HN2	H	0.248
4	HN3	H	0.248
5	CA	C.3H	0.000
6	CB	C.3H2	0.000
7	CG	C.3H2	0.000
8	CD	C.3H2	0.090
9	NE	N.pl3	–0.110
10	HE	H	0.240
11	CZ	C.2	0.340
12	NH1	N.pl3.5	–0.260
13	HH11	H	0.240
14	HH12	H	0.240
15	NH2	N.pl3.5	–0.260
16	HH21	H	0.240
17	HH22	H	0.240
18	C	C.2	0.270
19	O	O.2–	–0.635
20	OXT	O.2–	–0.635

Fig. 4. The set of partial atomic charges assigned to L-arginine molecule.

simulation included 219 SPC model water molecules. The wild-type ArgRc–L-citrulline simulation included 217 SPC model water molecules while mutant ArgRc(D128N)–L-citrulline included 216 model water molecules. All simulations for bound L-arginine and L-citrulline included 18 crystallographic water molecules. Both the simulation of L-arginine G1 and L-citrulline G1 free-states included 787 SPC model water molecules. The systems were assigned with modified GROMOS force field [22,23]. The L-arginine molecular fragment library file was created and a different set of partial atomic charges was assigned (Fig. 4). This is due to the fact that the standard L-arginine molecular fragment in the existing library file was defined as a peptide fragment and hence the library file is not suitable for L-arginine molecule. A similar L-citrulline molecular library file was also created (Fig. 5). L-citrulline urea angles were parameterized based on values provided by Lorna Smith (Lorna Smith, pers. comm.; Tom Hansson, pers. comm.)



Atom no.	name	type	charge
1	N	N.3	0.256
2	HN1	H	0.248
3	HN2	H	0.248
4	HN3	H	0.248
5	CA	C.3H	0.000
6	CB	C.3H2	0.000
7	CG	C.3H2	0.000
8	CD	C.3H2	0.000
9	NE	N.pl3	–0.280
10	HE	H	0.280
11	CZ	C.2	0.383
12	NH1	N.pl3.5	–0.480
13	HH11	H	0.240
14	HH12	H	0.240
15	OH2	O.2	–0.383
16	C	C.2	0.270
17	O	O.2–	–0.635
18	OXT	O.2–	–0.635

Fig. 5. The set of partial atomic charges assigned to L-citrulline molecule.

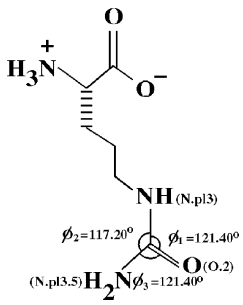


Fig. 6. The angles for L-citrulline urea moiety (Lorna Smith, pers. comm.). The atom-types were specified in brackets.

(Fig. 6). Qprep4 program was run iteratively to determine an 18 Å effective water radius for all simulation systems.

All simulations were run at 1 fs MD timestep. MD simulations were run for 865 ps for the wild-type ArgRc bound state and ligand free-state simulations. MD were run for 1160 ps for the mutant ArgRc(D128N) bound state simulations. The mutant protein–ligand interaction energies took longer time to converge. The first 115 ps MD were run to equilibrate the simulated systems. The simulation started with 1 K temperature at 1 fs temperature bath relaxation time (bath coupling time) and MD data collection stages were carried out at 300 K at 10 fs bath coupling time. All solute atoms (proteins and L-arginine/L-citrulline molecules, G, H, I, J, K and L) were subjected to a low force constraint of 5 kcal mol^{-1} $Å^{-2}$ from the beginning of the equilibration stage to 25 ps of simulation time, after which the constraint applied was removed. Cut-off radius was not applied to intermolecular interaction involving L-arginine/L-citrulline G1 ligand atoms. Cut-off radius of 10 Å was applied to protein atoms and L-arginine/L-citrulline H–L molecules for all simulations. Local reaction field (LRF) method [24] with 10 Å solvent–solvent interactions cut-off radius was used for all simulations. Shake constraints were applied for solvent molecules to keep the solvent molecules bond lengths and angles constant during dynamics simulation. Shake constraint was not applied to solute atoms. The interaction energies sampling and the trajectories sampling were done every 10 and 500 fs, respectively. In order to determine the protein–ligand binding free energy, a simulation on the water-solvated free-state L-arginine G1 and L-citrulline G1 molecules were carried out separately. In both free-state simulations, L-arginine and L-citrulline molecules were subjected to a 10 kcal mol^{-1} $Å^{-2}$ restraints throughout the dynamic simulation.

Methods to estimate the dynamic simulation duration and the standard error for the interaction energies sampling have been described on the standard basis of statistical inefficiency [14]. In this study, the simulations were run and MD data collection stage trajectories were selected for LIE calculation based on the analysis where the LJ interac-

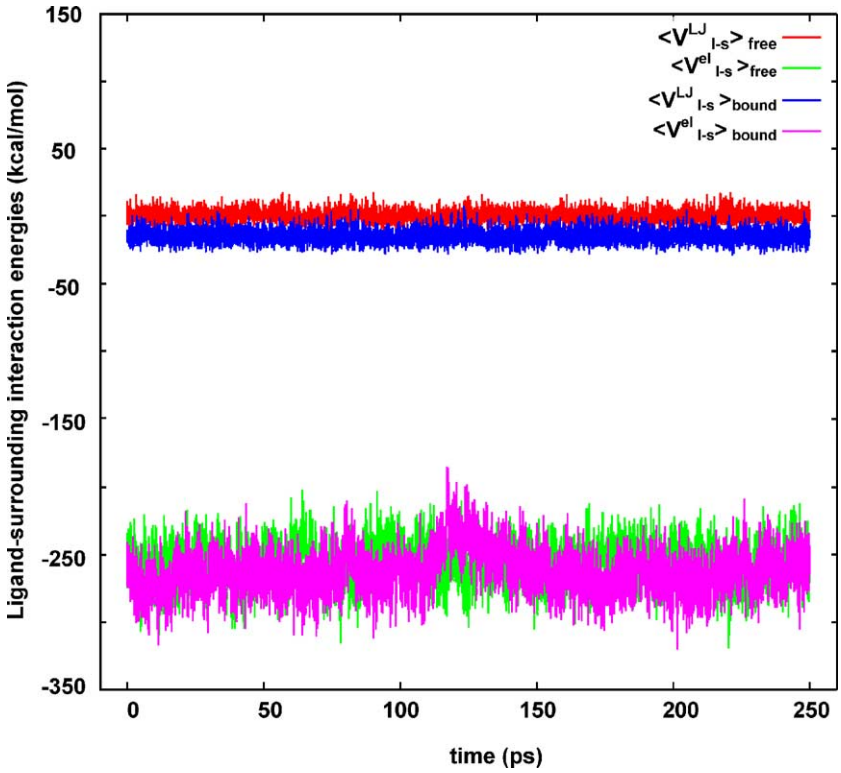


Fig. 7. The LJ and electrostatic interaction energies of L-arginine bound to wild-type ArgRc and L-arginine free-state simulations solvated in water molecules. The plotted 250 ps trajectory data block was well-converged and was used in LIE protein–ligand binding free energy calculation.

