PROTEINS: Structure, Function, and Genetics 47:258–264 (2002)

PREDICTION REPORT

C-Terminal Domain of Gyrase A Is Predicted to Have a β-Propeller Structure

Yuan Qi,[2] Jimin Pei,[2] and Nick V. Grishin[1,2]*
[1]*Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas*
[2]*Department of Biochemistry, University of Texas Southwestern Medical Center, Texas*

ABSTRACT **Two different type II topoisomerases are known in bacteria. DNA gyrase (Gyr) introduces negative supercoils into DNA. Topoisomerase IV (Par) relaxes DNA supercoils. GyrA and ParC subunits of bacterial type II topoisomerases are involved in breakage and reunion of DNA. The spatial structure of the C-terminal fragment in GyrA/ParC is not available. We infer homology between the C-terminal domain of GyrA/ParC and a regulator of chromosome condensation (RCC1), a eukaryotic protein that functions as a guanine-nucleotide-exchange factor for the nuclear G protein Ran. This homology, complemented by detection of 6 sequence repeats with 4 predicted β-strands each in GyrA/ParC sequences, allows us to predict that the GyrA/ParC C-terminal domain folds into a 6-bladed β-propeller. The prediction rationalizes available experimental data and sheds light on the spatial properties of the largest topoisomerase domain that lacks structural information. Proteins 2002; 47:258–264.** © 2002 Wiley-Liss, Inc.

Key words: topoisomerase; regulator of chromosome condensation; remote homology detection; non specific DNA binding; ParC; GyrA

INTRODUCTION

Topoisomerases are ubiquitous enzymes that catalyze cleavage and religation of DNA molecules allowing for the changes in DNA topological states.[1,2] Topoisomerases are involved in crucial cellular processes such as replication, transcription, and recombination, and thus have pharmaceutical importance.[3,4] Topoisomerases of type I and type II cleave one and two DNA strands, respectively. Type II enzymes require ATP for their activity and possess an ATPase domain or subunit. Most bacteria have two homologous type II enzymes: DNA gyrase (topoisomerase II, Gyr) and topoisomerase IV (Par). Each enzyme is composed of two subunits (Fig. 1). GyrA is involved in breakage and reunion of DNA and GyrB functions as an ATPase. Equivalent subunits in topoisomerase IV, ParC and ParE, share about 35% identity with GyrA and GyrB. Despite pronounced sequence similarity, gyrase and topo IV possess distinct cellular functions.[5,6] Gyrase introduces negative supercoils into DNA. Topo IV relaxes negative and positive DNA supercoils.[6]

The reaction mechanism of type II topoisomerases is relatively well understood and crystal structures for most of their domains are available. GyrB can be divided into two fragments [Fig. 1(a)]. The 43-kDa N-terminal portion of the *Escherichia coli* enzyme with known structure is composed of an ATPase domain related to MutL/Hsp90/histidine kinase and a ribosomal protein S5-like domain.[6–8] The 47-kDa C-terminal portion consists of a toprim Rossmann-like domain interrupted by an insertion and is homologous to the N-terminal segment of the yeast topoisomerase II with available structure.[9,10] Domain architecture of ParE is similar except that the insertion in the toprim domain is shorter [Fig. 1(b)].

GyrA is also composed of two fragments [Fig. 1(a)]. The structure of the 59K N-terminal fragment for *E. coli* enzyme has been determined and the position of the catalytic tyrosine has been localized.[11] The C-terminal 38K fragment of GyrA still remains the largest piece of the topoisomerase sequence without structural information. It has been shown that the C-terminal fragment can be expressed separately. It lacks catalytic activity, but can complement the N-terminal fragment upon mixing, which increases its supercoiling activity.[12] The C-terminal fragment acts as a non-specific DNA-binding protein and is probably involved in stabilization of the DNA-topoisomerase complex.[12] Without spatial structure information, this fragment remains poorly understood.

