Journal of Molecular Graphics and Modelling 29 (2011) 726–739

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Exploring the conformational changes of the ATP binding site of gyrase B from *Escherichia coli* complexed with different established inhibitors by using molecular dynamics simulation
Protein–ligand interactions in the light of the alanine scanning and free energy decomposition methods

Liane Saíz-Urra[a,b], Miguel Angel Cabrera[a], Matheus Froeyen[b,*]

[a] *Chemical Bioactive Center, Central University of Las Villas, Santa Clara, Villa Clara, C.P. 54830, Cuba*
[b] *Laboratory for Medicinal Chemistry, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10 blok X, bus 1030, 3000 Leuven, Belgium*

ARTICLE INFO

Article history:
Received 23 July 2010
Received in revised form 3 December 2010
Accepted 7 December 2010
Available online 17 December 2010

Keywords:
Inhibitors of GyrB
Molecular dynamics
MM-PBSA/MM-GBSA
Computational alanine scanning
Free energy decomposition

ABSTRACT

Currently, bacterial diseases cause a death toll around 2 million people a year encouraging the search for new antimicrobial agents. DNA gyrase is a well-established antibacterial target consisting of two subunits, GyrA and GyrB, in a heterodimer A_2B_2. GyrA is involved in DNA breakage and reunion and GyrB catalyzes the hydrolysis of ATP. The GyrB subunit from *Escherichia coli* has been investigated, namely the ATP binding pocket both considering the protein without ligands and bound with the inhibitors clorobiocin, novobiocin and 5′-adenylyl-β-γ-imidodiphosphate. The stability of the systems was studied by molecular dynamics simulation with the further analysis of the time dependent root-mean-square coordinate deviation (RMSD) from the initial structure, and temperature factors. Moreover, exploration of the conformational space of the systems during the MD simulation was carried out by a clustering data mining technique using the average-linkage algorithm. Recognizing the key residues in the binding site of the enzyme that are involved in the binding mode with the aforementioned inhibitors was investigated by using two techniques: free energy decomposition and computational alanine scanning.

The results from these simulations highlight the important residues in the ATP binding site and can be useful in the design process of potential new inhibitors.

1. Introduction

The increasing worldwide death toll in the world caused by bacterial infection diseases is a very worrying threat to humans nowadays together with the emergence of multi-drug resistant strains owing to overuse of antibiotics [1,2]. Also the side effects shown by many of the antibacterial drugs already established in the market can be very perturbing. The conventional methods of "trial and error" to obtain an appropriate antibacterial lead are very expensive and take a long time to be carried out [3] and once this is achieved, big ADMET (absorption, distribution, metabolism, elimination and toxicity) hurdles still remain.

The significant progresses of genomics and proteomics have deeply changed the approach used for the identification of new targets that are essential for the survival of bacteria. Such targets could deliver drugs with new mechanisms of action without preceding resistance problems. However, a feasible and less resource demanding alternative is to take advantage of existing validated targets and to design novel chemical structures that could overcome known resistance mechanisms for these targets.

DNA gyrase is a well-established antibacterial target. It is an essential, prokaryotic type II topoisomerase with no direct mammalian counterpart which is involved in the vital processes of DNA replication, transcription, and recombination. The enzyme consists of two subunits, GyrA and GyrB, they form a functional heterodimer A_2B_2 (see Fig. 1). The A subunit of DNA gyrase is involved in DNA breakage and reunion while the B subunit catalyzes the hydrolysis of ATP [4].

Inhibitors of GyrB bind to the ATP binding pocket in the N-terminal domain of GyrB and block the access of ATP. As a consequence, the source of energy is terminated and the enzyme loses its activity [4]. Due to the bacterial strains resistant to GyrA inhibitors fluoroquinolones [5,6], which are mainly used for bacterial infection treatments in hospitals and their side effects, there is an emergent interest in designing inhibitors of the ATPase cat-

* Corresponding author. Tel.: +32 16 33 73 79; fax: +32 16 33 73 40.
E-mail address: matheus.froeyen@rega.kuleuven.be (M. Froeyen).

1093-3263/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2010.12.005

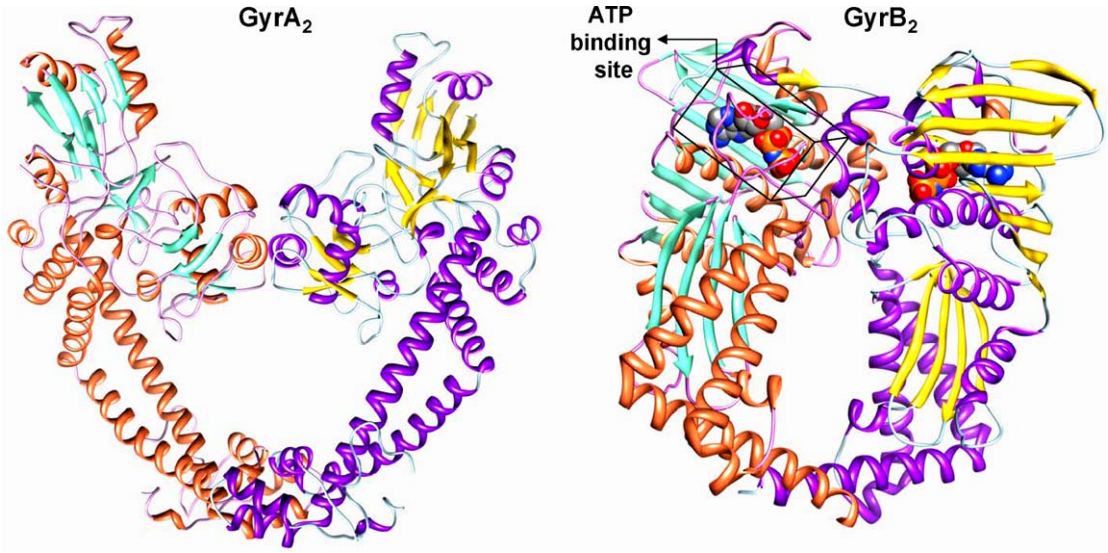


Fig. 1. DNA gyrase heterodimer structure highlighting the ATP binding site in GyrB.

alytic domain. In addition, at this moment no commercial antibiotic targets the B subunit of gyrase.

The strategy of structure-based design has accelerated many drug discovery projects and has already yielded several promising antibacterial leads acting towards different targets, including gyrase and topo IV [7]. Therefore, given the wealth of structural information about bacterial gyrases, the discovery of novel inhibitors of GyrB should not be limited exclusively to the existing chemotypes. A common feature described for the majority of GyrB ATPase inhibitors, is a H-bond interaction pattern between the central fragment (such as adenine, triazine, indazole, different azoles etc.), residue Asp73 and a conserved water molecule [8].

Other "*in silico*" fragment-based studies have identified additional low-molecular-weight fragments that could potentially bind to this central binding pocket of the N-terminal subdomain of GyrB (GyrB24) and subsequently be used for de novo design [9–11]. Furthermore, the performance of various scoring functions for virtual screening has been tested for the GyrB24 target system, where the use of the empirical PLP (piecewise linear potential)-scoring function produced the best results [12,13]. PLP is a simple four-parameter potential that is a piecewise linear approximation of a potential well for hydrogen bonds and lipophilic interactions without angular terms. As a consequence, virtual screening of corporate and/or commercial chemical libraries using the PLP-scoring function could yield even more diverse inhibitors of GyrB. Nevertheless, some of the high-throughput virtual screenings developed, are lacking in validation providing a large number of false positive [14].

Flexibility of the receptor owing to the two loops present in the ATP binding site of the GyrB subunit as well as solvation have been taken into account in the Schechner et al. [10] work. Three different conformations of the GyrB24 subdomain were analyzed and the effects of the solvent were estimated by a continuum dielectric model in the calculation of the binding free energy for every functional group. The results gave consensus maps indicating functional group binding sites that are insensitive to the specific protein conformation.

Recently, Hongtao Yu and Steven W. Rick have reported the use of thermodynamic integration computer simulations to calculate the free energy, enthalpy and entropy for the water molecules involved in hydrogen bonds with inhibitors novobiocin and clorobiocin and polar atoms of certain residues in the binding site of ATP. The entropically unfavorable effect of adding water molecules was demonstrated to be comparable to the difference in the changes of the entropy when novobiocin or clorobiocin bind to the wild type or mutant GyrB. Strong correlations between the thermodynamic parameters (ΔG, ΔH and ΔS) and the change in the hydrogen bonds made when a water molecule is added were found. One of the most important conclusions drawn by the authors was the advantage of designing new drugs for this target modifying the ligands so that, water molecules are removed from the site or the number of hydrogen bonds they can make is reduced [15].

In this work we attempt to explore the conformational changes in the protein accounting for the loops in the active site. For this aim, structures available in the Protein Data Bank (PDB) are used to analyze unliganded and complex structures involving different inhibitors. A "close-up" to the interactions between these ligands and the residues in the binding pocket is put forward in terms of binding energy. Combining results from computational alanine scanning mutagenesis and free energy decomposition methods, the contribution of van der Waals and electrostatic interactions as well as solvation free energy are investigated. Poisson–Boltzmann (PB) and Generalized Born (GB) solvation models are aimed at evaluating the polar component of the free energy of solvation. Moreover, the energy decomposition technique provides the possibility to study the effects of protein backbone and side chain.

2. Material and methods

2.1. Starting structures

Choosing the appropriate starting structures to perform molecular dynamics simulation involves checking many aspects to take into account such as availability, resolution of the structure, missing residues, source organism, mutations and the variety of possible inhibitors in complex with the protein. In this study we focused on the DNA GyrB subunit from *Escherichia coli* since this is one of the bacteria most abundant in the alimentary canal of mammals. In the case of the DNA gyrase subunit B, only three possible protein structures were available. The Protein Data Bank (PDB) codes for these X-ray structures are 1EI1 [16], 1AJ6 [17] and 1KZN [18]. They are complexed with inhibitors of the ATP binding site 5′-adenylyl-β-γ-imidodiphosphate (ADPND), novobiocin (NOV) and clorobiocin (CBN) respectively. A brief description of every structure is given in Table 1. It is worth noting that a correction was applied to the 1KZN file given a mistake in the numbering of the sequence frag-

Table 1
Main features of the protein structures used in the present work*.

Species	Fragment	Inhibitor	PDB entry	Resolution (Å)	Year, revision	Remarks
E. coli	43 kDa, A2–A392, B402–B792, dimer	ADPNP, (ANP)A394, B794	1EI1	2.30	23 Feb 2000, 24 Feb 2009	Mutant Y5S, without Mg ion. Asp198. Closed conformation for loop 2
E. coli	N-terminal 24 kDa, A12–A82, A88–A104, A112–A217	Novobiocin, NOV A1	1AJ6	2.30	15 May 1997, 24 Feb 2009	Mutant R136H. Asn198. Open conformation for loop 2
E. coli	24 kDa N-terminus domain, A15–A82, A88–A100, A116–A219	Clorobiocin, CBN A1	1KZN	2.30	07 Feb 2002, 24 Feb 2009	Wild type. Asn198. Open conformation for loop 2
T. thermophilus	43K domain A9–A392, B9–B392, dimer	Novobiocin, NOV A400, NOV B444	1KIJ	2.30	3 Dec 2001, 24 Feb 2009	Wild type. Open conformation for loop 2

* New gyrase B structures have been deposited into the PDB since writing this manuscript [19].

ment from amino acids 83 to 96. They were changed to 87–100 keeping the nature of the residues.

