International Journal of Multidisciplinary Research and Growth Evaluation
ISSN: 2582-7138
Received: 16-07-2021; Accepted: 02-08-2021
www.allmultidisciplinaryjournal.com
Volume 2; Issue 4; July-August 2021; Page No. 709-714

Insilico prediction of gyr A and gyr B in *Escherichia coli* insights the DNA-Protein interaction in prokaryotes

Sudheer Sivansankaran Menon [1], Shanmughavel Piramanayagam [2]
[1, 2] Department of Bioinformatics, Bharathiar University, Coimbatore, Tamil Nadu, India
Corresponding Author: **Sudheer Sivansankaran Menon**

Abstract
The high accessibility of genomic data and expression data has opened the possibilities for the target protein analysis. The full information about the protein structure affords well-founded hypothesis of the function of protein. Once the structure of the protein binding site is known, the applications of the drug discovery methods are done. The steric and complementarity of binding site of protein and drug molecule provides the complete information for structural design. We have employed protein threading and homology based prediction for the structure of DNA topoisomerase II from *Escherichia coli*. DNA topoisomerases catalyze topological interconversions: supercoiling relaxation, catenation decatenation and knotting unknotting of DNA, consist of GyrA and GryB subunit. In this paper we have described the docking for structure based drug design. Protein function can be predicted in number of ways from the sequence. Since the function of protein is based on protein domains, a number of databases like Pfam, PROSITE, PRINTS, ProDom, SwissPROT + TREMBL are used. Protein folds are evaluated by SCOP and CATH. The structural accepts predicted by the computational methods are expected to invade the search of target proteins and the development of new drugs.

Keywords: DNA Gyrase, GyrA, GyrB, *Escherichia Coli*

Introduction
DNA topoisomerases catalyze topological interconversions: supercoiling-relaxation, catenation-decatenation and knotting-unknotting of DNA. These results in the topological events occur during important cellular processes such as replication, transcription, recombination and chromosome segregation. Thus, the enzymes are essential for the cell survival, and hence are ubiquitous. The topoisomerases are classified into two distinct subclasses based on the mechanistics of the reaction.

Results
We have hypothesized the structure prediction of the subunit B of DNA gyrase. Functions of proteins are evaluated by Pfam, PROSITE, PRINTS, ProDom, SwissPROT + TREMBL. Protein folds are evaluated by SCOP and CATH. Base don the predicted structure of protein, drug design are done.

Conclusion
The present study highlights the complete evaluation of protein analysis. The structural and functional studies of this particular protein results in the development of new drugs.

Background
Protein-DNA interactions play an essential role in the genetic activities of life. Many structures of protein-DNA complexes are already known, but the common rules on how and where proteins bind to DNA have not emerged. Many attempts have been made to predict protein-DNA interactions using structural information, but the success rate is still about 80%. The studies on the shape of the molecular surface of the protein and DNA, along with the electrostatic potential on the surface and constructed a new statistical evaluation function to make predictions of DNA interaction sites on protein molecular surfaces. The shape of the molecular surface can be described by a combination of local and global average curvature, which has similarity to the small convex and concave and the large-scale concave curvatures of the protein surface, appearing at DNA-binding sites [1].
DNA topoisomerases catalyze topological interconversions mainly, supercoiling-relaxation, catenation-decatenation and knotting-unknotting of DNA. These results in the topological events occur during important cellular processes such as replication, transcription, recombination and chromosome segregation.

Thus, these enzymes are essential for the cell survival, and hence are ubiquitous. The topoisomerases are classified into two distinct subclasses based on the mechanistics of the reaction [12].