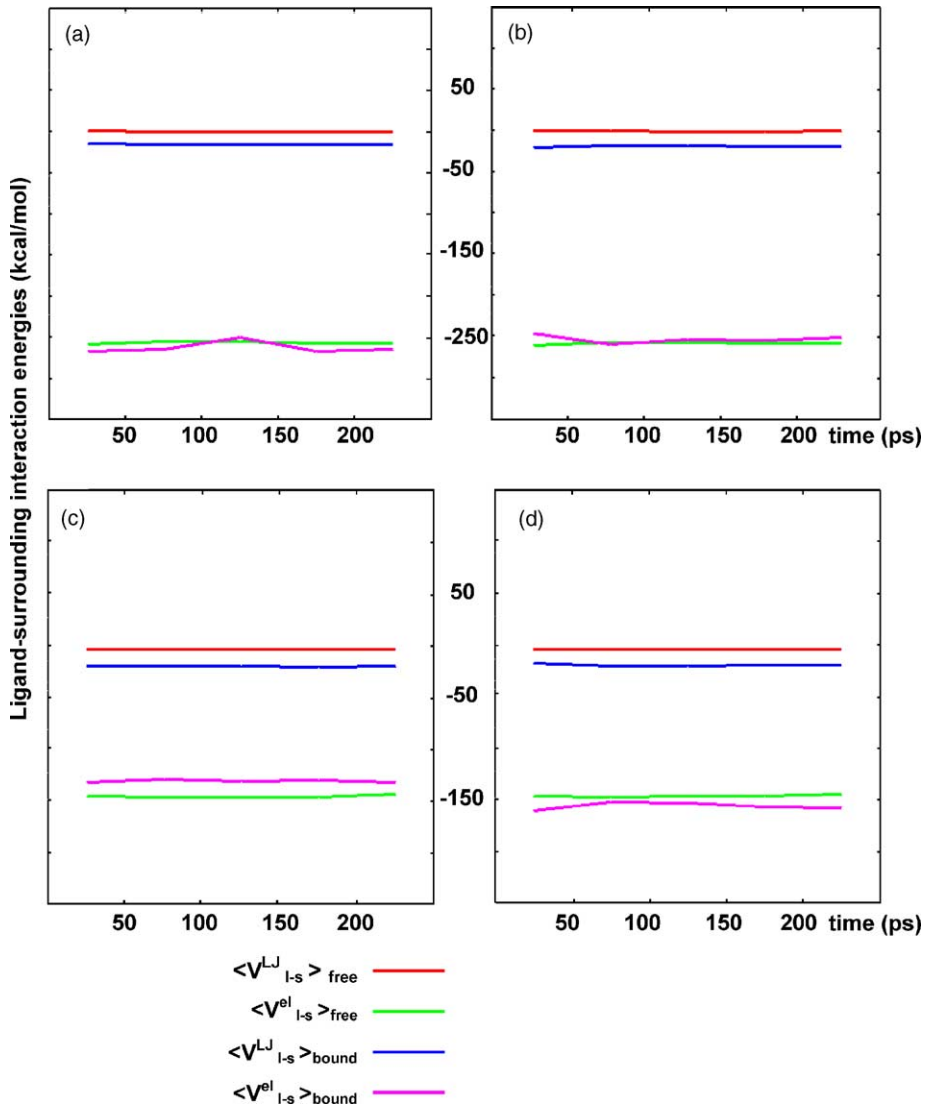


Fig. 8. 5×50 ps averages data of ligand-surrounding electrostatic and LJ interaction energies in protein–ligand bound state and ligand free-state simulations. LJ interaction energy in all protein–ligand bound state simulations were lower than that of L-arginine or L-citrulline in free-state simulations. (a) The interaction energies averages data plots of ArgRc–L-arginine bound state and L-arginine free-state simulations. ArgRc–L-arginine bound state simulation were lower in LJ and electrostatic interaction energies than that of L-arginine in free-state simulation; (b) The interaction energies averages data plots of ArgRc(D128N)–L-arginine bound state and L-arginine free-state simulations. ArgRc(D128N)–L-arginine was higher in electrostatic interaction energy than that of L-arginine in free-state simulation; (c) The interactions energies averages data plots of ArgRc–L-citrulline bound state and L-citrulline free-state simulations. ArgRc–L-citrulline bound state simulation was higher in electrostatic interaction energy than that of L-citrulline in free-state simulation; and (d) The interaction energies data plot of ArgRc(D128N)–L-citrulline bound state and L-citrulline free-state simulations. ArgRc(D128N)–L-citrulline bound state simulation was lower in electrostatic interaction energy than that of L-citrulline in free-state simulation.

tion energy convergence is less than 0.181 kcal mol^{-1} Å^{-2} and the electrostatic interaction energy convergence is less than 0.5 kcal mol^{-1} Å^{-2}. These convergence values were chosen since the LJ and electrostatic interaction energies coefficients used in the LIE equation were 0.181 and 0.5, respectively. In this calculation, a 250 ps block of MD trajectory data was determined using utility programs written in Perl and C scripts where the electrostatic and LJ interaction energies convergence values were estimated from the average difference between the first and the second halves of each single trajectory data block [10] (Tom Hansson, pers. comm). The earliest trajectory block detected was selected for LIE calculation. The source scripts were compiled and run on an SGI Octane R12000.

During selection of the 250 ps MD trajectory block, some of the earlier blocks selected by the program used might have fallen on higher energy level trajectories. This was observed from analysis of the whole time series MD trajectories interaction energies data plots. In this case, the 250 ps trajectory blocks with lower energy level were selected and used for LIE calculations.

3. Results

The electrostatic and LJ interaction energies were plotted using Gnuplot 3.6 (Figs. 7 and 8). The LJ and electrostatic interaction energies in ArgRc–L-arginine simulation (Fig. 7), ArgRc(D128N)–L-arginine, ArgRc–L-citrulline and ArgRc(D128N)–L-citrulline simulations were well-converged. The interaction energies trajectory averages data in Fig. 8a showed the LJ and electrostatic interaction energies for the ArgRc–L-arginine bound state to be lower than that of L-arginine in free-state, suggesting that binding of L-arginine to ArgRc was favoured. The results obtained for the other protein–ligand complexes indicated a lower LJ interaction energy in the bound state compared to ligand in the free-state (Fig. 8b, 8c and 8d). The ArgRc(D128N)–L-arginine and ArgRc–L-citrulline protein–ligand complexes showed a higher electrostatic interaction energy in the simulated bound state as compared to the ligand in free-state. This suggests that the binding of L-arginine to ArgRc(D128N) and the binding of L-citrulline to ArgRc were not favoured (Figs. 8b and c). The differences in electrostatic interaction energy averages data were small in the ArgRc(D128N)–L-arginine bound state compared to L-arginine free-state (Fig. 8b). A more significant differences in the electrostatic interaction energy averages data were observed in the ArgRc–L-citrulline bound state when compared to the L-citrulline free-state (Fig. 8c). The results obtained for the ArgRc(D128N)–L-citrulline protein–ligand complex indicated a lower electrostatic interaction energy in the bound state compared to L-citrulline in the free-state, suggesting binding of L-citrulline to ArgRc(D128N) (Fig. 8d).