Because of our interest in the ATP binding site, Domain I of the protein (p24 or GyrB24 subunit) was extracted from every X-ray structure for further calculations, considering the residues in common in all structures for chain A.

Two loops in the GyrB24 subunit, which are involved in the ATPase activity, have been described before [16,18]. Loop 1 encompasses residues 78–86 [18] (*E. coli* numbering) while loop 2 is made up of a larger set of amino acids 98–118 [18]. The main difference in Domain I between the complexes is the conformation of loop 2 which forms a lid over the ATP binding site in the structure with ADPNP and adopts an open and extended conformation in the GyrB24–NOV and GyrB24–CBN complexes. Unfortunately, both loops have not been structurally resolved in the complexes with novobiocin and clorobiocin.

To model loop 1, the structures from 1AJ6 and 1KZN were superimposed with 1EI1 by using the Dali server [20]. Due to the flexibility of the anchor points Ile82 and Val88 in the incomplete crystal structures (1AJ6 and 1KZN), the distances between them and the connector amino acids in 1EI1 (His83 and Gly87) after the superposition were about 1.81 and 2.68 Å respectively. Thus, to insert residues 83–87 would have implied to minimize the resulting structure in a larger extent. For this reason and taking care that the common residues keep the same secondary structure, coordinates for residues 77–90 were inserted into the GyrB24–NOV and –CBN structures.

In view of the fact that the conformation of loop 2 of complexes with NOV and CBN is open, differing from the one with ADPND which is closed, it was necessary to select another structure as template. The crystal structure of the *Thermus thermophilus* GyrB43–NOV dimer displays this open conformation (1KIJ [21], see Table 1). The same procedure explained above was carried out to insert the missing coordinates in both complexes. Residues 100–117 from *T. thermophilus* were inserted in the 1AJ6 and 1KZN structures but changing the numbers into 101–118 (*E. coli* numbering) according to the sequence alignments reported [21]. At the same time mutation of Glu104, Gln105, Gly106 and Ala107 (*T. thermophilus* numbering) was conducted into Asp105, Asp106, Asn107 and Ser108 (*E. coli* numbering) respectively to keep the *E. coli* sequence. The 1AJ6 pdb file was modified, mutating the His136 into Arg136 to obtain the wild type enzyme structure. Finally, the unbound or unliganded enzymes were created by deleting the respectively inhibitors from the pdb files.

2.2. Molecular dynamics simulations

Simulations of eight systems involving the complexed and unliganded enzyme were performed by using the AMBER software (version 10) [22]. Preparation of the ligands implicated applying the Antechamber program. Optimization of the initial geometry of the ligands was carried out with the AM1 approach implemented in the Gamess software [23] as well as calculating their electrostatic potentials at the HF/6-31G(d) level. Charges were generated by using the RESP (restrained electrostatic potential) approach [24].

The general AMBER force field (GAFF) was chosen for the ligands [25] and the ff03 force field for the protein [26,27]. The LEaP module was used to record the topology of the protein and the ligands [22]. All the systems were immersed in a truncated octahedral box of water molecules (TIP3PBOX) [28] keeping at the same time the water molecules from the X-ray experiments. Na^+ ions were added to obtain electrostatic neutrality. The assignment of the protonation state of the histidine residues in the active site was done based on the pKa calculations for the wild type protein reported by Schechner et al. [10], which also has been used by Yu and Rick [15]. The final protonation state was as follows (*E. coli* PDB numbering): 37 HID, 38 HIP, 55 HIE, 64 HIE, 83 HIE, 99 HIP, 116 HIP, 141 HIE, 147 HIP, 215 HIE, and 217 HIP, where HID is the Nδ tautomer, HIE is the Nε tautomer, and HIP is doubly protonated. Histidine 136 in the mutant 1AJ6 (R136H) was considered to be the Nδ tautomer enabling a hydrogen bond with Arg76.

The resulting systems to start simulations with added Na^+ and water molecules are depicted in Table 2, as well as the charged residues present in the protein subunit. The ADPNP ligand had a −4 charge and the coumarins NOV and CBN molecules were neutral.

The systems were minimized by 1000 steps changing the minimization method from steepest descent to conjugate gradient after 500 cycles. Subsequently, the systems were heated during 50 ps from 0 to 300 K, which temperature was kept for the rest of the simulation. To equilibrate the systems a short simulation was run during 600 ps at constant pressure of 1 atm.

The production simulations covered a period of 15 ns in total. The hydrogen atoms were constrained by performing the SHAKE method [29] and the time step was set at 2 fs. No extra restraints were set neither for the protein nor the ligand. The particle-mesh Ewald (PME) procedure [30] was used to estimate the long-range electrostatic interactions and a non-bonded cut-off distance of 8 Å was used. Long-range Lennard–Jones interactions were treated by a continuum model [31].

Periodic boundary conditions were applied during the simulation in all dimensions. Finally, coordinates of the different conformations generated during the simulations were registered every 200 steps (0.4 ps) for a total of 3750 snapshots per system. The same snapshots were used for clustering and energy analysis described in the next sections.

2.3. Clustering analysis

To get insight into exploration of the conformational space of the systems during the MD simulation, a clustering data mining technique was carried out over the 15 ns of trajectory. The structures were taken at 4 ps intervals for a total of 3750. The Average-Linkage

Table 2
Box condition of the simulated systems.

System	# of charged residues	# of Na^+ ions	# of water molecules
1EI1–ADPNP	18 Glu, 15 Asp, 11 Arg, 10 Lys, 5 Hip	11	4858
1EI1 unliganded	18 Glu, 15 Asp, 11 Arg, 10 Lys, 5 Hip	7	4884
1AJ6(R136H)–NOV	17 Glu, 14 Asp, 10 Arg, 11 Lys, 5 Hip	5	5552
1AJ6(R136H) unliganded	17 Glu, 14 Asp, 10 Arg, 11 Lys, 5 Hip	5	4991
1AJ6–NOV	17 Glu, 14 Asp, 11 Arg, 11 Lys, 5 Hip	4	5174
1AJ6 unliganded	17 Glu, 14 Asp, 11 Arg, 11 Lys, 5 Hip	4	4577
1KZN–CBN	18 Glu, 14 Asp, 11 Arg, 10 Lys, 5 Hip	6	6218
1KZN unliganded	18 Glu, 14 Asp, 11 Arg, 10 Lys, 5 Hip	6	5802

Note: Those PDB code written in the table with the abbreviation of the ligand next to it, represent a complex, otherwise it is the unbound enzyme. For 1AJ6, R136H means Arg136 is mutated into His. Systems derived from 1AJ6 structure from PDB used a sequence from Val12 to His217, the other ones included residues from Gly15 to Glu219, which explains the differences in the amino acid count in column 2 of the table.

algorithm [32,33] was chosen to produce from 2 to 20 clusters using the pairwise root-mean-square-deviations (RMSD) between frames as a metric comparing all the atoms from residues of the binding pocket (38, 42–43, 46–47, 49–50, 59, 71–73, 75–80, 82–87, 90–91, 94–95, 97–122, 136, 165–167, *E. coli* numbering, for complexed systems), and in a second study (unbound systems), only the secondary structure of the residues encompassed in the loops (82–87, 97–119) were considered. PDB files were dumped for the average and representative structures from the clusters.

The selection of the appropriate cluster count to be analyzed per system was done on the basis of the metrics values for the obtained clusters. These metrics plotted as a function of the cluster count can help locating the optimal cluster count, as it has been described in a recent article published by Shao et al. [34]. DBI counts for Davies–Bouldin index, pSF for pseudo F-statistic and SSR/SST is the percentage of variance explained, SSR is the sum of squares regression from each cluster and SST is the total sum of squares. The cluster count, pointed out by the majority of the metrics as the optimal one for the system under study, was chosen and the representative structures of these clusters were analyzed.

2.3.1. Structure deviation quantification approach

The visual inspection of the superimposition of the different representative conformations derived from the clustering analysis for every system can be very useful to "see" in an easier way the variation of the systems in time. However, how much do these conformations deviate from the starting point? Is loop 2 more closed or open? To answer those questions we opted for a simple semi-quantitative technique.

To set a protocol that could be followed always for this aim, the Ptraj program implemented in the Amber package version 10 [22] was used to carry out a RMSD calculation with the fit option on. It was needed to superimpose the backbone atoms of the common secondary structures, either α-helix or β-sheets, of the whole proteins to be compared. The fitting RMSD between the starting structures of the unbound enzymes having an open loop 2 conformation, 1KZN and 1AJ6 (the wild type and mutated R136H), was calculated. Then, the same procedure was done but fitting these systems to the 1EI1 structure (with a closed conformation of loop 2). Table 3 shows the results of this first analysis.

Then, the fitted structures were used to estimate threshold values of deviation between the structures with an open and closed conformation of the loops. They are defined as the values of the RMSD estimated for three different conditions of comparison (see Table 4 for final values) and are defined as follows:

i) the backbone atoms of the loop 2 comprising residues 97–119 (in the unbound enzymes).
ii) the residues in the active site (AS) and both loops 1 and 2 considering the atoms in main and side chains (in the complexed structures).
iii) the residues in the active site (AS) and the loops considering the backbone atoms involved (in the complexed structures).

Table 3
Results of the fitting RMSD calculation comparing the different systems.

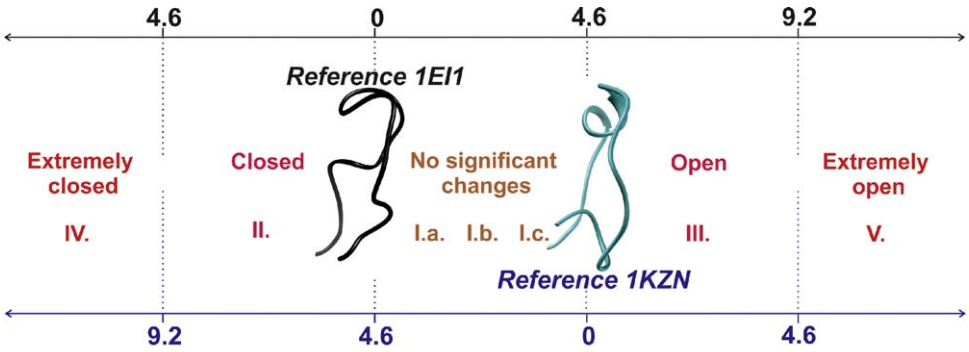
	1EI1	1KZN	1AJ6	1AJ6 (R136H)
1EI1		0.5	0.6	0.6
1KZN			0.3	0.3
1AJ6				
1AJ6 (R136H)				

Once the cluster count was determined for every simulation, a similar procedure was carried out to estimate the deviation of the representative structures of the clusters of the simulations, from the starting (reference) conformations. The fitting between the representative structures of the clusters and reference structures for open and closed conformation of loop 2 (1KZN and 1EI1 respectively) was done. Subsequently, the RMSD of the fitted representative structures with the reference structures was calculated considering the three criteria described above. The results of these calculations were compared to the threshold values from Table 4.