The type I topoisomerases break one strand of DNA and pass the other stand through the nick created and change the linking number in steps of one. On the other hand, type II enzymes cleave both stands of DNA and pass the duplex through the 'DNA gate' resulting in the change of linking number in steps of two [12-13]. All known topoisomerases form a transient covalent intermediate with DNA through a phosphotyrosine linkage and reseal after strand passage. The bacterium, *Escherichia coli* contains two type II topoisomerases besides two type I enzymes. Amongst all type II topoisomerases, only DNA gyrase has the ability to introduce negative supercoils into DNA in an ATP driven reaction. The heteromeric enzyme has been the subject of extensive study. The second bacterial type II enzyme, topoisomerase IV, has strong decatenation and weak relaxation activities. Both the bacterial type II topoisomerases have a similar architecture and also, share considerable sequence similarity [22].

Results and Discussion

A 3D model of Gyrase protein has been built using the 3D structure 1EI1 chain 'A' as template.This template shares 33.6% identities with your query sequence (using the ALIGN program).

Sequence was analyzed using Blast and Psi-Blast [Figure 4]. Since the function of protein is based on protein domains, a number of databases like Pfam, PROSITE, PRINTS, ProDom are used [Figure 1,2,3 and 6]. Protein folds are evaluated by SCOP and CATH [Table 1,2 and Figure 5]. Docking was performed using Hex.4.2. Crystal Structure of the Escherichia coli SbmC protein (AKA Gyrase Inhibitory Protein GyrI, AKA YeeB) (1KZN) and Crystal Structure of E. coli 24kDa Domain in Complex with Clorobiocin (1JYH) are used as receptor and ligand [Figure 7].The structural accepts predicted by the computational methods are expected to invade the search of target proteins and the development of new drugs.

The four DNA topoisomerases found in *E. coli* consist of the two type IA enzymes, DNA topoisomerases I and III, and the two type IIA enzymes, DNA gyrase and DNA topoisomerase IV [22]. Although some overlap of function has been shown genetically, each of the DNA topoisomerases appears optimized to carry out its own particular set of topological manipulations. DNA gyrase is the only known topoisomerase able to generate negative supercoiling at the expense of ATP hydrolysis and is responsible for global generation of negative supercoils in the bacterial chromosome [23]. Such global supercoiling in combination with the activity of the *E. coli* Muk proteins is essential for chromosome condensation leading to proper chromosome partitioning at cell division. Together, topoisomerases I and IV along with DNA gyrase set the steady-state level of negative supercoiling that is required for the initiation of replication and for transcription from at least some promoters [21]. Transcription itself generates positive supercoils ahead of and negative supercoils behind the translocating RNA polymerase that are rapidly resolved by DNA gyrase and DNA topoisomerase I, respectively. Fork movement during replication of a circular DNA can generate topological changes in both the unreplicated region ahead of the fork and in the already replicated region behind the fork [27]. Early after initiation, movement of a replication fork causes overwinding of the DNA in the unreplicated region of the theta intermediate, and the resulting positive supercoils are rapidly removed by DNA gyrase. Recent evidence confirms that excess helical windings generated by replication can be distributed both in front of and behind the replication fork. Therefore, fork movement during replication can be maintained not only by DNA gyrase relaxing positive supercoils in front of the fork but also through the "unwinding" of the daughter duplexes behind the replication fork. Of the two type II enzymes in *E. coli*, topoisomerase IV is much more effective at decatenating DNA than at relaxing positive supercoils whereas the converse is true for DNA gyrase.

The dimeric enzymes bind duplex DNA and cleave the opposing strands with a four base stagger. The Cleavage involves covalent attachment of each subunit of the dimer to the 50 end of the DNA through a phosphor tyrosine bond. A conformational change pulls the two ends of the cleaved duplex DNA apart to create an opening in what is referred to as the gated or G-segment DNA. A second region of duplex DNA from either the same molecule (relaxation, knotting or unknotting) or a different molecule (catenation or decatenation), referred to as the transported or T-segment, is passed through the open DNA gate. This feature of the reaction explains why the linking number is changed in steps of two when the supercoiling of a circular DNA is changed. The reactions require Mg(II), and ATP hydrolysis is required for enzyme turnover and rapid kinetics, although one cycle of relaxation or decatenation/catenation can occur in the presence of the nonhydrolyzable analog of ATP, ADPNP (50-adenylyl-Ø,° -imidodiphosphate).The crystal structures of several members, including the structurally distinct topoisomerase VI reveal that the active site tyrosines are situated in a helix-turn-helix (HTH) motif found within a domain that strongly resembles the DNA binding region of the *E. coli* catabolite activator protein (CAP).