Here, using consensus of probabilistic sequence comparison methods combined with hydrophobicity analysis, we detect sequence similarity between the C-terminal domain of bacterial gyrase A and regulator of chromosome condensation (RCC1)[13] and infer homology between them. We predict that GyrA/ParC C-terminal domain folds as a

*Correspondence to: Nick V. Grishin, Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390-9050. E-mail: grishin@chop.swmed.edu

Received 30 November 2001; Accepted 30 November 2001

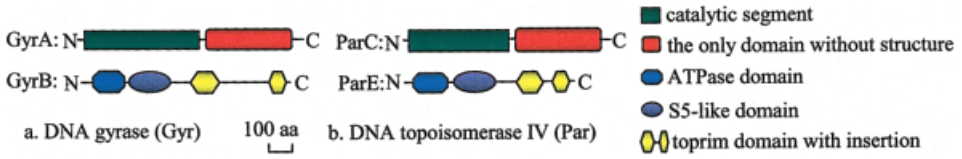

Fig. 1. Domain structures of (**a**) gyrase (topoisomerase II Gyr) and (**b**) topoisomerase IV (Par). Sequences shown are all from *Escherichia coli.*

6-bladed β-propeller. Functional implications of this homology prediction are discussed.

Regulator of chromosome condensation (RCC1) is the guanine-nucleotide-exchange factor for the nuclear G protein, Ran, which controls nucleocytoplasmic transport, mitotic spindle formation, and nuclear envelope assembly.[14] These functions depend on the association of RCC1 with DNA. Mutations in the yeast RCC1 gene affect pre-messenger RNA processing and transport, mating, initiation of mitosis, and chromatin decondensation. The crystal structure of RCC1 revealed that the molecule folds as a 7-bladed β-propeller, composed of seven four-stranded β-sheets (blades) arranged in a circular array.[13] The β-propeller proteins vary in the number of blades (from 4 to 8), share limited sequence similarity despite pronounced structural resemblance, and display extreme functional diversity.[15]

MATERIALS AND METHODS

Sequence Similarity Searches, Multiple Alignment, and Hydrophobicity Analysis

The PSI-BLAST program was used to search for homologues of the gyrase C-terminal fragment.[16] Residues 510–836 of *Mycoplasma genitalium* GyrA (gi|1346233) were selected as a query to search against the nonredundant (nr) database at NCBI (February 2001, 616,977 sequences, 195,057,269 total letters). The E-value threshold was set to 0.02. All other parameters were defaults.[16] PSI-BLAST was iterated until convergence. Found homologues were grouped by single linkage clustering (BLAST score threshold of 1 bit per site corresponding to about 50% identity) as implemented in the SEALS package,[17] and the representative sequences were used as new queries for subsequent PSI-BLAST iterations. Multiple sequence alignments were constructed using the T-COFFEE program[18] and adjusted manually based on the secondary structure predictions (discussed below) and the conserved residue patterns. Alignments for topo II sequences and RCC1 sequences were made separately and then merged based on the PSI-BLAST local alignments and hydrophobicity profiles. Propeller blades corresponding to sequence repeats were aligned to each other. The average hydrophobicity of residues in each of the four β-strands of the blades was calculated separately for topo II and RCC1 alignments using the scale from the mean values of 127 different hydrophobicity scales.[19]

Secondary Structure Prediction and Threading

Five representative (most diverse) topoisomerase C-terminal domain sequences (gi|68494, residues 537–875; gi|1346229, residues 538–922; gi|1346233, residues 514–836; gi|1835202, residues 528–907; gi|6655026, residues 517–755) were submitted to the JPRED2 consensus secondary structure prediction server (http://jura.ebi.ac.uk:8888/),[20] which returns the consensus of prediction results for six different secondary structure prediction methods, including PHD, NNSSP, DSC, PREDATOR, MULPRED, and ZPRED. These five sequences were also submitted to another secondary structure prediction server, SAM-T99 (http://www.cse.ucsc.edu/research/compbio/HMM-apps/T99-query.html).[21] JPRED2 secondary structure predictions were also carried out for 3 RCC1 sequences (gi|12325184, gi|132174, gi|7493765). Seven fold recognition (threading) methods were applied to five representatives of the gyrase C-terminal domain (gi|68494, gi|1346229, gi|121882, gi|1346235, gi|729651). The following methods were explored: (1) the hybrid fold recognition method of Fischer at the BioInBgu server (http://www.cs.bgu.ac.il/~bioinbgu/)[22]; (2) a method that combines multiple sequence profiles and knowledge of protein structures to provide enhanced recognition at the 3D-PSSM (three-dimension position-specific scoring matrix) server (http://www.bmm.icnet.uk/~3dpssm/)[23]; (3) the GenTHREADER program at the PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/)[24]; (4) Sausage (Sequence-Structure Alignment Using a Statistical Approach Guided by Experiment) server (http://rsc.anu.edu.au/~drsnag/TheSausageMachine.html)[25]; (5) the secondary structure prediction-based fold recognition server, TOPITS (http://www.embl-heidelberg.de/predictprotein/predictprotein.html)[26,27]; (6) FFAS (Fold and Function Assignment System) server (http://bioinformatics.ljcrf.edu/FFAS/)[28]; (7) sequence-structure homology recognition server that uses environment-specific substitution tables and structure-dependent gap penalties, FUGUE, at http://www-cryst.bioc.cam.ac.uk/~fugue/.[29]