Given the quadratic nature of the RMSD values, it is in principle impossible to know the direction of the deviation of loop 2 during the simulation with respect to the template structure used as reference. Aiming at a better interpretation of these results, a group of equations describing all the possible numerical combinations has been depicted as diagram in Fig. 2, based on RMSD values obtained by criterion *i* as stated higher. RMSD terms are related to the reference used in the calculation and the example, although specific for the case of the unbound enzymes 1EI1 and 1KZN, can be extrapolated to the rest of the analysis applying the corresponding threshold values set before.

Table 4
Threshold values for the deviation between the structures following criteria *i*, *ii* and *iii*.

System and criterion	1KZN			1AJ6			1AJ6(R136H)		
	i	*ii*	*iii*	*i*	*ii*	*iii*	*i*	*ii*	*iii*
1EI1	4.6	4.4	3.0	4.6	4.4	3.0	4.6	4.4	3.0
1KZN				0.1	1.0	0.2	0.1	0.1	0.1



I. a.	4.6 > RMSD(1KZN) > RMSD(1EI1) < 4.6	Closer to IEI1 reference
I. b.	4.6 > RMSD(1KZN) = RMSD(1EI1) < 4.6	Equidistant
I. c.	4.6 > RMSD(1KZN) < RMSD(1EI1) < 4.6	Closer to 1KZN reference
II.	4.6 ≤ RMSD(1KZN) > RMSD(1EI1) < 4.6	More closed than 1EI1 reference
III.	4.6 > RMSD(1KZN) < RMSD(1EI1) ≥ 4.6	More open than 1KZN reference
IV.	9.2 < RMSD(1KZN) > RMSD(1EI1) > 4.6	Extremely closed
V.	4.6 < RMSD(1KZN) < RMSD(1EI1) > 9.2	Extremely open

Fig. 2. Interpretation of the RMSD values derived from comparison between 1KZN and 1EI1 structures based on criterion *i* in Table 4.

2.4. Alanine scanning

In order to define the active site for every bound system under study, we collected information mainly from two important sources in this field, the review published by Oblak et al. [8] and the databases available from the **EMBL-EBI** website [35]. All the residues reported in these sources to be involved in receptor–ligand interactions, the ones that have been studied by site-directed mutagenesis proving to be essential for ATP binding and hydrolysis as well as those residues whose spontaneous mutations prompted coumarin resistance, are listed in Table 5. Moreover, amino acids in a zone of less than 5 Å from the ligand were selected by using the Chimera software [20,36] to guarantee not to miss any possible important protein–ligand interaction.

In general, the computational alanine scanning technique was applied in 30 out of a total of 44 residues to study the effects of mutating these residues into alanine, on the binding free energy. Amino acids equal to alanine or smaller like glycine were discarded. Also, proline residues were not included in the study due to the difference they display in the conformation of the backbone in comparison to alanine [37,38].

As can be observed in the Table 5, the residues to be included in the active site can vary depending on the kind of inhibitor bound. In addition, residues that were not located in a range of 5 Å from the ligand and consequently did not participate in the non-covalent ligand–receptor interactions, were excluded from the post-processing energy calculation. Examples of this are Ser121 and Ala100 for the systems complexed with coumarins.

All the alanine mutant structures were obtained by truncating the residue in question at Cγ in the template systems. Then the modified parameter files were generated again by using the LEaP module [22]. This was extrapolated to the snapshots collected from the trajectories resulting from the MD simulations by using the perl script mm_pbsa.pl implemented in the AMBER package. Descriptions of this procedure and the different input variables have been previously given by Massova and Kollman [38].

2.5. Binding free energy and energy decomposition per residue

Molecular mechanics – Poisson-Boltzman surface area (MM-PBSA) and generalized Born (MM-GBSA) methods were used to calculate the binding free energy in the complexed systems. For this aim the free energy of the complex, receptor and ligand were estimated by the following equations:

$$\Delta G_{\text{bind}} = \Delta G_{\text{comp}} - \Delta G_{\text{rec}} - \Delta G_{\text{lig}} \quad (1)$$

$$\Delta G_{X=\text{comp,rec,lig}} = E_{\text{gas}} + \Delta G_{\text{solv}} - T\Delta S \quad (2)$$

$$E_{\text{gas}} = E_{\text{elect}} + E_{vdW} + E_{\text{intern}} \quad (3)$$

$$\Delta G_{\text{solv}} = G_{\text{PB/GB}} + G_{\text{nonP}} \quad (4)$$

where E_{gas} counts for the absolute energy in the gas phase which encompasses the terms for electrostatic energy (E_{elect}), the van der Waals interaction energy (E_{vdW}) and the internal energy (E_{intern}). In this case the difference in internal energy is zero because a single MD trajectory is used to derive the energy components for the complex as well as receptor and ligand.

The solvation free energy (ΔG_{solv}) includes the contribution of the polar ($G_{\text{PB/GB}}$) and non-polar components (G_{nonP}). As can be seen in Eq. (4) the polar contribution is calculated either by a numerical solver for Poisson–Boltzmann method (PB) or by a generalized Born model (GB) [39,40]. The non-polar component is defined by the following equation:

$$G_{\text{nonP}} = \gamma \cdot \text{SASA} + \beta \quad (5)$$

in which γ and β are empirical constants and they were set in this work as 0.0072 and 0.00 kcal/mol respectively, while SASA is the solvent-accessible-surface area.

For the PB approach, it was needed to set for the electrostatic component calculation the solvent probe radius to 1.4, compatible with the radii values in the prmtop files in AMBER (default). The dielectric constant for the solute was 1 and 80 for the surrounding solvent. The SASA term was estimated by using molsurf [41]. For

Table 5
List of amino acids considered as part of the active site. Residues interacting with the ligand have been labeled with: **a** source is Oblak et al. [8], **b** source is the **EMBL-EBI** website [35]. **c** is related to residues whose mutations cause loss of the ATPase activity or coumarin resistance [8]. **d** Corresponds to amino acids in a range of less than 5 Å from the ligand. Cells in gray stand for residues that are not related to the ligand according to the criteria stated.

Residue ID	1EI1	1KZN	1AJ6
Glu42	a,c,d		
Val43	d	a,b,d	a,d
Asn46	a,b,c,d	a,b,d	a,b,d
Ala47	d	a,b,d	b,d
Glu50	b,c,d	a,b,d	a,b,d
Ile59		d	
Val71		a,b,d	
Gln72		d	
Asp73	a,b,c,d	a,b,c,d	a,b,c,d
Gly75	d	d	d
Arg76	c,d	a,b,d	a,b,d
Gly77	a,b,c,d	c,d	b,c,d
Ile78	a,b,c,d	a,b,c,d	a,b,c,d
Pro79	c,d	a,b,d	a,b,d
Thr80		d	
His83		d	d
Pro84		d	d
Ala90		d	d
Ala91		d	d
Ile94	a,b,d	a,b,c,d	a,b,c,d
Met95		d	d
Ala100	b,d	c	c
Gly101	b,d		
Gly102	a,b,d		
Lys103	a,b,c,d	d	d
Phe104		d	d
Asn107		d	d
Ser108	d		
Tyr109	a,b,d		
Lys110		d	d
Gly113	d		
Gly114	b,d		
Leu115	a,b,d		
His116	a,b,d		
Gly117	a,b,d		
Val118	b,d	d	d
Gly119	a,b,d		
Val120	a,b,d	a,c,d	a,c,d
Ser121	d	c	c
Val122	d		
Arg136/His136		a,b,c,d	a,c,d
Thr165	a,b,c,d	a,b,c,d	a,c,d
Met166		d	
Val167	d	a,d	d

the GB method, the conditions for the calculation of the electrostatic components were the same while the Linear Combinations of Pairwise Overlaps method (LCPO) was used for the SASA [42].

Estimating the contribution of the entropy in the binding energy can be computationally expensive. For this reason we took only 1 out of 1250 snapshots for a total of 30 [43,44]. The nmode module from AMBER was used to estimate the entropy (ΔS), which includes the translational, rotational, and vibrational entropy of the solute. The vibrational modes were obtained under harmonic approximation after energy minimization in a distance-dependent dielectric environment, with $\varepsilon = 4r$.

Another advantage of drawing on MM-PBSA/GBSA approaches implemented in AMBER, is the possibility to calculate the contribution of different fragments from either receptor or ligand to the free binding energy. This is the so-called energy decomposition technique and is based on the estimation of the energy at atomic level. Given that the free energy is an additive property, it is then possible to obtain the contribution of key residues and even to split this into backbone and side chain which can be very useful to get a better insight into their relative importance in the interactions within the complex [43].

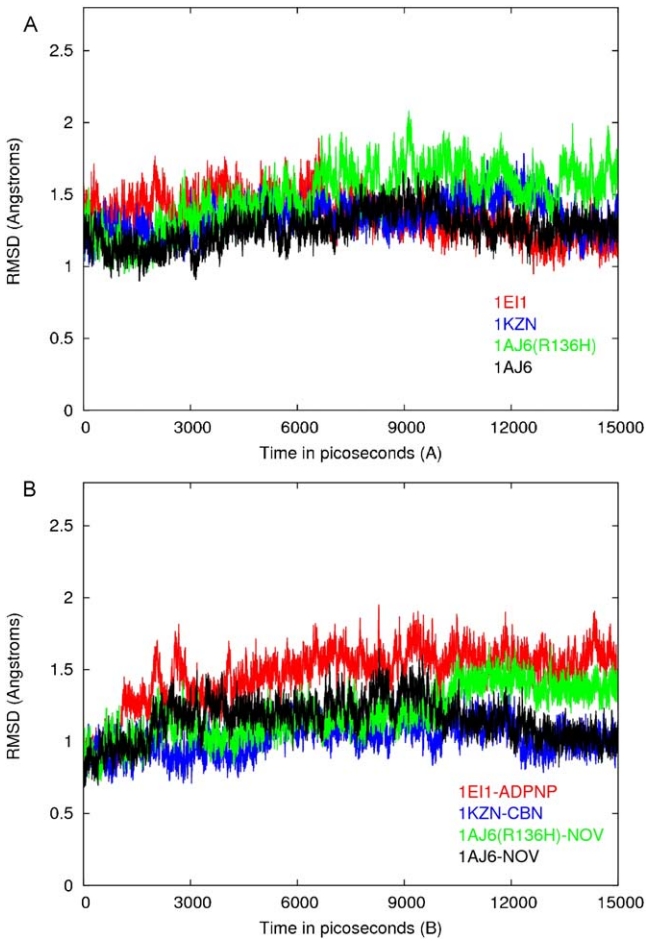


Fig. 3. RMSD plot for the backbone atoms during the MD simulations after the equilibration in every system (A) unliganded and (B) complexed. Plots were done with the Gnuplot tool [45].

For this aim, decomposition was performed on a per-residue basis. We opted for the GB approach to calculate the electrostatic component in the solvation energy and the LCPO for the non-polar one. Once again, the difference in internal energy was not taken into account, according to the reasons aforementioned.