All of the type II enzymes from both prokaryotic domains contain two different subunits and are therefore are heterotetrameric in structure, whereas the eukaryotic enzymes are homodimers. Among all of the known type II enzymes, DNA gyrase stands alone as the only enzyme capable of using the energy from ATP hydrolysis to introduce negative supercoils into the DNA. Finally, different members of the type II family can be distinguished by their relative proficiency at DNA relaxation versus decatenation (or catenation), and this property likely reflects their specialized roles in the cell.

DNA topoisomerase (*GyrB)*

DNA topoisomerases are enzymes essential for DNA replication, transcription, recombination and repair. They control the level of supercoiling by cleaving and resealing the phosphodiester backbone of DNA. The topoisomerases are classified into type I (EC 5.99.1.2) and type II (EC 5.99.1.3) according to their enzymatic properties.

The bacterial DNA gyrase is a type II topoisomerase that is capable of introducing negative supercoils into a relaxed closed circular DNA molecule. This reaction is coupled to the ATP hydrolysis. DNA gyrase can also relax supercoiled DNA without the ATP hydrolysis. The DNA gyrase comprises two proteins in the quaternary structure of A_2B_2.

The A protein (GyrA) has a molecular weight of approximately 100 kDa while the B protein (GyrB) has a

molecular weight of either 90 kDa or 70 kDa. The comparison of the structure of the 90 kDa class and that of the 70 kDa class revealed that the 90 kDa type has an insertion of about 170 amino acids in the region of residue 560 in the 70 kDa-type sequence. The N-terminal portion of the B protein was thought to catalyze the ATP-dependent supercoiling of DNA while the C-terminal portion to support the complexation with the A protein and the ATP-independent relaxation. The crystal structure of the N-terminal 43-kDa fragment of the B protein was resolved. The 43-kDa protein monomer comprises two domains, the ATP-binding site being located at the center of the first domain.

DNA gyrase introduces negative supercoiling into circular DNA by catalyzing the passage of one DNA segment through another. The efficiency of the reaction is many times higher than that of other topological transformations. We analyze, by a computer simulation, the reaction selectivity for a model of DNA gyrase action that assumes existence of a free loop between the G- and T- DNA segments participating in the reaction. A popular model of this type assumed that the selectivity can be provided by the conformation of the DNA segment wrapped around the enzyme into the right-handed helix turn (G-segment).

Conclusions

DNA gyrase is a bacterial enzyme that introduces negative supercoiling into circular DNA. It belongs to the type II DNA topoisomerases that catalyze passing one double-stranded DNA segment through another one. The free energy required for supercoiling comes from ATP hydrolysis coupled with the strand-passing reaction. The enzyme is important for maintaining a certain level of DNA supercoiling inside bacterial cells and for DNA replication [11].

The most appreciated advances in the biological component of drug development is a shifted to catalyze the strategies and tactics that underlie the drug discovery process [18]. New information has evolved to describe disease states at the molecular rather than organismic level, which in turn presents those involved in drug development with a large array of well-defined targets. Additionally, economic factors are driving the need for a shorter lead-to-drug development time. The most important step in preparing the target for virtual screening is the identification of the proper ligand binding site. Ideally, the ligand binding site is well-defined and capable of specifically binding a small molecule that will modulate its function. In many cases, such as enzymes, the targeted ligand binding site is well-known, in other cases, such as small molecules that disrupt protein: protein interactions, it is more obscure [19].