RESULTS

PSI-BLAST Searches

GyrA and ParC homologues were found in PSI-BLAST searches initiated from the C-terminal fragment of *M. genitalium* GyrA as described in Materials and Methods. Inspection of local alignments generated by PSI-BLAST revealed the presence of multiple high-scoring pairs (HSPs) for as many as 80% of the found homologues, indicating the presence of sequence repeats. In other words, the same segment of the query sequence was aligned to several different segments in the same subject sequence with reliably high E-values (below 0.02). Multiple alignment

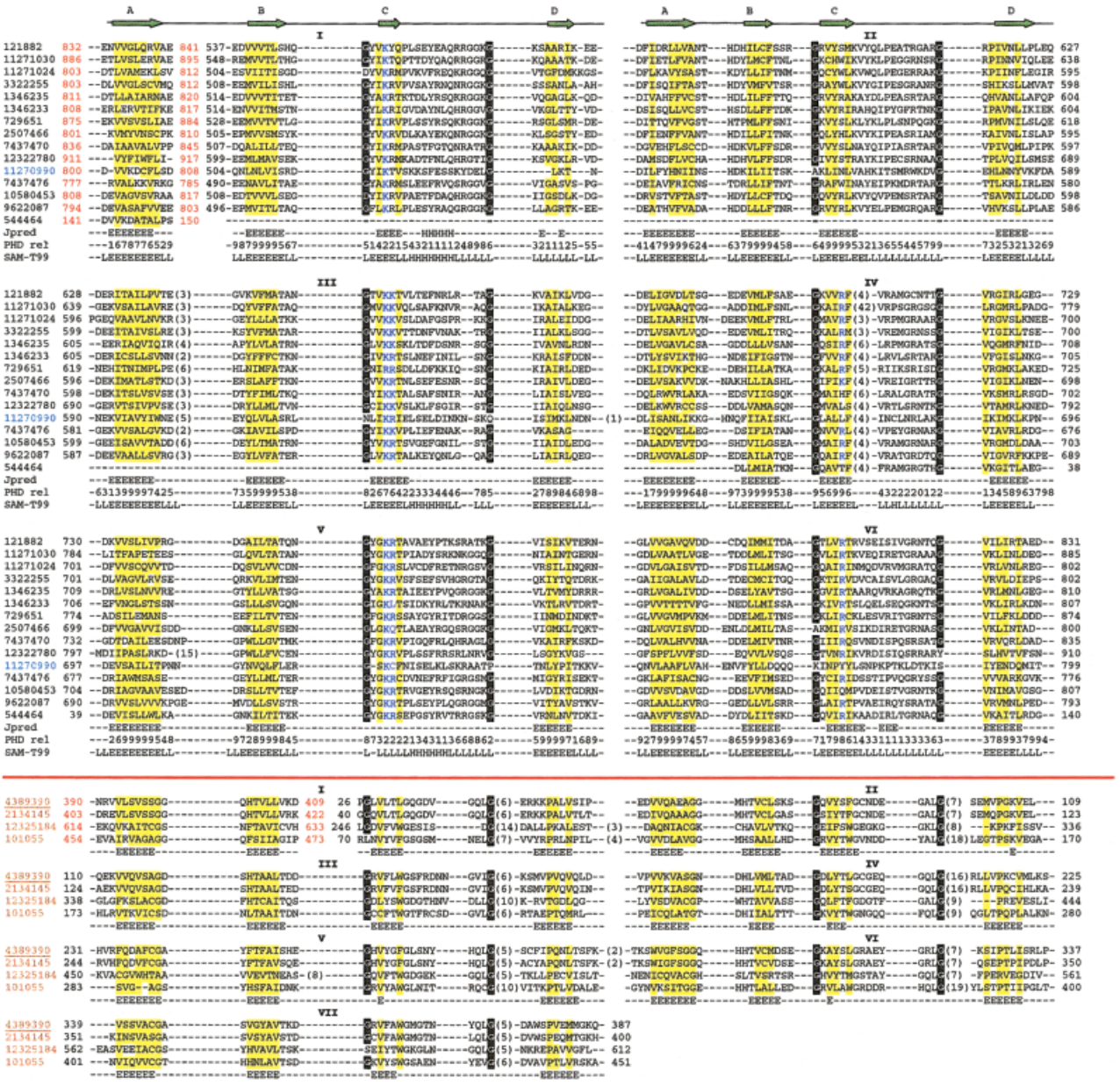

Fig. 2. Multiple sequence alignment of C-terminal of gyrase subunit A/RCC1 domain. Each sequence is labeled by its NCBI gene identification (gi) number. The gi numbers of GyrA/ParC C-terminal domains and the gi numbers of RCC1 domains are in black and brown, respectively. The gi number of the topoisomerase IV subunit A sequence (gi|11270990) is in blue. The gi number of the sequence with known structure is underlined (gi|4389390; PDB entry 1A12, chain A). The alignment is arranged in such a way that each row of sequences contains two blades. The blades are numbered from above using Roman numbers. The last β-strand A of GyrA/ParC C-terminal domain was placed in front of the first β-strand B of blade I to complete the blade. The sequences of RCC1 were rearranged in the same manner. The first and last residue numbers of each row of sequences are indicated. The first and last residue numbers of the rearranged C-terminal segments are marked in red. Long insertions in loop regions are not shown, but the numbers of the omitted residues are in parentheses. Uncharged residues at mainly hydrophobic positions are shaded yellow. The conserved glycine residues are shown in white on a black background. Conserved positively charged residues in β-strand C are shown in blue. The JPRED secondary structure prediction results are the first lines shown below each row of the alignment. The PHD prediction confidence values of every position for GyrA/ParC are shown on the second line under the predictions. The third lines under the GyrA/ParC alignment are the secondary prediction results from SAM-T99. The diagram of the secondary structure elements in each blade, according to the RCC1 X-ray structure, is shown at the top. Species names: gi|121882, *Escherichia coli*; gi|11271030, *Neisseria