3. Results and discussion

3.1. Molecular dynamics simulation aspects

Root-mean-square-deviations (RMSD) of the MD trajectories resulting from the simulations of the eight systems were calculated to assess the stability in comparison with the respective X-ray structures. Only the backbone atoms (N, CA, C and O) were taken into account, without the two loops due to their flexibility (Fig. 3). In general, the trajectories resulted stable with an average RMSD value of about 1.0–1.75 Å for the unbound systems and 0.7–1.6 Å for the complexed ones. Along the simulations the systems tend to keep a similar behavior. A comparison between the liganded systems involving NOV illustrates how after 9 ns the wild type enzyme is more stable than the mutated protein (R136H). However, this difference finds its origin in the movement of the random coils including residues 172–184 and residues 196–201, not because of the mutation.

The results from the analysis of the temperature factors are shown in Fig. SM1 in the supplementary material. The loops are the most flexible fragments in every system, especially in the unbound

Table 6
Clustering metrics of the selected clusters per system.

System ID	Number of cluster	DBI	pSF	SSR/SST	Critical distance	Cluster size
a Unlinganded						
1EI1	6	1.13	4333.87	0.85	1.85	249, 740, 953, 757, 982, 69
1KZN	3	0.94	3671.14	0.66	2.43	511, 2396, 843
1AJ6	3	0.88	2250.40	0.55	1.99	100, 1453, 2197
1AJ6(R136H)	2	0.67	17536.84	0.82	3.28	738, 3012
b Complexed						
1EI1–ADPNP	2	1.25	9212.23	0.71	2.04	817, 2933
1KZN–CBN	2	1.42	2100.59	0.36	2.40	1285, 2465
1AJ6–NOV	4	2.11	7653.95	0.86	1.39	272, 2437, 850, 191
1AJ6(R136H)–NOV	2	1.67	7460.08	0.67	1.91	1267, 2483

enzymes. RMSD calculations show a higher stability for the complexed system, which might be justified by the influence of possible interactions of the inhibitors with the residues in the active site.

3.2. Clustering results

A clustering based on the average-linkage algorithm [32,33] was carried out. A quantitative assessment of the clustering quality was done based on the metrics pseudo F-statistic (pSF), the Davies–Bouldin index (DBI), the SSR/SST ratio, and the critical distance, to choose the optimal number of clusters per system. Since a detailed explanation about these metrics has been reported before by Shao et al. [34] we will focus our discussion on their application.

DBI is related to the distance of the points inside the cluster to its centroid and the intercluster separation, therefore it is a useful metric that measures the compactness and the separation of the clusters. It is expected that low values of DBI correspond to better clustering. pSF relates the intracluster variance with the residual variance over all points and contrary to DBI, high values suggest better clustering. Another way to identify the optimal number of clusters is by plotting the SSR/SST ratio versus the cluster count where a change in the slope known as "the elbow criterion" indicates the optimal cluster count. Finally, the critical distance, which is defined as the distance between the last clusters split or merged, can illustrate the appropriate cluster count when a steep change in the slope of the plot of the critical distance versus cluster count is found.

In the unbound protein system 1EI1, only DBI provided a very clear pattern that suggested a cluster count of six while for the protein binding ADPNP, DBI and pSF pointed out two clusters. For the protein in the KZN system, DBI, pSF and the SSR/SST show their best value at three clusters. However, the decision is not so easy to make when the CBN ligand is bound to the protein. In this case DBI shows similar values for two, three and four clusters, pSF has a large peak at two clusters and a very discrete one at seven clusters, whereas SSR/SST shows smooth changes in the slope at five and seven clusters, while the critical distance does not show a significant change in the curve. Wild type protein conformations derived from the simulations of system 1AJ6 can be grouped into three clusters according to the consensus of the performance of metrics DBI, pSF and SSR/SST, although the most appreciable change in the critical distance is going from two to three clusters. The NOV ligand affects the number of clusters, increasing this to four based on the DBI value which provides the clearest pattern among the metrics. Finally, the analysis for the mutated protein 1AJ6 system led to select a cluster count of two based on the data for DBI, pSF and a change in the critical distance at two clusters, despite three different elbows are found when plotting SSR/SST at three, six and 15 clusters. Plots of most of the metrics are almost flat when NOV is bound to the mutated protein, only pSF shows a peak at cluster count of two. The values of the different clustering metrics have been collected in Table 6. For more details, a graph of the metrics versus number of clusters is available in the supplementary material section (Fig. SM2).

Given the large amount of data generated from the RMSD calculation based on the three criteria explained above, the resulting values have been stated in the supplementary material in Table SM1 where the significant deviated values have been highlighted in red.

From the clustering results, it can be concluded that in agreement with the results from the temperature factors and the RSMD analysis, the complexed systems remain more stable than the other ones without significant changes in the flexible loops as well as the rest of the residues in the active site. More specifically, conformations that stemmed from the 1EI1 unbound protein simulations tend to be more closed than the starting one and in some cases (clusters four, five and six, Table SM1a) are significantly closed. An illustration of this can be consulted in Fig. 4 where both references have been shown along with a representative protein structure derived from cluster four. For a better appraisal of the implications

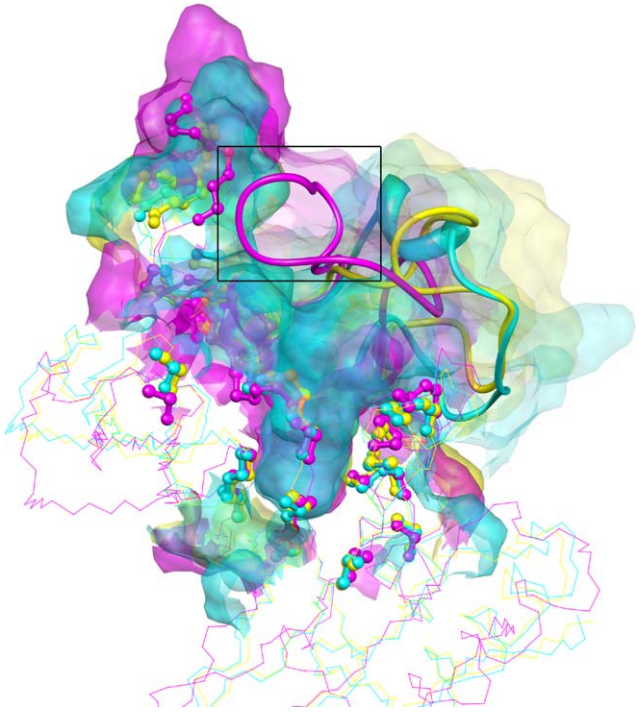


Fig. 4. Superposition of the template enzyme structures 1EI1 (yellow) and 1KZN (cyan) with the representative structure of cluster 4 (in magenta) derived from simulating the 1EI1 unbound enzyme. The surface color is in agreement with the structures. Residues of the active site have been depicted as ball and sticks and the ones involved in the loops as ribbons. The rectangle highlights the surface of loop 2 from the representative structure of cluster 4. The picture was generated by using the Chimera software [20,36].(For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

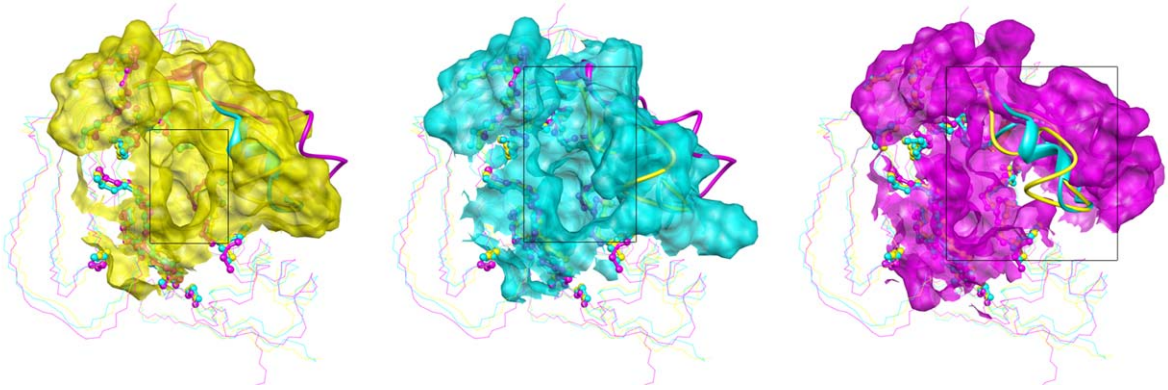


Fig. 5. Superposition of the template enzyme structures 1EI1 and 1KZN with the representative structure of cluster 3 derived from simulating 1KZN unbound enzyme. The colors have been kept as in Fig. 4. Rectangles have been outlined to frame the surface of the entrance of the binding pocket highlighting the changes of the loop 2 from a closed conformation in the 1EI1 structure to the extremely open one in the representative structure of cluster 3 derived from simulating 1KZN unbound enzyme. The picture was generated by using the Chimera software [20,36].

of this extremely closed conformation, the surface of the active site and the loop has been represented in the picture, so that, it can be noticed how the cavity for this cluster is completely closed, disabling the possible use of such a structure in further virtual screening experiments.

The opposite tendency was observed in the conformations resulting from simulating the 1KZN (Table SM1c) and the mutated 1AJ6(R136) unbound systems (Table SM1g) where the tendency is a shift of loop 2 to a more open conformation. However, loop 2 of the wild type protein 1AJ6 is significantly more stable (Table SM1e). For a visual inspection, Fig. 5 can be consulted.

3.3. Binding free energy analysis

According to descriptions found in the literature, DNA GyrB is one of those proteins with ordered water molecules (relative to the liquid) located at the protein–ligand interface that affect the binding affinity for the ligands because of the entropic cost [15,46]. The stability of these water molecules in the active site were analyzed along the simulation trajectory to identify the ones that were stable enough to be included as explicit molecules in the binding free energy calculations.

In the 1EI1–ADPNP simulations the main interactions with water molecules in the protein–ligand interface involved Wat223, Wat221 and Wat3213 where the numbers refer to the Amber numbering adopted by the molecules after running the Leap program. Molecules Wat223 and Wat221 can be identified in the X-ray structure with the PDB number 1601 and 1578 while Wat3213 is the result of immersing the protein in a truncated octahedral box of water molecules (TIP3PBOX). On the other hand, Wat220 (with PDB number 1560) which was expected to form a bridge between residue Glu42(OE1) and ADPNP(O02), was about 30 Å away from both residues, after snapshot 3500 (1400 ps of simulation). Structure 1EI1 taken from the PDB does not contain the counter ion Mg^{2+} as explained before (Table 1). This ion is essential for the orientation of the phosphate groups of the ligand in the active site [8,10,16]. In our simulation, Na^+ ions were added to the structure. The Na^+ 207 was found to be important as part of the interaction net in the active site, including with all the phosphate groups of ADPNP (Fig. SM3-A4 in the supplementary material).

Complexes involving the novobiocin ligand showed similar results in the simulation regarding to the interacting water molecules. Waters 1 and 11, (PDB numbering), were expected to be located between the nitrogen atom of the carbamate group attached to the sugar ring on novobiocin and polar atoms of Val43. A hydrogen bond between Asp73 and carboxylate of the novobiocin has been described as well [15]. However, water 11 was exchanged by water 10 after the equilibration in both systems 1AJ6(R136H)–NOV and 1AJ6–NOV and during the simulation water 1 and 10 swapped their position while water 7 was stable. (Amber number for waters in systems 1AJ6–NOV: Wat1 = Wat212, Wat11 = Wat222, Wat10 = Wat221, Wat7 = Wat218 and 1AJ6(R136H)–NOV: Wat1 = Wat213, Wat11 = Wat223, Wat10 = Wat222, Wat7 = Wat219).