Once the virtual library is created and the target is prepared, including the specification of the ligand binding site, the library must be docked into the target site and evaluated for goodness-of-fit. The two stages represented in this step are 1) docking – the search for the conformation and configuration of the ligand in the binding site and 2) scoring-the evaluation of the interaction energy between the target and ligand. Many previous reviews have extensively covered the aspects of docking and scoring. This review will not serve to reiterate that material, but will cover some of the essential questions raised in considering a docking problem. Specifically, the flexibility of ligands during docking, the treatment of ligands as fragments and the flexibility of the target will be discussed.

Methods

Structure and function determination of proteins can be done using protein threading and homology-based prediction methods. Here, the prediction methods can exercise their strengths, which lie in being used interactively by experts and making suggestions that can be followed up by succeeding experimentation, rather than being required to provide proven fact. The process of going from structure to function is far from being automated.

Analysis of protein function

Protein function is predicted based on similarity. Sequence alignment tools like BLAST and PSI-BLAST are used for this purpose [24]. If the homology is above 40 % and above, functionally important motifs are conserved then its function are quiet similar to that of query sequence. As the level of similarity decreases, the conclusions on functions that can be drawn from sequence similarity become less and less reliable [25]. Protein functions are determined by Pfam, PROSITE, PRINTS, ProDom, SwissPROT + TREMBL [27-29, 33].

Analysis of protein structure

Computational methods for predicting protein structure from sequence, known as mini-threading, which are based on the assembly of fragments. In contrast, modeling protein structures after folds that have been seen before has become quite a powerful method for protein structure prediction [36]. Here, the query sequence is aligned (threaded) to a model sequence whose three-dimensional structure is known (the template protein). All proteins in a given protein structure database are usually, an appropriate representative set of structures are tried and each template is ranked using heuristic scoring functions. The score reflects the likelihood that the query sequence assumes the template structure. The approach of modelling a protein structure after a known template is called *homology-based modelling* and the selection of a suitable template protein is often done via *protein threading*.

There are strong efforts to render the quality of protein structure prediction methods more transparent and easier to evaluate. Protein structure prediction methods on *blind* predictions and aims at developing standardized and generally agreed upon assessment procedures both for fold identification and the evaluation of alignment accuracy as well as homology models. A blind prediction is a prediction of the three-dimensional structure for a protein sequence at a time, at which the actual structure of the protein is not known [31]. After the structure has been resolved, the prediction is compared with the actual structure.

Protein structure aspects are useful for drug design studies typically have to involve three-dimensional structure. Predicting the secondary structure of the protein is done for the analysis of the structure. Even the similarity of the three dimensional structures of two proteins cannot be taken as an indication for a similar function of these proteins. The molecular function of the protein is tied to local structural characteristics pertaining to binding pockets on the protein surface [1].

Drug design based on protein structure

The most important objective of drug design is to find or develop a, mostly small, drug molecule that tightly binds to

the target protein, moderating its function or competing with natural substrates of the protein. Such a drug can be best found on the basis of knowledge of the protein structure. If the spatial shape of the site of the protein is known, to which the drug is supposed to bind, then docking methods can be applied to select suitable receptor compounds that have the potential of being fitted to drugs. In order to screen really large drug databases with several hundred thousand compounds docking methods that can handle single protein/drug pairs within seconds are used. This energy model is hence called as *scoring* function that rates the protein–ligand complex energetically.

Authors Contributions

SSM participated in the conception and design of the study, in the acquisition, analysis and interpretation of data, carried out the bioinformatics tasks and wrote the draft of the manuscript.

PV participated in the conception and design of the study, in the analysis of the data, and in the critical review of the manuscript.