The LIE binding free energy values of L-arginine G1 molecule bound to the wild-type ArgRc and the mutant ArgRc(D128N) were calculated using Eq. (1) (Section 1) and the data obtained are summarised in Table 2. In this calculation, the LJ coefficient α value used was 0.181 [10,12–15]. β value of 0.5 was used for protein systems bound to the charged L-arginine molecule, [10] while β value of 0.43 was used for protein systems bound to the uncharged L-citrulline molecule [13]. The values for protein–ligand binding free energy obtained for the wild-type ArgRc–L-arginine and the mutant ArgRc(D128N)–L-citrulline protein systems were -4.757 and $-6.145\,\mathrm{kcal\,mol^{-1}}$, respectively (Table 2), indicating a favourable protein–ligand association. The values for protein–ligand binding free energy obtained for the mutant ArgRc(D128N)–L-arginine and the wild-type ArgRc–L-citrulline mutant protein–ligand complexes were $+0.362$ and $+4.375\,\mathrm{kcal\,mol^{-1}}$, respectively, suggesting poor protein–ligand association.

Estimated dissociation constant (K_d and log K_d) values of ArgRc–L-arginine, ArgRc–L-citrulline, ArgRc(D128N)–L-arginine and ArgRc(D128N)–L-citrulline were also calculated. The estimated K_d value obtained for ArgRc–L-arginine, ArgRc(D128N)–L-arginine, ArgRc–L-citrulline and ArgRc-(D128N)–L-citrulline were 3.428×10^{-4}, 1.835, 1.538×10^{3} and 3.340×10^{-5}, respectively (Table 2). A lower estimated K_d value for the ArgRc(D128N)–L-citrulline mutant complex when compared to the wild-type ArgRc–L-arginine complex indicated that L-citrulline binding to the mutant ArgRc(D128N) mutant protein was stronger than L-arginine binding to the wild-type ArgRc protein.

Further analysis were carried out on selected conformations from the MD trajectories. Average converged structures were obtained from 5×50 ps blocks MD trajectory data used for the LIE protein–ligand binding free energy calculations. Backbone superimposition of L-arginine and L-citrulline molecules from MD trajectory structures showed that L-citrulline molecule bound to the mutant ArgRc(D128N) protein conform to the similar linear shape as the L-arginine molecule bound to the wild-type ArgRc protein (Fig. 9).

To analyse the wild-type and the mutant protein–ligand interactions from MD trajectories, the crystal structure of ArgRc–L-arginine protein–ligand interactions involving ArgRc residues Gln 106, Asp 113, Thr 124, Ala126, Asp 128 and Asp 129 were used (Fig. 10) [6]. The wild-type and mutant protein–ligand interactions could be categorised either as intra-trimer interactions or trimer–trimer interactions. In this case, protein–ligand intra-trimer interaction refers to the protein–ligand interaction that link two protein subunits of the same trimer, while protein–ligand trimer–trimer interaction refers to protein–ligand interaction that link two protein subunits of two opposing trimers. Intra-trimer interactions involve protein residues Gln106,

Table 2
Protein–ligand binding free energy values of wild-type and mutant protein–ligand complexes

Protein–ligand complexes	Lennard–Jones (kcal mol^{-1})	Electrostatic (kcal mol^{-1})	ΔG_{bind} (kcal mol^{-1})	ΔG_{corr} (kcal mol^{-1})	Estimated K_d	Estimated log K_d
ArgRc–L-arginine	-2.781 ± 0.011	-2.976 ± 0.492	-5.757 ± 0.503	-4.757	3.428×10^{-4}	-3.465
ArgRc(D128N)–L-arginine	-3.262 ± 0.020	$+2.624 \pm 0.495$	-0.638 ± 0.516	$+0.362$	1.835	$+0.264$
ArgRc–L-citrulline	-2.970 ± 0.041	$+6.345 \pm 0.194$	$+3.375 \pm 0.235$	$+4.375$	1.538×10^{3}	$+3.187$
ArgRc(D128N)–L-citrulline	-2.782 ± 0.041	-4.363 ± 0.259	-7.145 ± 0.300	-6.145	3.340×10^{-5}	-4.476

The fifth column showed the corrected protein–ligand binding free energy values where a correction value of $+1.0\,\mathrm{kcal\,mol^{-1}}$ obtained by Coulomb's law calculations using a uniform dielectric constant of 80 were included [10]. The last two columns showed the wild-type and mutant protein–ligand complexes estimated values of dissociation constant. The dissociation constant obtained for our analysis is about 10^{-4} which is in agreement with the estimated K_d value of *E.coli* ArgR for arginine [27]. The estimated K_d values obtained indicated that L-citrulline binding to ArgRc(D128N) mutant protein was stronger than L-arginine binding to ArgRc wild-type protein.

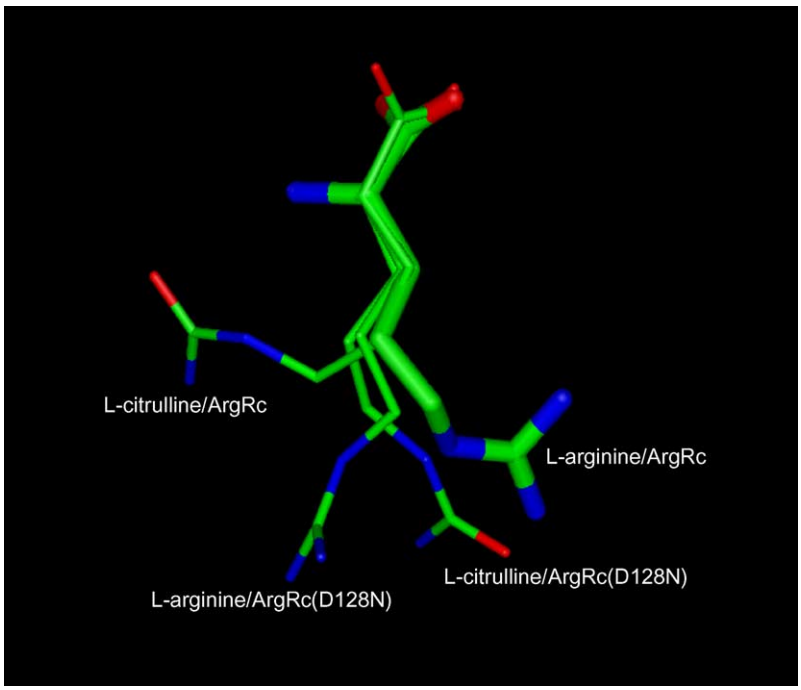


Fig. 9. Backbone superimposed ligand molecules that conformed to specific shapes in interactions with protein atoms as observed from MD trajectories structures. L-Citrulline in mutant ArgRc(D128N)–L-citrulline MD structure conformed to a similar linear shape to that of L-arginine in wild-type ArgRc–L-arginine MD structure.