The main difference between wild type GyrB and the R136H mutant structure bound to novobiocin is water molecule 12 (PDB identifier, Wat224 using Amber numbering) that occupies the space left by the absence of the guanidinium group of the arginine residue when it is mutated into a histidine. Conversely, this water molecule was not stable in the active site and it was not possible to include it in the energy calculations.

In the complex 1KZN–CBN the X-ray water molecule 1001 (Wat213 in Amber numbering) was stable in the interaction with Asp73 and the clorobiocin ligand. As it has been reported for this complex, the bulky pyrrole moiety expulses the two water molecules 1 and 10 present in the novobiocin complexes. Fig. SM3 in the supplementary material shows the analysis of the distances between the water molecules, residues and atoms of the ligands in the different systems. In order to keep the same format, the water molecules have been identified by the Amber residue number adopted after running the Leap software.

Binding free energy calculations were carried out in the complexed systems keeping the stable waters to account for the changes in the entropy. The binding free energy of the complex 1EI1–ADPNP was calculated including Na^+ 207 and without. Results are collected in Table 7 dissected into contributions from the different terms involved in the mathematical equations defining this energy. The predicted binding energies found by MM-PBSA and MM-GBSA methods are around 5 kcal/mol lower for the complexion of the protein with CBN compared to NOV, which is in agreement with the results from thermodynamic experiments previously reported for the binding of these coumarins to Gyr24 [17,18,47,48].

Experimental reports from Holdgate et al. [17] demonstrated that the binding of novobiocin to the mutant R136H gyrase ($\Delta H = -14.3$ kcal/mol) shows a more favorable enthalpy change of -2.1 kcal/mol than when it binds to the wild type protein ($\Delta H = -12.2$ kcal/mol). In our calculations a difference of about -3.4 kcal/mol is reproduced by the MM-PBSA approach when comparing the relative binding energies for these systems (Table 7). MM-GBSA estimates a difference of 0.6 kcal/mol.

Table 7
Binding free energy (kcal/mol) for the four complexed systems under study.

	1EI1–ADPNP (with Na^+)		1EI1–ADPNP (without Na^+)		1KZN–CBN		1AJ6–NOV		1AJ6(R136H)–NOV	
	Mean[a]	STD[b]	Mean[a]	STD[b]	Mean[a]	STD[b]	Mean[a]	STD[b]	Mean[a]	STD[b]
E_{elect}	−361.0	48.0	14.8	48.6	−40.4	9.9	−49.1	6.2	−56.6	5.9
E_{vdW}	−47.9	4.9	−56.2	4.3	−70.4	6.1	−64.0	3.4	−58.1	3.9
E_{gas}	−408.9	47.3	−41.4	48.0	−110.9	13.3	−113.1	6.8	−114.6	6.2
G_{nonP}	−6.2	0.1	−6.2	0.1	−8.5	0.7	−8.2	0.2	−8.3	0.2
G_{PB}	341.2	39.1	79.7	37.4	76.7	11.2	83.0	7.1	81.2	5.8
$G_{(PB)-subt}$	−73.8	11.3	32.1	15.7	−42.6	7.4	−38.3	4.9	−41.7	5.4
G_{GB}	320.2	38.7	13.0	39.0	60.9	9.3	66.0	5.0	68.2	4.4
$G_{(GB)-subt}$	−94.8	11.9	−34.6	12.0	−58.4	6.0	−55.3	3.4	−54.7	4.0
TSTRA	−13.3	0.0	−13.3	0.0	−13.6	0.0	−13.5	0.0	−13.5	0.0
TSROT	−11.1	0.0	−11.1	0.0	−11.9	0.1	−11.6	0.0	−11.6	0.0
TSVIB	−0.4	6.5	3.0	7.4	2.1	7.6	0.4	7.5	0.3	6.8
TSTOT	−24.9	6.5	−21.5	7.4	−23.4	7.7	−24.7	7.5	−24.8	6.8
$G_{(PB)-total}$	−49.0		53.6		−19.2		−13.6		−16.9	
$G_{(GB)-total}$	−69.9		−13.1		−35.0		−30.6		−29.9	

[a] Mean corresponds to $\Delta X = \Delta X_{comp} - \Delta X_{rec} - \Delta X_{lig}$ where X is every term displayed in the first column.
[b] STD is the standard deviation. All the terms have been explained above in equations (1)–(4). In the case of $G_{(PB/GB)-subt} = E_{gas} + G_{solv}$ the influence of the entropy has not been taken into account, therefore is considered a relative binding free energy. $G_{solv} = G_{nonP} + G_{PB/GB}$. TSTRA, TSROT, TSVIB and TSTOT are the parameters related to the entropy (TS = temperature multiplying the entropy) and stand for the translational, rotational, and vibrational entropy of the solute respectively. $G_{(PB)-total}$ and $G_{(GB)-total}$ are the final binding free energy values since the effect of the entropy is included. Details for complex and receptor can be consulted in the supplementary material Table SM2.

On the other hand, it has been described that, despite the fact that the complexation between the mutant protein and novobiocin shows a better binding enthalpy, there is an entropy penalty ($T\Delta S = -6.1$ kcal/mol versus $T\Delta S = -2.1$ kcal/mol) produced by an ordered water that is located in lieu of the side chain of the Arg136 of the wild type gyrase [17]. This results in a better binding free energy for the wild type GyrB with novobiocin complex ($\Delta G_{wild\ type} = -10.1$ kcal/mol and $\Delta G_{mutant} = -8.1$ kcal/mol). In Table 7 can be seen that all binding reactions give a decrease in entropy. However, the 1AJ6(R136H)–NOV does not show any significant difference with the 1AJ6–NOV complex. This can be explained by the absence of water 12 (PDB code) interacting with His136 and the novobiocin ligand, which could not be included in the energy calculations. Yu and Rick [15] predicted this entropic contribution correctly by a more accurate alchemical free energy calculation approach. In our simulation, we use a significant amount of approximations intrinsic in the two end-points MM/PB(GB)SA approach [49].

The relative binding free energy for the 1EI1–ADPNP system, where the ligand has a charge equal to −4 is highly influenced by Na^+ ion 207 interacting with residue Asn46, the charged oxygen atoms of the phosphate groups and the waters 221 and 3213. These favorable electrostatic interactions contribute to stabilize the binding reaction for about −367.46 kcal/mol in the gas phase which compensates the high values of the solvation free energy ($G_{PB(with\ Na)} - G_{PB(without\ Na)} = 261$ kcal/mol, $G_{GB(with\ Na)} - G_{GB(without\ Na)} = 307.22$ kcal/mol). On the other hand, it can be seen that when the Na^+ ion is included, the difference between the results derived from both MM-PBSA and MM-GBSA approaches is about 20.97 kcal/mol which is similar to the results yielded in the other systems (with clorobioicin and novobiocin ligands), whereas excluding this ion leads to a larger difference of 66.74 kcal/mol. In addition, there is a penalty in the entropy of about 3.37 kcal/mol due the ordered Na^+. Thus, it can be concluded that including the counter ion more reliable results can be achieved in these energy calculations.

3.4. Residue based description and comparison between the systems: energy decomposition and computational alanine scanning techniques

Recognizing the key residues in the binding site of the enzyme that are involved in the binding mode with the aforementioned inhibitors was carried out by means of two techniques: free energy decomposition and alanine scanning. Both techniques combine explicit molecular mechanical energies and continuum solvation models for calculating receptor–ligand interaction free energies. Free energy decomposition is a non-perturbing alternative approach which partitions solvation free energies and gas-phase energies between residues at an atomic level and provides a description of the contribution of the substructural elements side chain and backbone. On the other hand, alanine scanning relies on the assumption of minimal global conformational changes after the structural modification in the residues of interest which is considered the main limitation of the method [38]. Its major attraction consists in evaluating the binding free energies and predicts favorable modifications on residues of the active site to enhance non-covalent interactions in receptor–ligand systems [38]. To verify whether mutating this protein in the active site conforms to this assumption, a superposition of the average structures obtained after 15 ns of MD simulation of systems 1AJ6–NOV and 1AJ6(R136H)–NOV was done. As it has been described before, a "manual mutation" of residue His136 into Arg136 in 1AJ6 pdb file was carried out to obtain the wild type sequence in the protein. Fig. 6 shows the secondary structure of the amino acids in the

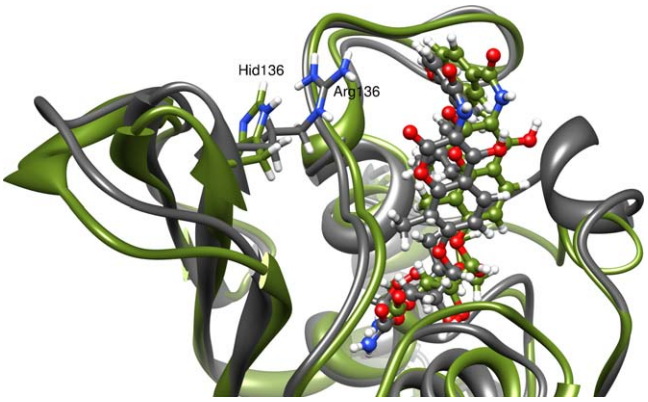


Fig. 6. Superposition of the average structures from the MD simulations of the 1AJ6–NOV (in gray) and 1AJ6(R136H)–NOV (in green) systems. NOV is depicted in stick representation as well as Arg136 and His136. This picture was generated by using the Chimera software [20,36].(For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

binding pocket and the novobiocin structure highlighting residues Arg136 and His136 involved in the mutation procedure. As can be seen, only the region comprising the flexible loops displays some difference. Thus, no significant conformational changes are caused after the mutation so we believe that it is possible to apply computational alanine scanning to study the binding energy changes caused by ala mutations in GyrB24.

The most significant results will be explained. The energy values related to the binding reaction which support the explanations have been stated in Tables SM4 and SM5 of the supplementary material for the energy decomposition technique and computational alanine scanning respectively because of the large volume of data generated by the calculations. The energy decomposition technique was not applied to the 1EI1–ADPNP system, since the Na^+ could not be included in the calculations.

3.4.1. Energy decomposition technique

In broad terms, the energy decomposition approach was able to recognize key residues involved in the stabilization of the non-covalent interactions in the GyrB24 complexes. Derived from the simulations and in accord with the experimental data, residues Ile78 and Pro79 are implicated in stabilizing the isopentenyl group of the coumarin inhibitors by hydrophobic interactions. Ile78 is involved in van der Waals interactions with the noviose sugar while Pro79 contacts with phenol moiety and the coumarin ring. In this hydrophobic interaction grid is also included residue Ile94 which interacts with the dimethyl and hydroxymethyl substituents in the noviose sugar framework, and Arg76 whose side chain stacks to the coumarinic ring of the ligands. The energy decomposition technique illustrates their favorable contributions to the relative binding free energy where the side chain showed the lowest values of energy. This effect is mainly supported by low values of the van der Waals energy.