Figures

[743 residues]

Trusted matches - domains scoring higher than the gathering threshold (A)

Domain	Start	End	Bits	Evalue	Mode
HATPase_c	3	113	43.72	5.3e-10	ls
DNA_gyraseB	161	331	330.61	2.4e-97	ls
Toprim	356	465	28.81	1.6e-06	ls
DNA_gyraseB_C	665	731	168.12	2e-46	ls

Matches to Pfam-B

Domain	Start	End
Pfam-B_228	359	412
Pfam-B_370	426	477
Pfam-B_3140	483	597
Pfam-B_2227	603	663

Potential matches - Domains with E values above the cutoff

Domain	Start	End	Bits	Evalue	Mode
PaRep2a	496	617	-6.21	0.36	ls
Hepar_II_III	124	153	5.91	0.34	fs
DUF199	272	285	4.82	0.25	fs
CN_hydrolase	498	512	2.93	0.92	fs
Phenol_Hydrox	553	583	3.21	0.82	fs
Glyco_transf_20	564	579	3.92	0.33	fs

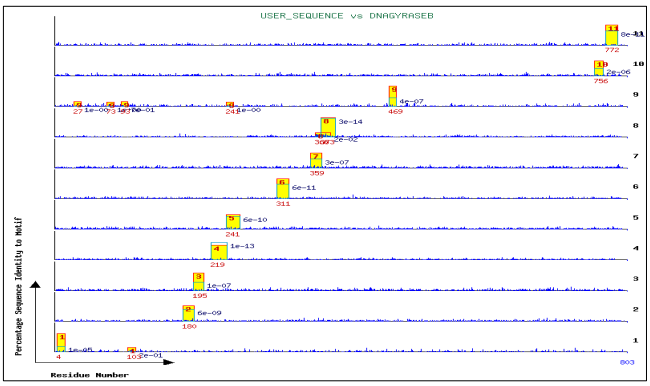


Fig 1: Analysis by Pfam.

Alignment of GyrB in *E.coli:*

Hits by patterns with a high probability of occurrence or by user-defined patterns: [41 hits (by 6 distinct patterns) on 1 sequence]

Fig 2: Analysis by PRINTS

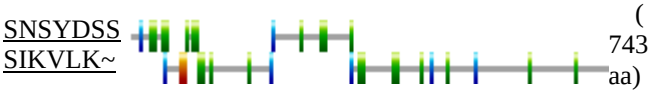


Fig 3: Analysis By PROSITE showing the six distinct patterns.

SNSYDSSSIKVLKGLDAVRKRPGMYIGDTDDGTGLH
HMVFEVVDNAIDEALAGHCKEIIVTIHADNSVSVQDD
GRGIPTGIHPEEGVSAAEVIMTVLHAGGKFDDNSYKV
SGGLHGVGVSVVNALSQKLELVIQREGKIHRQIYE----
HGVPQAPLAVTGETEKTGTMVRFWPSLETFTNVTEF
EYEILAKRLRELSFLNSGVSIRLRDKRDGKEDH---------
---------FHYEGGIKAFVEYL-NKNKTPIHPN-

IFYFSTEKDG--------IGVEVALQWN-
DGFQENIYCFTNNIPQRDGGTHLAGFRAAMTRTLNA
YMDKEGYSKKAK-
VSATGDDAREGLIAVVSVKVPDPKFSSQTKDKLVSSE
VKSAVEQQMNELLAEYLLENPTDAKIVVGKIIDAAR
AREAARRAREMTRRKGALDLAGLPGKLADCQERDP
ALSELYLVEGDSAGG

SAKQGRNRKNQAILPLKGKILNVEKARFDKMLSSQE
VATLITALGCGIGRDEYNPDKLRYHSIIIMTDADVDGS
HIRTLLLTFFYRQMPEIVERGHVYIAQPPLYKVKKGK
QEQYIKDDEAMDQYQISIALDGATLHTNASAPALAG
EALEKLVSEYNATQKMINRMERRYPKAMLKELIYQP
TLTEADLSDEQTVTRWVNALVSELNDKEQHGSQWK
FDVHTNAEQNLFEPIVRVRTHGVDTDYPLDHEFITGG
EYRRICTLGEKLRGLLEEDAFIERGERRQPVASFEQAL
DWLVKESRRGLSIQRYKGLGEMNPEQLWETTMDPES
RRMLRVTVKDAIAADQLFTTLMGDAVEPRRAFIEEN
ALKAANIDI

Fig 4: GyrB of *Ecoli* (Alignment using psi-BLAST).