Asp113, Thr124, Ala126 and Asp129, while trimer–trimer interactions involve wild-type protein residue Asp128 or the substituted mutant protein residue Asn128.

In both ArgRc–L-arginine and the ArgRc(D128N)–L-arginine protein–ligand complexes, L-arginine amino functional group formed interactions with Thr124, Asp129 and Asp113 residues (Figs. 11 and 12). However, no interactions were observed for L-citrulline amino functional group with the above-mentioned residues in both L-citrulline complexed with the wild-type ArgRc or with the mutant ArgRc(D128N) proteins conformations examined (Figs. 13 and 14).

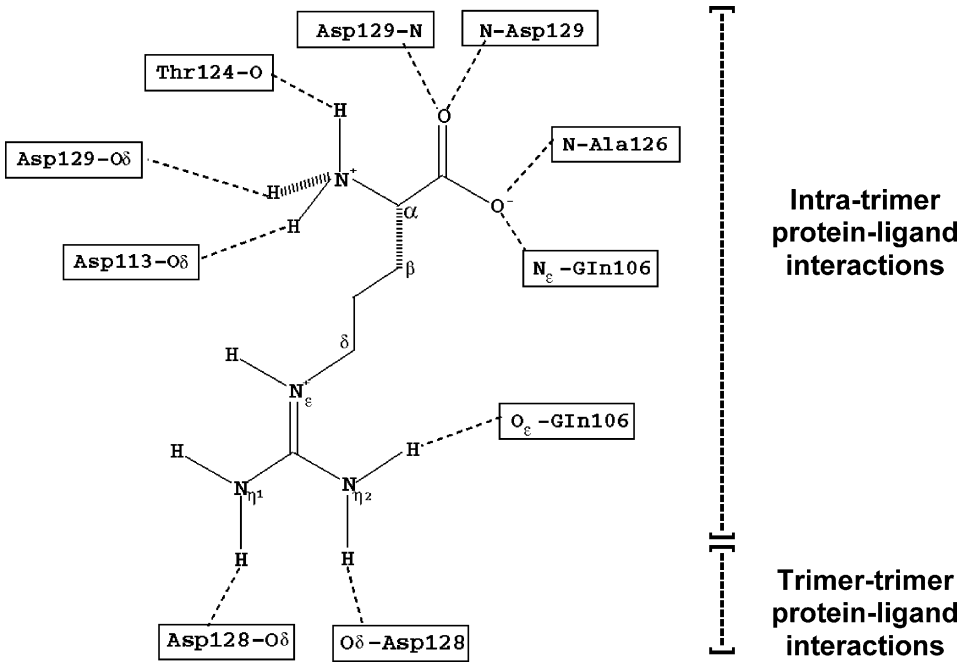


Fig. 10. The ArgRc–L-arginine crystal structure protein–ligand interactions involving ArgRc residues Gln106, Asp113, Thr124, Ala126, Asp128 and Asp129 as reported by Van Duyne and co-workers [6].

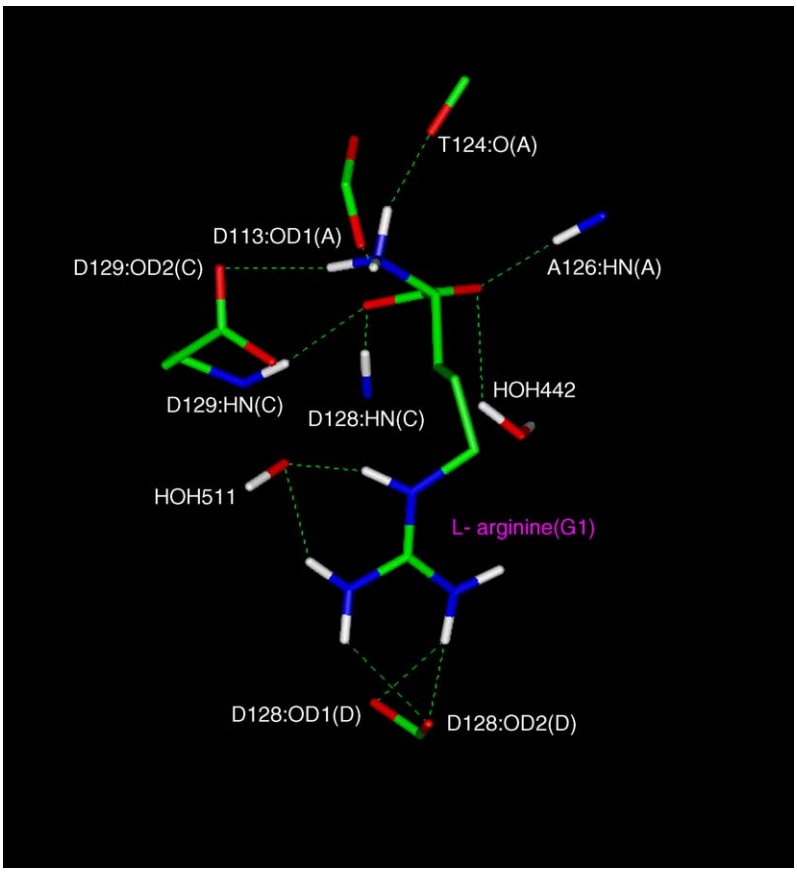


Fig. 11. The wild-type ArgRc–L-arginine protein–ligand molecular interactions. Protein–ligand interactions observed in the ArgRc–L-arginine MD structures were similar to ArgRc–L-arginine crystal structure interactions. Bifurcated hydrogen bond was observed in protein–ligand trimer–trimer interaction involving L-arginine guanidino functional group hydrogen atom and Asp128 side-chain oxygen atoms of ArgRc subunit D of the opposing trimer.

L-arginine and L-citrulline carboxylate functional group were observed to form hydrogen bond interactions with Ala126 and Asp129 residues in all the conformations studied (Figs. 11–14). However, no interactions were observed between L-arginine and L-citrulline carboxylate and guanidino functional groups, as well as L-citrulline urea moiety with Gln106 in all the protein–ligand complexes conformations studied although such interactions were observed in the crystal structure ArgRc–L-arginine complex as shown in Fig. 10.