Intriguingly, despite the well described role of residues Val43, Ala47, Val71, Val167 and Val120 encompassing a deep hydrophobic pocket where the pyrrol fragment of CBN is anchored, energy decomposition technique yielded a small but stabilizing contribution for most of these residues in any of the systems including the GyrB24–CBN complex. Only Ala47 displayed a modest contribution to the relative binding free energy of −1.48 kcal/mol mainly contributed by the van der Waals interactions energy of −1.59 kcal/mol (system 1KZN–CBN). An examination of the average structure of the GyrB24–CBN complex extracted from the MD simulation shows appropriate distances for residues Val43, Val71 and Val167 to the pyrrole of the CBN inhibitor to establish van der Waals interactions (see Fig. 7). Although all these residues have individually small contributions, the sum of them stabilizes the ligand in this pocket.

In the text that follows, the symbols and abbreviations used are the same as described in the footnote of Table 7. Additionally, (S), (B) and (T) stand for side chain, backbone and total residue. Asn46 is one of the important residues involved in the protein–ligand interactions. In the case of inhibitors like novobiocin and clorobiocin, a hydrogen bond is formed to the novobiose sugar. Residue Asn46 has one of the largest stabilizing contributions to the binding energy (see Table SM4). In the complexes 1KZN–CBN ($\Delta G_{(GB)-subt(T)} = -3.38$ kcal/mol), 1AJ6–NOV ($\Delta G_{(GB)-subt(T)} = -1.89$ kcal/mol) and 1AJ6(R136H)–NOV ($\Delta G_{(GB)-subt(T)} = -3.55$ kcal/mol), this contribution is attributed to its favorable van der Waals and electrostatic interaction with the ligands provided mainly by the backbone of the residue. Hydrogen bond occupancy of 99.30%, 98.68% and 5.13% were found for Asn46@O...H86-CBN@O49, Asn46@O...H53-NOV@O10 in 1AJ6–NOV and 1AJ6(R136H)–NOV respectively (Table SM3). Additionally this residue established a hydrogen bond with a significant occupancy of 92.89% in the complex involving the R136H protein with the nitrogen of the carbamate moiety attached to the noviose sugar (Asn46@O...H48-NOV@N03).

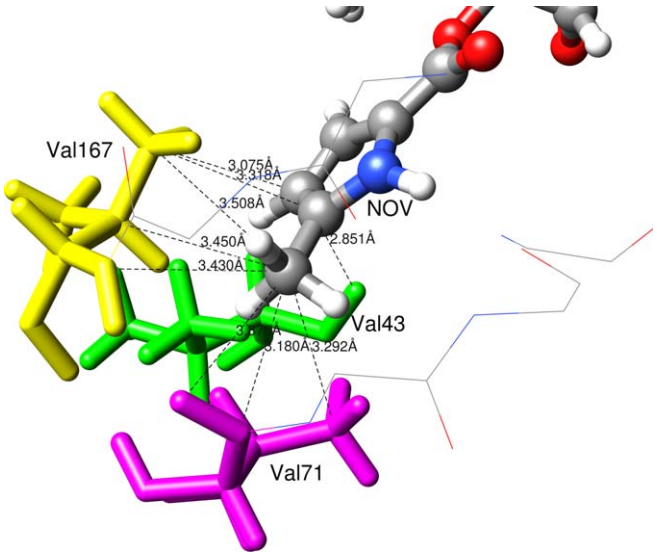


Fig. 7. Distance of residues Val43, Val71 and Val167 to the pyrrole of CBN inhibitor in the hydrophobic pocket of the average structure of GyrB24 derived from MD simulation. CBN is depicted in ball and stick representation while Val43, Val71 and Val167 are in stick. Picture was generated by using the Chimera software [20,36].

The energy decomposition technique predicts favorable van der Waals interactions between Phe104 and the coumarin ligands NOV and CBN. An analysis of the structures indicated that this interesting result follows from the van der Waals interaction that the Phe104 phenyl ring establishes with the dimethyl substituents in the sugar moiety and 3-methylbut-2-enyl attached to the benzamido part in the coumarin inhibitors (1KZN–$CBN_{(Phe104)}$: $\Delta G_{(GB)-subt(S)} = -1.15$; 1AJ6–$NOV_{(Phe104)}$: $\Delta G_{(GB)-subt(S)} = -2.33$; 1AJ6(R136H)–$NOV_{(Phe104)}$: $\Delta G_{(GB)-subt(S)} = -2.09$).

Some peculiarities were found when comparing the 1AJ6–NOV and 1AJ6(R136H)–NOV systems. The main difference between 1AJ6–NOV and 1AJ6(R136H)–NOV is the mutation of the Arg136 into His136. The analysis of the contribution to the total binding free energy for these residues revealed that they are not significant (see Table SM4 in the supplementary material). The hydrogen bond NOV@O26...HH12-Arg136@NH1 showed a 46.46% of occupancy during the MD simulation in the wild type GyrB complexed with novobiocin (Table SM3c in the supplementary material) along with a large favorable value of the electrostatic interaction energy ($\Delta E_{elect(T)} = -7.29$ kcal/mol). However, this contribution is offset by the high desolvation value for this residue (7.48 kcal/mol) leading to a not significant total value of relative free binding energy (−0.51 kcal/mol). On the other hand, His136 does not interact with the ligand, since the side chain points to the solvent and therefore the contribution of every term to the relative binding free energy is not significant.

The stability of the residues contained in the loops was extracted from the results of the RMSD analysis of the active site and the two loops in the representative structures of the clusters derived from the simulation (Stated in Table SM1f and SM1h from the supplementary material). However, a few residues of the loops showed different tendencies in every system and their contribution to the binding free energy of the complex is discussed in following paragraph.

Contributions of loop 1 residues (83–87) to the binding energy are very small, although not negligible. An example of this is residue Pro84 showing lower values of the relative free binding energy in 1AJ6(R136H)–NOV: $\Delta G_{(GB)-subt(T)(1AJ6-NOV)} = -0.6$ kcal/mol and $\Delta G_{(GB)-subt(T)(1AJ6(R136H)-NOV)} = -1.3$ kcal/mol. Shorter

distances between NOV and the side chain of this residue in the R136H system, compared to the wild type GyrB24 complex with NOV, improve hydrophobic contacts ($\Delta E_{vdW-Pro84(T)(1AJ6-NOV)} = -0.49$ kcal/mol and $\Delta E_{vdW-Pro84(T)(1AJ6(R136H)-NOV)} = -1.08$ kcal/mol).

Likewise, residue Lys110 from the second loop (101–118) displays different possibilities of interaction with the ligand in both systems, which are reproduced by the techniques used for the calculation of the energies. In accordance, this amino acid shows a better energy contribution in the mutated protein bound to novobiocin ($\Delta G_{(GB)-subt(T)(1AJ6-NOV)} = 0.27$ kcal/mol and $\Delta G_{(GB)-subt(T)(1AJ6(R136H)-NOV)} = -1.25$ kcal/mol), which may be explained by the favorable van der Waals and electrostatic interactions found for the side chain of Lys110 in the complex. For example, only in the 1AJ6(R136H)–NOV system a hydrogen bond was ascertained between the novobiocin and Lys110 (NOV@O12...HZ3-Lys110@NZ) with 38% occupancy (See Table SM3d and SM4 in the supplementary material for details).

3.4.2. Computational alanine scanning technique

Experimental extensive site-directed mutagenesis studies have been previously performed [50–53] to identify crucial residues implied in ATP binding and hydrolysis. A notably decline of the ATPase activity of the *E. coli* DNA gyrase was found after mutating residues His38, Glu42, Asn46, Glu50, Asp73, Arg76, Gly77, Ile78 and Thr165 to alanine. In addition, when residues Pro79 and Lys103 were substituted by alanine, about half of the ATPase activity was conserved, nevertheless displayed no DNA supercoiling activity, hinting that both residues might be essential for linking of the ATP hydrolysis to the DNA supercoiling activity. Herein we state the most significant results derived from applying the computational alanine scanning technique (see Table SM5 in supplementary material section for details).

Qualitatively, in the coumarin simulations, energies obtained by alanine scanning for residues Asn46, Ile78, Ile94, Val120, Val43, Val71, Arg76, Arg136, Phe104, Lys110 and Lys103 are in agreement with the results found with the energy decomposition technique. However, only computational alanine scanning indicates the stabilizing role that the side chain of Glu50 has over the coumarin inhibitors. In the case of the mutant 1AJ6(R136H) we see that the inhibitor shifts in the pocket resulting in different distances. For example, the distance between the carboxylate of the Glu50 and the carbonyl in the carbamate of novobiocin is shorter in the case of the mutant R136H leading to unfavorable electrostatic interactions.

The analysis of the complex 1EI1–ADPNP will be explained apart due to the particularities of: the −4 charged ligand, the closed conformation of loop 2 in the protein and a Na^+ ion included in the calculations. Alanine scanning data suggested residues Asn46, Arg76, Ile78 and Ile94 to be important for the binding affinity in a similar way to the GyrB–coumarin complexes explained above. Mutating Asn46 to alanine results in a destabilization of the binding energy for ADPNP from −73.83 to −59.13 kcal/mol and from −94.80 to −81.82 kcal/mol for the PB and GB model respectively. One of the contributions may be the hydrogen bond of 99.47% of occupancy between the side chain of the Asn46 and the ligand (ADPNP@O11...HD21-Asn46@ND2, see Table SM3b).

Binding sites for coumarins NOV and CBN show a certain degree of overlap with the one for ADPNP involving the novobiose sugar of the coumarins and the adenine ring of ADPNP. Nevertheless, there are regions of the binding sites for these compounds that are unique. Flexible loop 2 (residues 101–118) undergoes a considerable conformational change between the ADPNP bound and the coumarin bound structures. Residues that are part of this loop such as Gly102, Lys103, Leu115, HIP116, Gly117, Val118, Gly119 and Val120 have been described to be essential in the interactions with ADPNP [8].

Mutating Val120 and Val118 into alanine destabilizes the binding reaction while not significant changes were found when mutating Leu115 (Table SM5). Although these residues established very stable hydrogen bonds between the N of the backbone and the phosphate groups of ADPNP along the simulation (see Table SM3b of the supplementary material for values), these H-bonds are also kept in the alanine residue. However, the side chain of the analyzed valine residues can interact with the phosphate groups whereas in Leu115, the side chain points out of the active site.

The favorable influence of residues with a positively charged side chain such as Hip116 and Lys103 was also verified by using the computational alanine scanning technique. These residues are involved in a salt bridge between the charged group of the side chain and the phosphate groups (38.50% and 44.7% of occupancy during the simulation for ADPNP@O03...Hip116@ND1 and ADPNP@O06...Lys103@NZ respectively), which are not possible with alanine.

Mutating the negatively charged residue Glu42 by an alanine, results in an energetic stabilization (electrostatic contribution to the binding energy that is not compensated by the unfavorable solvation contribution to the free energy). Nevertheless, Glu42 has been proposed to be a catalytic base that is stabilized by His38 and accepts a proton from a nearby water molecule. The resulting OH^- interacts with the γ-phosphate of the ATP producing ADP and inorganic phosphate [50,54]. In our study, the interaction between Glu42 and the double protonated Hip38 was reproduced along the simulation.