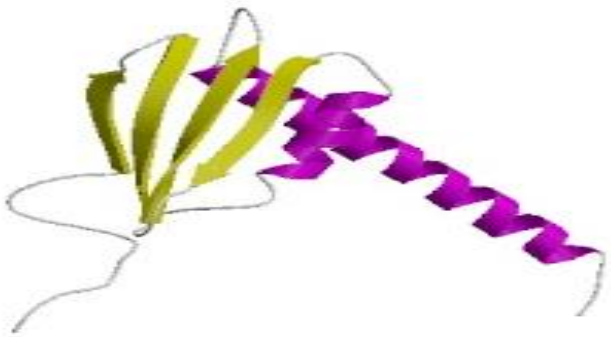
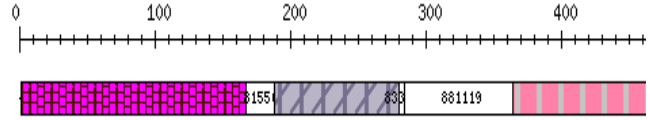


Fig 5: (Structure analysis by CATH showing the protein folds) [Topology (3.30.1360]

The following is the graphical representation of the HSP found by BLAST (The HSPs are sorted from highest to lowest scores, so that lower scoring HSPs may be hidden).



Align subsequence with ProDom domains

Fig 6: Graphical results and forms to other applications

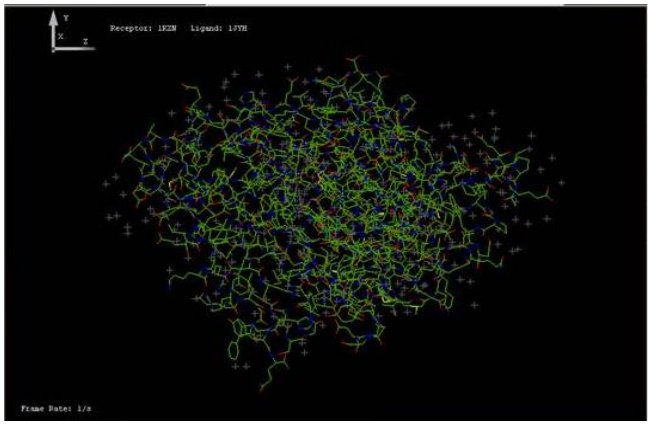


Fig 7: Docking using Hex.4.2. Crystal Structure of the Escherichia coli SbmC protein (AKA Gyrase Inhibitory Protein GyrI, AKA YeeB) (1KZN) and Crystal Structure of E. coli 24kDa Domain in Complex with Clorobiocin (1JYH) are used as receptor and ligand.

Table 1: (Analysis of gyrB using SCOP).

Serial No.	Properties of proteins analysed by SCOP.
1	Globin-like core: 6 helices; folded leaf, partly opened
2	Long alpha-hairpin 2 helices; antiparallel hairpin, left-handed twist
3	Type I dockerin domain tandem repeat of two calcium-binding loop-helix motifs, distinct from the EF-hand
4	LEM/SAP HeH motif helix-extended loop-helix; parallel helices
5	KRAB domain (Kruppel-associated box, Pfam 01352) 2 helices: one short, one long; aromatic-rich iterface
6	Cytochrome c core: 3 helices; folded leaf, opened
7	DNA/RNA-binding 3-helical bundle core: 3-helices; bundle, closed or partly opened, right-handed twist; up-and down
8	Another 3-helical bundle topologically similar to the DNA/RNA-binding bundles; distinct packing
9	RuvA C-terminal domain-like 3 helices; bundle, right-handed twist
10	S13-like H2TH domain core: 3-4 helices
11	Putative DNA-binding domain core: 3 helices; architecture is similar to that of the "winged helix" fold buttopology is different
12	Spectrin repeat-like 3 helices; bundle, closed, left-handed twist; up-and-down
13	immunoglobulin/albumin-binding domain-like 3 helices; bundle, closed, left-handed twist; up-and-down; mirror topology to the spectrin-like fold