Bifurcated hydrogen bond interaction where more than one hydrogen bond acceptor group are present, was observed in the trimer–trimer interaction involving L-arginine guanidino functional group hydrogen atom to Asp128 side-chain oxygen atoms of ArgRc subunit D of the opposing trimer (Fig. 11). The configuration of such bifurcated hydrogen bond interaction mentioned above has long been observed and are quite common in crystal structures of many biological small molecules, e.g. α-glycine [25]. Trimer–trimer interactions were not observed in the ArgRc(D128N)–L-arginine (Fig. 12) and the ArgRc(D128N)–L-citrulline (Fig. 14) protein–ligand complexes. However, in the ArgRc–L-citrulline protein–ligand complex (Fig. 13), a single hydrogen bond between the $-NH_2$ terminal of the L-citrulline urea moiety and the side-chain of ArgRc Asp128 residue of subunit D was observed in three out of the five MD trajectory conformations examined.

Hydrogen bonds were observed between L-arginine guanidino functional group and water molecules in the mutant ArgRc(D128N)–L-arginine complex (Fig. 12). L-citrulline urea moiety oxygen was observed to form hydrogen bonds with water molecules in the wild-type ArgRc–L-citrulline complex (Fig. 13). Interactions with water molecules involving L-citrulline urea moiety, amino and carboxylate functional groups were observed in the mutant ArgRc(D128N)–L-citrulline complex (Fig. 14). The interactions of L-arginine and L-citrulline with water molecules arise due to the lack of protein–ligand hydrogen bond interactions in ArgRc(D128N)–L-arginine, ArgRc–L-citrulline and ArgRc(D128N)–L-citrulline protein–ligand complexes.

Unphysical numbers of hydrogen bonds involving a single L-arginine/L-citrulline carboxylate oxygen atom as shown in Figs. 12 and 14 are indication of possible hydrogen bond interactions that could be involved. The approximation of the InsightII Viewer program used to visualise the hydrogen bond interactions is constrained to geometrical criteria where the distance between the proton on the donor atom and the

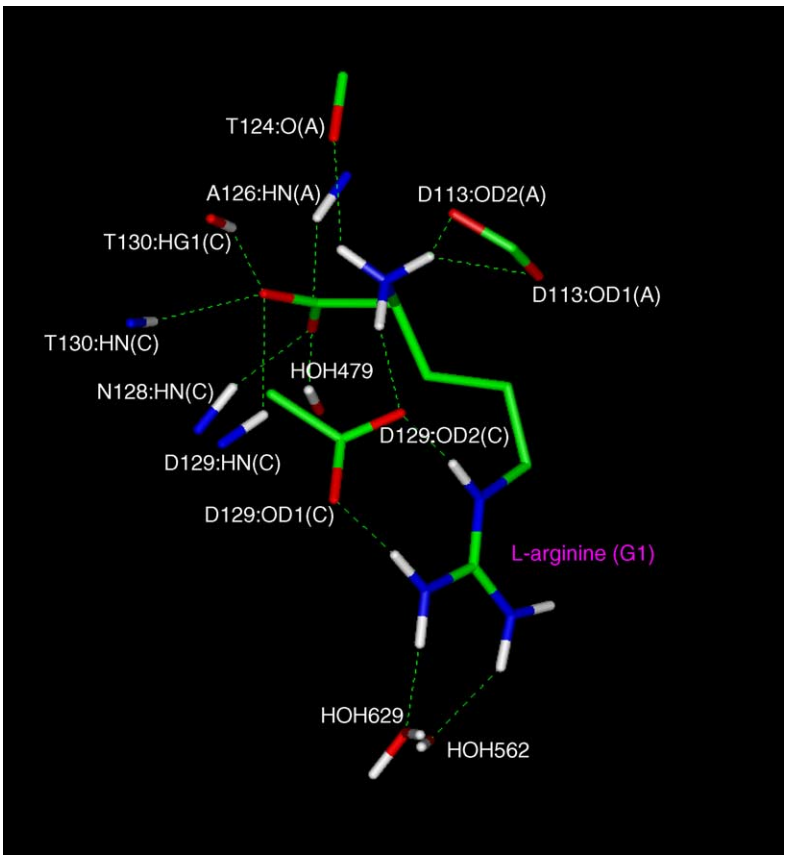


Fig. 12. The mutant ArgRc(D128N)–L-arginine protein–ligand molecular interactions. Similar to ArgRc–L-arginine interactions, L-arginine amino functional group in ArgRc(D128N)–L-arginine complex formed intra-trimer interactions with Thr124, Asp129 and Asp113 residues, while L-arginine carboxylate functional group formed intra-trimer interactions with Ala126 and Asp129 residues. Protein–ligand trimer–trimer interactions were not observed, however, L-arginine guanidino functional group formed hydrogen bonds with water molecules.

heavy atom acceptor must be less than 2.5 Å and the angle between the heavy atom donor, the proton, and the heavy atom acceptor must be between a minimum angle of 120 and 180° [26].

4. Discussion

The binding free energy values of ArgRc–L-arginine, ArgRc(D128N)–L-arginine, ArgRc–L-citrulline and ArgRc-(D128N)–L-citrulline protein–ligand complexes using the LIE method by MD simulation were calculated in order to predict the binding behaviour of L-arginine G1 and L-citrulline G1 molecules to the wild-type and mutant ArgRc proteins. There are several factors to be considered in order for the LIE binding free energy calculations obtained to be consistent with experimental data. The substitution of the charged Asp128 residue with uncharged asparagine residue changed the overall charge of the protein atoms. L-arginine substitution with L-citrulline also changed the overall charge of the protein atoms since the defined non-Q atoms for ArgRc also included the L-arginine molecules H, I, J, K and L. For the wild-type ArgRc–L-arginine protein–ligand LIE binding free energy calculation, the ArgRc protein atoms were modelled to electro-neutral state. For charged ligand molecule such as L-arginine, it is necessary to ensure that both protein–ligand bound state and ligand free-state simulations have the same overall charge to prevent the Born terms from entering into the binding energy calculation [10].

The wild-type ArgRc electro-neutral state model used for ArgRc–L-arginine bound simulation, was used as a template to reproduce the mutant ArgRc(D128N) mutant model where Asp128 was substituted with asparagine residue. At this stage, the net charge of ArgRc(D128N)–L-arginine complex within the 18 Å spherical boundary was +4. Since mutations that changed the overall charge of the protein system resulted in large errors in protein–ligand LIE calculations, substitutions involving distant amino acid residues were carried out (Tom Hansson, pers. comm.). ArgRc(D128N) amino acid residues within the outermost 3 Å spherical boundary, Ala147 of subunit A, Ala105 of subunit B and Ala126 of subunit D residues were each substituted with a charged aspartate residue. These steps were taken to ensure that the ArgRc(D128N) mutant protein system to be not only electro-neutral but also to resemble the wild-type protein