It was not possible to study quantitatively the effects of glycine residues on the energy calculations neither by using energy decomposition technique because of the reasons abovementioned, neither by alanine scanning since this is a smaller residue. Nevertheless, the analysis of the hydrogen bond occupancy for these residues suggests that they contribute to the phosphate moiety stabilization (see Table 8).

Asp73 has been considered in many of the computational studies about DNA gyrase inhibitors as a cornerstone in the interactions to take into account when designing new drugs against this target [8–10]. For this reason, we took a closer look to the energy contributions. Let us first consider the interaction of residue 73 with the coumarin inhibitors (1KZN and 1AJ6). The positive free energy of solvation in the Asp73 and Ala73 systems (Table 9) must be explained by mainly desolvation penalty of these residues. The final binding reaction of the coumarins is stabilized more by the Asp than the Ala residue, mainly by the electrostatic component. The side chain of Asp73 is part of a very favorable H-bond network formed by the coumarin ligand and a water molecule [8,55,56]. During the simulation these interactions were reproduced and translated into better values of electrostatic interactions in comparison with the mutant proteins with alanine substituting this important residue. There is a direct H-bond between Asp73 and the pyrrole with a 42% of occupancy while another H-bond guides the binding between the carboxylate linked to the methylpyrrole of clorobiocin and Asp73 through Wat213. In the protein with R136H mutation as well as in the wild type one, it is revealed that Asp73 establishes one hydrogen bond with the N of the carbamate in the novobiocin ligand (99.91% of occupancy in 1AJ6–NOV and 53.15% in 1AJ6(R136H)), and another one via an intermediate water molecule with the carboxylate from the same moiety. The water molecules are Wat218 and Wat219 for 1AJ6–NOV and 1AJ6(R136H)–NOV respectively. Additionally, two more water molecules interact with the Asp73 side chain as part of the favorable net of H-bonds with the ligand and Val43 as explained above in the binding free energy analysis section (see Fig. SM3 for the analysis of the role of the waters in the net and Table SM3 of supplementary material for the H-bond occupancy).

Table 8
Hydrogen bond occupancy for the glycine residues around the phosphate moiety of ADPNP throughout the simulation. For every system the cut-off used for H-bond distance and angle was 3.5 Å and 120° respectively.

HB acceptor	HB donor	Occupancy (%)	Distance (Å)	Angle (°)
:ADPNP@O04	Gly119@N...H	85.46	3.131 (0.17)	27.80 (12.11)
:ADPNP@O04	Gly117@N...H	58.54	2.930 (0.14)	46.17 (10.24)
:ADPNP@O18	Gly102@N...H	49.05	3.071 (0.17)	25.60 (13.89)
:ADPNP@O12	Gly119@N...H	10.86	3.421 (0.07)	41.42 (11.88)
:ADPNP@O10	Gly119@N...H	8.25	2.951 (0.16)	38.04 (9.71)
:ADPNP@O12	Gly117@N...H	3.25	3.373 (0.09)	57.33 (2.39)
:ADPNP@O07	Gly117@N...H	2.35	3.378 (0.10)	53.15 (5.44)
:ADPNP@O03	Gly117@N...H	2.18	3.398 (0.10)	26.94 (9.87)
:ADPNP@O02	Gly119@N...H	2.14	3.379 (0.11)	44.57 (7.19)
:ADPNP@O11	Gly119@N...H	1.98	3.228 (0.15)	56.63 (3.00)
:ADPNP@O07	Gly102@N...H	1.72	3.316 (0.15)	44.24 (11.87)
:ADPNP@O20	Gly102@N...H	0.75	3.380 (0.10)	46.80 (8.83)

Table 9
Contribution of residue 73 to the relative binding free energies for the basic systems and for the respective alanine mutant structures. The energies are in kcal/mol. Energy values in this table are already based on differences. The binding energies are calculated as $\Delta G_{PB/GB} = G_{PB/GB-complex} - G_{PB/GB-receptor} - G_{PB/GB-ligand}$.

	Asp73	ΔE_{elect}	ΔE_{vdW}	ΔE_{gas}[a]	$\Delta G_{nonP-PB}$[b]	ΔG_{PB}[b]	$\Delta G_{(PB)-subt}$[c]	$\Delta G_{nonP-GB}$[b]	ΔG_{GB}[b]	$\Delta G_{(GB)-subt}$[c]
1KZN-CBN	Asp	−40.4	−70.4	−110.9	−8.5	76.7	−42.6	−8.5	60.9	−58.4
	Ala	−29.1	−71.0	−100.0	−8.5	77.2	−31.4	−8.5	56.6	−52.0
	ΔΔ[d]	**−11.4**	**0.6**	**−10.8**	**0.1**	**−0.5**	**−11.3**	**0.1**	**4.4**	**−6.4**
1AJ6-NOV	Asp	−49.1	−64.0	−113.1	−8.2	83.0	−38.3	−8.2	66.0	−55.3
	Ala	−35.4	−64.9	−100.3	−8.2	87.4	−21.1	−8.2	60.5	−48.0
	ΔΔ[d]	**−13.7**	**0.9**	**−12.8**	**0.0**	**−4.4**	**−17.2**	**0.0**	**5.5**	**−7.3**
1AJ6(R136H)–NOV	Asp	−56.6	−58.1	−114.6	−8.3	81.2	−41.7	−8.3	68.2	−54.7
	Ala	−43.7	−59.1	−102.8	−8.3	83.2	−27.9	−8.3	63.7	−47.4
	ΔΔ[d]	**−12.9**	**1.0**	**−11.9**	0.0	**−2.0**	**−13.8**	**0.0**	**4.5**	**−7.3**
1EI1–ADPNP	Asp	−361.0	−47.9	−408.9	−6.2	314.2	−73.8	−6.2	320.2	−94.8
	Ala	−469.2	−47.5	−516.7	−6.2	442.5	−80.4	−6.2	427.6	−95.3
	ΔΔ[d]	**108.3**	**−0.4**	**107.9**	**0.1**	**−101.3**	6.6	**0.1**	**−107.4**	0.5

[a] $\Delta E_{gas} = \Delta E_{elect} + \Delta E_{vdW}$, the sum of the electrostatic and van der Waals contribution of residue 73 to the binding energy of the respective inhibitors to the enzyme.
[b] ΔG_{PB}, ΔG_{GB}, $\Delta G_{nonP-PB}$, $\Delta G_{nonP-GB}$ = electrostatic polar and nonpolar solvation contribution to the binding energy of the ligand calculated by using PB or GB for every complex respectively.
[c] $\Delta G_{(PB)-subt} = \Delta G_{PB} + \Delta G_{nonP-PB} + \Delta E_{gas}$, $\Delta G_{(GB)-subt} = \Delta G_{GB} + \Delta G_{nonP-GB} + \Delta E_{gas}$.
[d] ΔΔ The difference between the energy values of wild type and (Asp73–Ala73) mutated system.

In the 1EI1–ADPNP complex, the Ala73 electrostatic component is more stabilizing for the binding of ADPNP than the one estimated for Asp73. This can be explained by the repulsion between Asp and ADPNP. The solvation contribution to the binding of ADPNP, is more favorable for the Asp residue than for the Ala. A direct H-bond was found along the simulation between Asp73 side chain and the amine group linked to the ring at position 6 (52.32% of occupancy). An extra H-bond was ascertained via the water molecule Wat223 between the Asp73 and the nitrogen atom of the adenine ring at position 1 of the ADPNP. However, the better solvation of Asp73 does not compensate the more favorable electrostatic effect of Ala73, making Ala73 the most stabilizing residue for the binding free energy in the 1EI1–ADPNP complex when using the MM-PBSA approach, while for MM-GBSA the difference between Asp73 and Ala73 is not significant (Table 9).

Supporting also these explanations, the energy decomposition technique shows the stabilizing effect that the bridging water molecules Wat213, Wat218 and Wat219 have in the interaction with the ligand in the systems 1KZN–CBN, 1AJ6–NOV and 1AJ6(R136H)–NOV respectively. The values are shown in Table 10.

Table 10
Energy decomposition results for the bridging water molecules interacting with the ligand.

System	1KZN–CBN	1AJ6–NOV	1AJ6(R136H)–NOV
	Wat213	Wat218	Wat219
$\Delta E_{vdW(T)}$	−0.08(0.37)	0.46(0.54)	0.11(0.46)
$\Delta E_{elect(T)}$	−1.54(0.79)	−2.64(0.68)	−1.45(1.16)
$\Delta G_{GB(T)}$	0.97(0.57)	0.97(0.24)	0.85(0.56)
ΔG_{nonP}	−0.07(0.02)	−0.06(0.02)	−0.04(0.02)
$\Delta G_{(GB)-subt(T)}$	−0.72(0.54)	−1.27(0.34)	−0.52(0.57)

4. Conclusions

In summary, the GyrB24 subunit from *E. coli* has been investigated, namely the ATP binding pocket both in free ligand state and bound with the well-established inhibitors clorobiocin, novobiocin and 5′-adenylyl-β-γ-imidodiphosphate. Molecular dynamics simulations were run in the different possible systems for energetically accessible conformational searching. Then a clustering data mining technique was used to group conformations to explore the possible changes in the conformational space using the information from the trajectories. Analysis of the time dependent root-mean-squared coordinate deviation (RMSD) to the initial structure, temperature factors and obtained clusters led to the same conclusion: (1) the loops are the most flexible fragments in every system, especially in the unbound enzymes. (2) The higher stability is for the complexed system, which may be justified by the influence of the possible interactions of the inhibitors with the residues in the active site. (3) Although a small deviation from the starting conformation of the protein is observed when studying the representative structures of the obtained clusters, a similar shift is detected in the ligands, so that important interactions with the residues in the binding pocket are conserved. This fact shows that for further modeling studies using the GyraseB24 ATP binding site to design new inhibitors by docking or virtual screening, the 1KZN structure where loop 1 and 2 have been added (loop 2 in the open conformation) can be used. Also the 1EI1 structure can be used to account for the closed conformation (of loop 2) structure. Recognizing the key residues in

the binding site of the enzyme that are involved in the binding mode with the aforementioned inhibitors was carried out by means of two techniques: free energy decomposition and computational alanine scanning.

The assumption of minimal global conformational changes after the structural modification in the residues of interest to use the computational alanine scanning was verified by superposing the average structures obtained after 15 ns of MD simulation of systems 1AJ6–NOV and 1AJ6(R136H)–NOV.

Qualitatively, energies obtained by alanine scanning for residues Asn46, Ile78, Ile94, Val120, Val43, Val71, Arg76, Arg136, Phe104, Lys110 and Lys103 are in agreement with the results found with the energy decomposition technique. When comparing results obtained by these techniques for 1KZN–CBN system with 1AJ6–NOV and 1AJ6(R136H)–NOV, it can be seen how residues like Asn46, Asp73, Arg76, Ile78, Pro79, Ile94 and Phe104 have been predicted with similar contributions to the binding free energy and role in the non-covalent interactions with these two coumarins.