Acknowledgements
We thank those who helped directly or indirectly to contribute the success of this work.

References

1. Tatiana A Soboleva, David A Jans. Melanie Johnson-Saliba, Proteins: Structure, Function, and Bioinformatics. 2004; 55(4):885-894. Wiley-Liss, Inc.).
2. Madhusudan K, Nagaraja V. J Biosci. 1996; 21(5):613-629.
3. Bates AD, A Maxwell. DNA gyrase can supercoil DNA circles as small as 174 base-pairs. EMBO J. 1989; 8:1861-1866.
4. Brian AA, HL Frisch, LS Lerman. Thermodynamics and equilibrium sedimentation analysis of the close approach of DNA molecules and a molecular ordering transition. Biopolymers. 1981; 20:1305-1328.
5. Brown PO, NR Cozzarelli. A sign inversion mechanism for enzymatic supercoiling of DNA. Science. 1979; 206:1081-1083.
6. Gellert M, K Mizuuchi, MHO' Dea, HA. Nash. DNA gyrase: an enzyme that introduces superhelical turns into DNA. Proc. Natl. Acad. Sci. USA. 1976; 73:3872-3876.
7. Heddle JG, S Mitelheiser, A Maxwell, NH. Thomson. Ucleotide binding to DNA gyrase causes loss of DNA wrap. J Mol. Biol. 2004; 337:597-610.
8. Kirkegaard K, JC Wang. Mapping the topography of DNA wrapped around gyrase by nucleolytic and chemical probing of complexes of unique DNA sequences. Cell. 1981; 23:721-729.
9. Levene SD, DM Crothers. Topological distributions and the torsional rigidity of DNA. A Monte Carlo study of DNA circles. J Mol. Biol. 1986; 189:73-83.
10. Liu LF, JC Wang. DNA-DNA gyrase complex: the wrapping of the DNA duplex outside the enzyme. Cell. 1978a; 15:979-984.
11. 11) Orphanides G, A Maxwell. Evidence for a conformational change in the DNA gyrase-DNA complex from hydroxyl radical footprinting. Nucleic Acids Res. 1994; 22:1567-1575.
12. Reece RJ, A Maxwell. DNA gyrase: structure and function. Crit. Rev. Biochem. Mol. Biol. 1991; 26:335-375.
13. Wang JC. DNA topoisomerases. Annu. Rev. Biochem. 1996; 65:635-695.
14. Vologodskii AV, W Zhang, VV Rybenkov, AA Podtelezhnikov D. Subramanian JD, *et al*. Mechanism of topology simplification by type II DNA topoisomerases. Proc. Natl. Acad. Sci. USA. 2001; 98:3045-3049.
15. Hagerman PJ. Flexibility of DNA. Annu. Rev. Biophys. Biophys. Chem. 1988; 17:265-286.
16. Hammermann M, C Stainmaier, H Merlitz, U Kapp, W Waldeck, G Chirico, *et al*. Salt effects on the structure and internal dynamics of superhelical DNAs studied by light scattering and Brownian dynamic. Biophys. J. 1997; 73:2674-2687.
17. Klenin KV, AV Vologodskii, VV Anshelevich, AM Dykhne, MD Frank-Kamenetskii. Effect of excluded volume on topological properties of circular DNA. J Biomol. Struct. Dyn. 1988; 5:1173-1185.
18. Klenin KV, AV Vologodskii, VV Anshelevich, AM Dykhne, MD Frank-Kamenetskii. Computer simulation of DNA supercoiling. J. Mol. Biol. 1991; 217:413-419.
19. Vologodskii AV, NR Cozzarelli. Modeling of long-range electrostatic interactions in DNA. Biopolymers. 1995; 35:289-296.