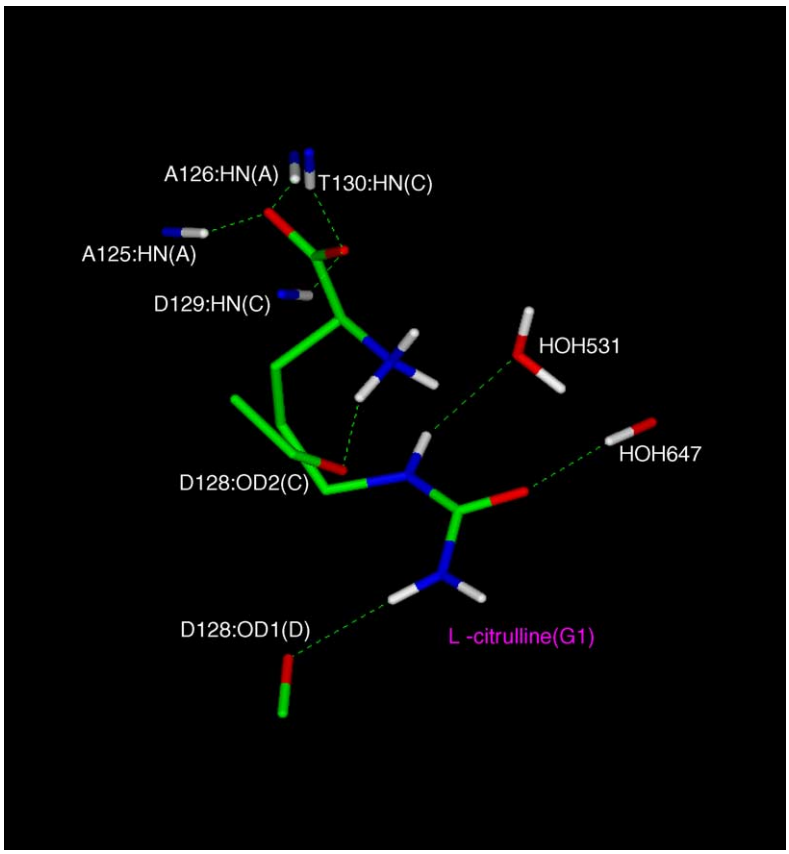


Fig. 13. The wild-type ArgRc–L-citrulline protein–ligand interactions. Intra-trimer interactions between ligand's amino functional group and Thr124, Asp129 and Asp113 residues that were observed in both L-arginine liganded systems, ArgRc–L-arginine and ArgRc(D128N)–L-arginine were not observed in ArgRc–L-citrulline complex. A single trimer–trimer interaction involving L-citrulline urea moiety and the side-chain of ArgRc Asp128 subunit D residue was observed.

system in terms of its charge configuration (within the unrestrained 15 Å spherical boundary).

For LIE calculation involving uncharged ligand molecules such as L-citrulline, the overall charge for bound state and free-state simulations model systems were originally not expected to be the same. However, the simulations carried out did not produce results that were consistent with the experimental data. Therefore, for the L-citrulline liganded systems, the electro-neutral model of the wild-type ArgRc protein system was again used as the template model. The net charge within the 18 Å spherical boundary of ArgRc–L-citrulline complex was −4. Ala94, Pro135 and Ala136 of subunit A as well as Ala109 of subunit E residue were each substituted with a lysine residue. In the ArgRc(D128N)–L-citrulline protein–ligand complex where the net charge was −1, Ala94, Pro135 and Ala136 of subunit A as well as Ala109 of subunit E residue were substituted with lysine, respectively, and Ala147 of subunit A, Ala105 of subunit B and Ala126 of subunit D were substituted with aspartic acid, respectively. A more reliable LIE data were obtained after all the substitution were carried out.

The dissociation constant (K_d) obtained for our analysis was 3.428×10^{-4} which is exceptionally close to the estimated K_d value of *E. coli* ArgR for L-arginine (3.6×10^{-4}) [27]. For ArgRc(D128N)–L-citrulline, Mass and co-workers reported that mutant ArgRc(D128N) protein preferentially bind L-citrulline in the same manner as wild-type ArgR binds L-arginine (no experimental K_d values reported) [20]. In this study, the estimated K_d value obtained for ArgRc(D128N)–L-citrulline was 3.340×10^{-5}, indicating favourable binding of L-citrulline to ArgRc(D128N).

The LIE protein–ligand binding free energy that was obtained for ArgRc–L-arginine complex can be associated with protein–ligand molecular interactions from MD trajectory structures examined. Overall, the protein–ligand interactions observed in the ArgRc–L-arginine MD structures were similar to crystal structure [6] and minimised model structure [28] described for wild-type ArgRc–L-arginine complex. L-citrulline interactions with water molecules were observed in the ArgRc(D128N)–L-citrulline complex, while only a single intra-trimer L-arginine interaction with water molecule was observed in the wild-type ArgRc–L-arginine complex. More L-citrulline interactions with water molecules in the ArgRc(D128N)–L-citrulline complex could be observed when compared to L-arginine and L-citrulline interactions with water molecules in ArgRc(D128N)–L-arginine and ArgRc–L-citrulline protein–ligand complexes, respectively.

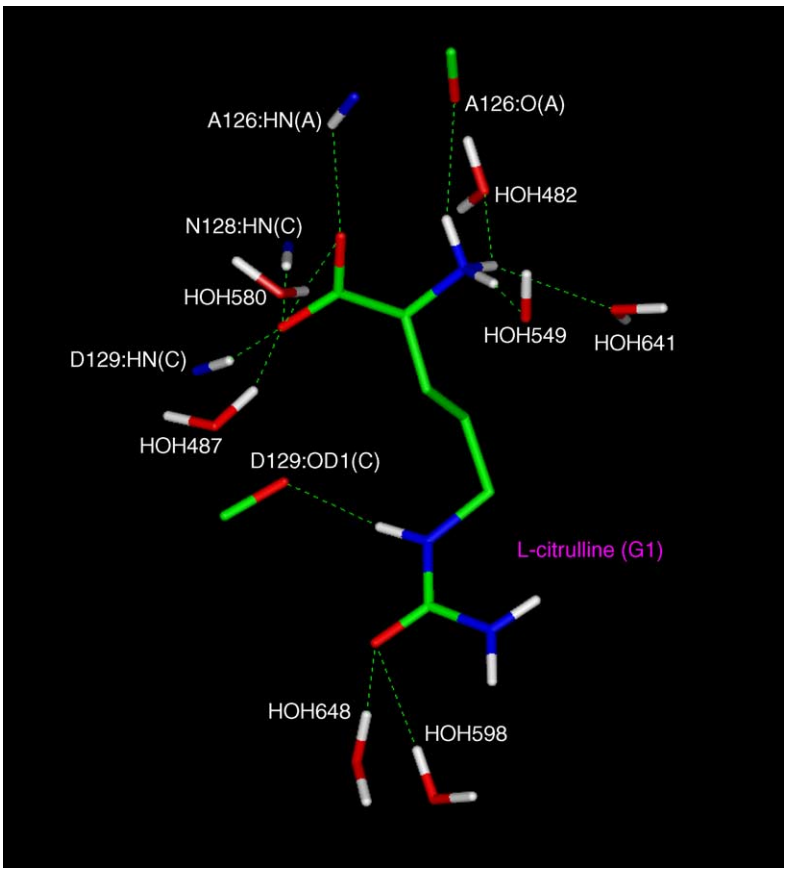


Fig. 14. The mutant ArgRc(D128N)–L-citrulline protein–ligand interactions. More L-citrulline interactions with water molecules in ArgRc(D128N)–L-citrulline complex were observed than L-arginine and L-citrulline interactions with water molecules in ArgRc–L-arginine, ArgRc(D128N)–L-arginine and ArgRc–L-citrulline protein–ligand complexes.