Binding sites for coumarins NOV and CBN show a certain degree of overlap with the one for ADPNP involving novobiose sugar of the coumarins and the adenine ring of ADPNP. However, residues from loop 2 such as Gly101, Gly102, Lys103, Gly114, Leu115, His116, Gly117, Val118 and Gly119 have been pointed out by energy decomposition technique as significant only in interaction with ADPNP which is expected from the different conformation of loop 2. Asp73 was predicted as one of the most important residues in the binding mode with conserved hydrogen bond interactions in all the inhibitors under study.

Two different solvation models were used to determine the polar component of the solvation free energy, Poisson–Boltzmann (PB) and Generalized Born (GB). Despite a similar tendency in the results can be appraised, there is still, some inconsistency when comparing both methods. Despite these methods have been used widely and resulted useful in this kind of studies, it is worthy to remind there are many approximations included that can affect the accuracy of the results.

Due to the different philosophy behind the techniques alanine scanning and energy decomposition, we considered important to combine them to obtain more information about the role of the residues interacting with the inhibitors. This has proven to be a useful and powerful tool to predict the possible binding mode of potential new inhibitors with an adequate accuracy.

Acknowledgments

We are grateful to the Flemish Interuniversity Council for financial support to the project “Strengthening postgraduate education and research in Pharmaceutical Sciences” where this research is included. We are also grateful to the Vlaams Supercomputer Centrum (VSC) for the use of the linux cluster.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2010.12.005.

References

[1] S.B. Levy, B. Marshall, Antibacterial resistance worldwide: causes, challenges and responses, Nat. Med. 10 (2004) S122–S129.
[2] R.L. Monaghan, J.F. Barrett, Antibacterial drug discovery – then, now and the genomics future, Biochem. Pharmacol. 71 (2006) 901–909.
[3] D.J. Payne, M.N. Gwynn, D.J. Holmes, D.L. Pompliano, Drugs for bad bugs: confronting the challenges of antibacterial discovery, Nat. Rev. Drug Discov. 6 (2007) 29–40.
[4] A. Maxwell, DNA gyrase as a drug target, Trends Microbiol. 5 (1997) 102–109.
[5] J.S. Wolfson, D.C. Hooper, The fluoroquinolones: structures, mechanisms of action and resistance, and spectra of activity in vitro, Antimicrob. Agents Chemother. 28 (1985) 581–586.
[6] R. Stahlmann, Clinical toxicological aspects of fluoroquinolones, Toxicol. Lett. 127 (2002) 269–277.
[7] J.J. Barker, Antibacterial drug discovery and structure-based design, Drug Discov. Today 11 (2006) 391–404.
[8] M. Oblak, M. Kotnik, T. Solmajer, Discovery and development of ATPase inhibitors of DNA gyrase as antibacterial agents, Curr. Med. Chem. 14 (2007) 2033–2047.
[9] H.J. Boehm, M. Boehringer, D. Bur, H. Gmuender, W. Huber, W. Klaus, et al., Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening, J. Med. Chem. 43 (2000) 2664–2674.
[10] M. Schechner, F. Sirockin, R.H. Stote, A.P. Dejaegere, Functionality maps of the ATP binding site of DNA gyrase B: generation of a consensus model of ligand binding, J. Med. Chem. 47 (2004) 4373–4390.
[11] S. Firth-Clark, N.P. Todorov, I.L. Alberts, A. Williams, T. James, P.M. Dean, Exhaustive de novo design of low-molecular-weight fragments against the ATP-binding site of DNA-gyrase, J. Chem. Inf. Model. 46 (2006) 1168–1173.
[12] M. Stahl, M. Rarey, Detailed analysis of scoring functions for virtual screening, J. Med. Chem. 44 (2001) 1035–1042.
[13] G.M. Verkhivker, D. Bouzida, D.K. Gehlhaar, P.A. Rejto, S. Arthurs, A.B. Colson, et al., Deciphering common failures in molecular docking of ligand-protein complexes, J. Comput. Aided Mol. Des. 14 (2000) 731–751.
[14] D.A. Ostrov, J.A. Hernandez Prada, P.E. Corsino, K.A. Finton, N. Le, T.C. Rowe, Discovery of novel DNA gyrase inhibitors by high-throughput virtual screening, Antimicrob. Agents Chemother. 51 (2007) 3688–3698.
[15] H. Yu, S.W. Rick, Free energies and entropies of water molecules at the inhibitor-protein interface of DNA gyrase, J. Am. Chem. Soc. 131 (2009) 6608–6613.
[16] L. Brino, A. Urzhumtsev, M. Mousli, C. Bronner, A. Mitschler, P. Oudet, et al., Dimerization of *Escherichia coli* DNA-gyrase B provides a structural mechanism for activating the ATPase catalytic center, J. Biol. Chem. 275 (2000) 9468–9475.
[17] G.A. Holdgate, A. Tunnicliffe, W.H. Ward, S.A. Weston, G. Rosenbrock, P.T. Barth, et al., The entropic penalty of ordered water accounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase: a thermodynamic and crystallographic study, Biochemistry 36 (1997) 9663–9673.
[18] D. Lafitte, V. Lamour, P.O. Tsvetkov, A.A. Makarov, M. Klich, P. Deprez, et al., DNA gyrase interaction with coumarin-based inhibitors: the role of the hydroxybenzoate isopentenyl moiety and the 5′-methyl group of the noviose, Biochemistry 41 (2002) 7217–7223.
[19] A.J. Schoeffler, A.P. May, J.M. Berger, A domain insertion in *Escherichia coli* GyrB adopts a novel fold that plays a critical role in gyrase function, Nucleic Acids Res. 38 (2010) 7830–7844.
[20] L. Holm, J. Park, DaliLite workbench for protein structure comparison, Bioinformatics 16 (2000) 566–567.
[21] V. Lamour, L. Hoermann, J.M. Jeltsch, P. Oudet, D. Moras, An open conformation of the *Thermus thermophilus* gyrase B ATP-binding domain, J. Biol. Chem. 277 (2002) 18947–18953.
[22] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C. Walker, W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, P.A. Kollman, AMBER 10, University of California, San Francisco, 2008.
[23] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, et al., General atomic and molecular electronic structure system, J. Comput. Chem. 14 (1993) 1347–1363.
[24] Ch.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, J. Phys. Chem. 97 (1993) 10269–10280.
[25] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174.
[26] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, et al., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, J. Comput. Chem. 24 (2003) 1999–2012.
[27] M.C. Lee, Y. Duan, Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model, Proteins 55 (2004) 620–634.
[28] W.L. Jorgensen, J. Chandrasekhar, J. Madura, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[29] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (1977) 327–341.
[30] T. Darden, D. York, L. Pedersen, Particle mesh Ewald-an Nlog(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.
[31] D.A. Case, T.E. Cheatham 3rd, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr., et al., The Amber biomolecular simulation programs, J. Comput. Chem. 26 (2005) 1668–1688.
[32] L.A. Kelley, S.P. Gardner, M.J. Sutcliffe, An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies, Protein Eng. 9 (1996) 1063–1065.

[33] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM Comp. SurV. 31 (1999) 264–323.
[34] J. Shao, S.W. Tanner, N. Thompson, T.E. Cheatham 3rd., Clustering molecular dynamics trajectories: characterizing the performance of different clustering algorithms, J. Chem. Theory Comput. 3 (2007) 2312–2334.
[35] http://www.ebi.ac.uk.
[36] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, et al., UCSF chimera – a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[37] T. Clackson, M.H. Ultsch, J.A. Wells, A.M. de Vos, Structural and functional analysis of the 1:1 growth hormone: receptor complex reveals the molecular basis for receptor affinity, J. Mol. Biol. 277 (1998) 1111–1128.
[38] I. Massova, P.A. Kollman, Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies, J. Am. Chem. Soc. 121 (1999) 8133–8143.
[39] R. Luo, L. David, M.K. Gilson, Accelerated Poisson-Boltzmann calculations for static and dynamic systems, J. Comput. Chem. 23 (2002) 1244–1253.
[40] D. Bashford, D.A. Case, Generalized born models of macromolecular solvation effects, Annu. Rev. Phys. Chem. 51 (2000) 129–152.
[41] D. Sitkoff, K.A. Sharp, B. Honig, Accurate calculation of hydration free energies using macroscopic solvent models, J. Chem. Phys. 98 (1994) 1978–1988.
[42] J. Weiser, P.S. Shenkin, W.C. Still, Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO), J. Comput. Chem. 20 (1999) 217–230.
[43] H. Gohlke, C. Kiel, D.A. Case, Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes, J. Mol. Biol. 330 (2003) 891–913.
[44] J. Lee, J.S. Kim, C. Seok, Cooperativity and specificity of Cys2His2 zinc finger protein-DNA interactions: a molecular dynamics simulation study, J. Phys. Chem. B. 114 (2010) 7662–7671.
[45] http://www.gnuplot.info.
[46] J.E. Ladbury, Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design, Chem. Biol. 3 (1996) 973–980.
[47] F.T. Tsai, O.M. Singh, T. Skarzynski, A.J. Wonacott, S. Weston, A. Tucker, et al., The high-resolution crystal structure of a 24-kDa gyrase B fragment from *E. coli* complexed with one of the most potent coumarin inhibitors, clorobiocin, Proteins 28 (1997) 41–52.
[48] N.A. Gormley, G. Orphanides, A. Meyer, P.M. Cullis, A. Maxwell, The interaction of coumarin antibiotics with fragments of DNA gyrase B protein, Biochemistry 35 (1996) 5083–5092.
[49] D.L. Mobley, A.P. Graves, J.D. Chodera, A.C. McReynolds, B.K. Shoichet, K.A. Dill, Predicting absolute ligand binding free energies to a simple model site, J. Mol. Biol. 371 (2007) 1118–1134.
[50] A.P. Jackson, A. Maxwell, Identifying the catalytic residue of the ATPase reaction of DNA gyrase, Proc. Natl. Acad. Sci. USA 90 (1993) 11232–11236.
[51] M.H. O'Dea, J.K. Tamura, M. Gellert, Mutations in the B subunit of *Escherichia coli* DNA gyrase that affect ATP-dependent reactions, J. Biol. Chem. 271 (1996) 9723–9729.
[52] C.V. Smith, A. Maxwell, Identification of a residue involved in transition-state stabilization in the ATPase reaction of DNA gyrase, Biochemistry 37 (1998) 9658–9667.
[53] C.H. Gross, J.D. Parsons, T.H. Grossman, P.S. Charifson, S. Bellon, J. Jernee, et al., Active-site residues of *Escherichia coli* DNA gyrase required in coupling ATP hydrolysis to DNA supercoiling and amino acid substitutions leading to novobiocin resistance, Antimicrob. Agents Chemother. 47 (2003) 1037–1046.
[54] A. Maxwell, D.M. Lawson, The ATP-binding site of type II topoisomerases as a target for antibacterial drugs, Curr. Top. Med. Chem. 3 (2003) 283–303.
[55] E.J. Gilbert, A. Maxwell, The 24 kDa N-terminal sub-domain of the DNA gyrase B protein binds coumarin drugs, Mol. Microbiol. 12 (1994) 365–373.
[56] R.J. Lewis, O.M. Singh, C.V. Smith, T. Skarzynski, A. Maxwell, A.J. Wonacott, et al., The nature of inhibition of DNA gyrase by the coumarins and the cyclothialidines revealed by X-ray crystallography, EMBO J. 15 (1996) 1412–1420.