20. Vologodskii AV, W Zhang, VV Rybenkov, AA Podtelezhnikov, D Subramanian, JD Griffith, *et al*. Mechanism of topology simplification by type II DNA topoisomerases. Proc. Natl. Acad. Sci. USA. 2001; 98:3045-3049.
21. Wang JC. Moving one DNA double helix through another by a type II DNA topoisomerase: the story of a simple molecular machine. Quart. Rev. Biophys. 1998; 31:107-144.
22. Peng H, KJ Marians. The interaction of Escherichia coli topoisomerase IV with DNA. J. Biol. Chem. 1995; 270:25286-25290. Vologodskii, A. V., A. V. Lukashin, MD. Frank-Kamenetskii, and V. V.
23. Rybenkov VV, NR Cozzarelli, AV Vologodskii. Probability of DNA knotting and the effective diameter of the DNA double helix. Proc. Natl. Acad. Sci. USA. 1993; 90:5307-5311.
24. Altschul SF, Gish W, Miller W, *et al*. 'Basic local alignment search tool', J Mol. Biol. 1990; 215(3):403-410. http://ncbi.nlm.nih. gov/BLAST/.
25. Altschul SF, Madden TL, Schaffer AA, *et al*. 'Gapped Blast and of protein database search programs', Nucleic Acids Res. 1997; 25(17):3389-3402. http: ncbi.nlm.nih.gov/blast/psiblast.cgi.
26. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 'The COG database: a tool for genome-scale analysis of protein functions and evolution', *Nucleic Acids Res*., Vol. 28(1), pp. 33–36.
27. 27) Tatusov, R. L., Koonin, E. V. and Lipman, D. J. (1997), 'A genomic perspective on protein families', *Science*, Vol. 278(5338), pp. 631–637.
28. 28) Corpet, F., Servant, F., Gouzy, J. and Kahn, D. (2000), 'ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons', Nucleic Acids Res. 2000; 28(1):267-269.
29. Bateman A, Birney E, Durbin R, *et al*. 'The Pfam protein families database', Nucleic Acids Res. 2000; 28(1):263-266.
30. Schultz J, Milpetz F, Bork P, Ponting CP. 'SMART, a simple modular architecture research tool: identification of signaling domains', Proc. Natl Acad. Sci. USA. 1998; 95(11):5857-5864.
31. Schultz J, Copley RR, Doerks T, *et al*. 'SMART: A web-based tool for the study of genetically mobile domains', Nucleic Acids Res. 2000; 28(1):231-234.
32. Attwood TK, Croning MD, Flower DR, *et al*. 'PRINTS-S: the database formerly known as PRINTS', Nucleic Acids Res. 2000; 28(1):225-227.
33. Henikoff S, Henikoff JG, Pietrokovski S. 'Blocks+: A non-redundant database of protein alignment blocks derived from multiple compilations', Bioinformatics. 1999; 15(6):471-479.
34. Yona G, Linial N, Linial M. 'ProtoMap: Automatic classification of protein sequences and hierarchy of protein families', Nucleic Acids Res. 2000; 28(1):49-55.
35. Yona G, Linial N, Linial M. 'ProtoMap: automatic classification of protein sequences, a hierarchy of protein families, and local maps of the protein space', Proteins. 1999; 37(3):360-378.
36. Rose GD. 'Hierarchic organization of domains in globular proteins', J Mol. Biol. 1979; 134(3):447-470.