Molecular interactions does not support the energy values obtained from the analysis of LIE protein–ligand binding free energy calculation for ArgRc(D128N)–L-citrulline complex, i.e. no important protein–ligand contacts were observed for ArgRc(D128N)- L-citrulline interaction. Since L-citrulline interactions with water molecules were observed in the ArgRc(D128N)–L-citrulline complex, it is presumed that the L-citrulline non-polar interaction contribution to solvation affected the LIE calculations, resulting in a favourable binding of the ArgRc(D128N)–L-citrulline complex. Van der Waals interaction and non-polar contribution to solvation provided the basis for favourable absolute free energy of complex as reported for theophylline binding to an aptamer [29] and hapten binding to antibody 48G7 [30]. In these studies, the molecular mechanics Poisson–Boltzman surface area (MM–PBSA) [31] method was used where the total binding free energy was separated into electrostatic and van der Waals solute–solute and solute–solvent interactions. Hence, the role of electrostatic in the complexation process can be examined by considering the electrostatic component of molecular-mechanical energy together with the electrostatic contribution to solvation [29–31]. In the LIE method where the contributions from electrostatic and LJ interaction energies to the total binding energy between the ligand and surrounding atoms are considered, the solvation contribution is only involved in the calculation in an indirect manner. The development of LIE method in the future could be improved by taking into account other factors not currently considered, such as solvation effect on the electrostatic contribution.

The binding of L-arginine and it's analogues such as L-citrulline, L-canavanine, L-lysine and L-homoarginine to the wild-type ArgRc protein using docking method was reported [32]. Our results presented in Fig. 11 showed similar wild-type ArgRc–L-arginine interaction pattern as observed in the low estimated free energy of binding conformations from the docking results. Trimer–trimer interactions involving the $–NH_2$ terminal of the L-citrulline urea moiety and the Asp128 side-chain of the opposing trimer residue that was observed in the MD trajectory conformations in this study was also observed in the docking results of L-citrulline binding to the wild-type ArgRc protein system [32].

5. Conclusions

Having analysed the protein–ligand systems of ArgRc–L-arginine, ArgRc(D128N)–L-arginine, ArgRc–L-citrulline and ArgRc(D128N)–L-citrulline, the results obtained

suggest that the mutant protein system ArgRc(D128N)–L-citrulline exhibit better binding than the others. The binding behaviour in protein–ligand systems followed the order of: ArgRc(D128N)–L-citrulline > ArgRc–L-arginine > ArgRc(D128N)–L-arginine > ArgRc–L-citrulline.

The results obtained from the LIE methods were in agreement with reported experimental data and this shows that the LIE method can be applied to estimate the binding free energy of the wild-type and mutant *E. coli* ArgRc–L-arginine and ArgRc–L-citrulline protein–ligand complexes and possibly their homologues in other bacteria [33–40] as well or other repressor–co-repressor protein–ligand complex systems.

Acknowledgements

We especially thank Prof. Johann Aqvist, Dr. Tomas Hansson, Dr Karin Kolmodin and the Q-support team for their suggestions and critical comments of this work. We acknowledge Dr. Lorna Smith of Oxford Centre for Molecular Sciences, University of Oxford for providing us with the urea angle parameters. We acknowledge Assoc. Prof. Dr. Rauzah Hashim for the use of the Computational Chemistry Research Group facilities during the early stage of this study, Mohd. Rosli and Mohd. Talib for advice on the utilities programs and scripts used in this study and Zulkifli Merican of Division of Chemistry, University of New England, Australia for helpful comments. This study is supported by the Biotechnology Top Down Grant under the National Biotechnology Program of the Ministry of Science, Technology and Environment Malaysia (29-02-03-0001).

References

[1] N. Glansdorff, Biosynthesis of arginine and polyamines in *Escherichia Coli* and *Salmonella typhymurium*. Neidhart et al. (Eds.), Cellular and Molecular Biology, second ed., vol. B, American Society for Microbiology, Washington, DC, 1996, pp. 408–433.
[2] C.J. Stirling, G. Szatmari, G. Stewart, M.C.M. Smith, D.J. Sherratt, The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the ColE1 *cer* locus, EMBO J. 7 (1998) 4389–4395.
[3] W.K. Maas, The arginine repressor of *Escherichia coli*, Microbiol. Rev. 58 (1994) 631–640.
[4] M. Burke, A.F. Merican, D.J. Sherratt, Mutant *Escherichia coli* repressor proteins that fail to bind L-arginine, yet retain the ability to bind their their normal DNA-binding sites, Mol. Microbiol. 13 (1994) 609–618.
[5] G. Tian, W.K. Maas, Mutational analysis of the arginine repressor of *Escherichia coli*, Mol. Microbiol. 13 (1994) 599–608.
[6] G.D. van Duyne, G. Ghosh, W.K. Maas, P.B. Sigler, Structure of the oligomerization and L-arginine binding domain of the arginine repressor of *Escherichia coli*, J. Mol. Biol. 256 (1996) 377–391.
[7] M. Sunnerhagen, M. Nilges, G. Otting, J. Carey, Solution structure of the DNA-binding domain and model for the complex of multifunctional hexameric arginine repressor with DNA, Nat. Struct. Biol. 4 (1997) 819–826.
[8] S.H. Chen, A.F. Merican, D.J. Sherratt, DNA binding of *Escherichia coli* arginine repressor mutants altered in oligomeric state, Mol. Microbiol. 24 (1997) 1143–1156.
[9] V. Ajay, M.A. Murcko, Computational methods to predict binding free energy in ligand–receptor complexes, J. Med. Chem. 38 (1995) 4953–4967.
[10] J. Aqvist, Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions, J. Comput. Chem. 17 (1996) 1587–1597.
[11] D.K. Jones-Hertzog, W.L. Jorgensen, Binding affinities for sulfonamide inhibitors with human thrombin using monte carlo simulations with a linear response method, J. Med. Chem. 40 (1997) 1539–1549.
[12] T. Hansson, J. Marelius, J. Aqvist, Ligand binding affinity prediction by linear interaction energy methods, J. Comput.-Aided Mol. Des. 12 (1998) 27–35.
[13] J. Marelius, T. Hansson, J. Aqvist, Calculation of ligand binding free energies from molecular dynamics simulations, Int. J. Quant. Chem. 69 (1998) 77–88.
[14] J. Marelius, M. Graffner-Nordberg, T. Hansson, A. Hallberg, J. Aqvist, Computation of affinity and selectivity: Binding of 2,4-diaminopteridine and 2,4-diaminoquanizoline inhibitors to dihydrofolate reductases, J. Comput.-Aided Mol. Des. 12 (1998) 119–131.
[15] J. Marelius, K. Kolmodin, I. Feierberg, J. Aqvist, Q: A Molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems, J. Mol. Graphics Modell. 16 (1998) 213–225.
[16] A.R. Ortiz, M.T. Pisabarro, F. Gago, R.C. Wade, Prediction of drug binding affinities by comparative binding energy analysis, J. Med. Chem. 38 (1995) 2681–2691.
[17] M.K. Holloway, J.M. Wai, T.A. Halgren, P.M.D. Fitzgerald, J.P. Vacca, B.D. Dorsey, R.B. Levin, W.J. Thompson, L.J. Chen, S.J. deSolms, N. Gaffin, A.K. Ghosh, A.K. Giuliani, S.L. Graham, J.P. Guare, R.W. Hungate, T.A. Lyle, W.M. Sanders, T.J. Tucker, M. Wiggins, C.M. Wiscount, O.W. Woltersdorf, S.D. Young, P.L. Darke, J.A. Zugay, A prior prediction of activity for HIV-1-Protease inhibitors employing energy minimization in the active site, J. Med. Chem. 38 (1995) 305–317.
[18] H.J. Bohm, Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs, J. Comput.-Aided Mol. Des. 12 (1998) 309–323.
[19] R. Wang, L. Liu, L. Lai, Y. Tang, Score: a new empirical method for estimating the binding affinity of a protein–ligand complex, J. Mol. Model. 4 (1998) 379–394.
[20] H. Niersbach, R. Lin, G.D. Van Duyne, W.K. Maas, A superrepressor mutant of the arginine repressor with a correctly predicted alteration of ligand binding specificity, J. Mol. Biol. 279 (1998) 753–760.
[21] G. King, A. Warshel, Investigation of the free energy functions for electron transfer reactions, J. Chem. Phys. 93 (1990) 8682–8693.
[22] W.F. van Gunsteren, H.J.C. Berendsen, Groningen Molecular Simulation (GROMOS) Library Manual, Biomos BV, Groningen, The Netherlands, 1987.
[23] J. Aqvist, M. Fothergill, A. Warshel, Computer simulation of the carbon dioxide–hydrogen carbonate interconversion step in human carbonic anhydrase I, J. Am. Chem. Soc. 115 (1993) 631–635.
[24] F.S. Lee, A. Warshel, A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations, J. Chem. Phys. 97 (1992) 3100–3114.
[25] G.A. Jeffrey, W. Saenger, Hydrogen Bonding in Biological Structures, Springer, Berlin, Hiedelberg, 1991.
[26] Biosym/MSI/Accelrys, InsightII 95.0 Molecular Modeling System User Guide, San Diego, 1995.
[27] G. Tian, D. Lim, J.D. Oppenheim, W.K. Maas, Explanation for different types of regulation of arginine biosynthesis in *Escherichia coli* B and *Escherichia coli* K12 caused by a difference between their arginine repressors, J. Mol. Biol. 235 (1994) 221–230.
[28] A.F. Merican, A.M. Asi, N.A. Rahman, Molecular modeling of C-terminal arginine repressor protein (ArgRc) interactions with its co-repressor L-arginine, AsPac J. Mol. Biol. Biotechnol. 8 (2000) 13–25.

[29] H. Gouda, I.D. Kuntz, D.A. Case, P.A. Kollman, Free energy calculations for theophylline binding to an RNA aptamer: comparison of MM-PBSA and thermodynamic integration methods, Biopolymer 68 (2003) 16–34.
[30] L.T. Chong, Y. Dong, L. Wang, I. Massova, P.A. Kollman, Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7, PNAS 96 (1999) 14330–14335.
[31] J. Srinivasan, T.E. Cheatham, P. Cieplak, P.A. Kollman, D.A. Case, Continuum solvent studies of the stability of DNA, RNA and phosphoramidate–DNA helices, J. Am. Chem. Soc. 120 (1999) 9401–9409.
[32] R. Kueh, N.A. Rahman, A.F. Merican, Computational docking of L-arginine and its structural analogues to C-terminal domain of *Escherichia coli* arginine repressor protein (ArgRc), J. Mol. Model. 9 (2003) 88–98.
[33] A.K. North, M.C. Smith, S. Baumberg, Nucleotide sequence of a *Bacillus subtilis* arginine regulatory gene and homology of its product to the *Escherichia coli* arginine repressor, Gene 80 (1) (1989) 29–38.
[34] C.C.A. Dennis, N.M. Glykos, M.R. Parsons, S.E. Phillips, The structure of AhrC, the arginine repressor/activator protein from *Bacillus subtilis*, Acta Crystallogr. D. Biol. Crystallogr. 58 (2002) 421–430.
[35] C.D. Lu, J.E. Houghton, A.T. Abdelal, Characterization of the arginine repressor from *Salmonella typhimurium and* its interactions with the *carAB* operator, J. Mol. Biol. 225 (1992) 11–24.
[36] R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, B.A. Dougherty, J.M. Merrick, K. McKenny, G. Sutton, W. Fitzhugh, C. Fields, J.D. Gocayne, J. Scott, R. Shirley, L.I. Liu, A. Glodek, J.M. Kelley, J.F. Weidman, C.A. Phillips, T. Spriggs, E. Hedblom, M.D. Cotton, T.R. Utterback, M.C. Hanna, D.T. Nguyen, D.M. Saudek, R.C. Brandon, L.D. Fine, J.L. Fritchman, N.S.M. Geoghagen, C.L. Gnehm, L.A. McDonald, K.V. Small, C.M. Fraser, H.O. Smith, J.C. Venter, Whole genome random sequencing and assembly of *Haemophilus influenzae* Rd, Science 269 (5223) (1995) 496–512.
[37] M. Dion, D. Charlier, H. Wang, D. Gigot, A. Savchenko, J.N. Hallet, N. Glansdorf, V. Sakanyan, The highly thermostable arginine repressor *Bacillus stearmophilus*: gene cloning and repressor-operator interactions, Mol. Microbiol. 25 (1997) 385–398.
[38] N. Jianping, V. Sakanyan, D. Charlier, N. Glansdorf, G.D. van Duyne, Structure of the arginine repressor from *Bacillus stearmophilus*, Nat. Struct. Biol. 6 (1999) 427–432.
[39] A. Rodriguez-Garcia, M. Ludovice, J.F. Martin, P. Liras, Arginine boxes and the *argR* gene in *Streptomyces clavuligerus*: evidence for a clear regulation of the arginine pathway, Mol. Microbiol. 25 (1997) 219–228.
[40] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, B.G. Barrel, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.