meningitides*; gi|11271024, *Chlamydia muridarum*; gi|3322255, *Treponema pallidum*; gi|1346235, *Mycobacterium tuberculosis*; gi|1346233, *Mycoplasma genitalium*; gi|729651, *Rickettsia prowazekii*; gi|2507466, *Helicobacter pylori*; gi|7437470, *Synechocystis sp*; gi|12322780, *Arabidopsis thaliana*; gi|11270990, *Ureaplasma urealyticum*; gi|7437476, *Archaeoglobus fulgidus*; gi|10580453, *Halobacterium sp.*; gi|9622087, *Thermus thermophilus*; gi|544464, *Fibrobacter succinogenes*; gi|4389390, *Homo sapiens*; gi|2134145, *African clawed frog*; gi|12325184, *Arabidopsis thaliana*; gi|101055, *Schizosaccharomyces pombe*.

analysis established the presence of 6 sequence repeats in GyrA and ParC C-terminal domains (Fig. 2).

PSI-BLAST iterations initiated from most of the GyrA and ParC sequences converged within the type II topoisomerase family and did not result in structural predictions. However, the 3rd iteration with the query gi|544464, which is annotated as *Fibrobacter succinogenes* GyrA, yielded one non-topoisomerase sequence with an E-value

of 0.017 (bit score 40, NCBI nr database, September 2001, 751,829 sequences, 239,148,880 total letters). This sequence, human cell cycle regulatory protein (gi|87057, residues 80–206), is a variant of human RCC1, which has a known three-dimensional structure (gi|4389390, PDB entry 1a12)[13,30] and can offer a fold prediction for the C-terminal fragment of GyrA/ParC. RCC1 folds as a 7-bladed β-propeller, with blades being coded by sequence repeats. Each blade is composed of 4 antiparallel β-strands. No sequences from other families were found with significant E-values.

Secondary Structure Predictions and Fold Recognition

JPRED2 secondary structure predictions[20] obtained for several gyrase sequences strongly suggest that they are all-beta proteins (Fig. 2). Most of the β-strands were predicted with high confidence level (PHD confidence 7–9; Fig. 2). SAM-T99 secondary structure prediction yielded similar results (Fig. 2). β-Strands 5 residues long on average were predicted along the sequence with spacing of about 2–20 residues between them. Secondary structure predictions were similar for the sequence repeats with the consensus prediction of 4 β-strands per repeat (Fig. 2). The secondary structure prediction for RCC1 sequences were similar and in agreement with the crystal structure of RCC1. Furthermore, the secondary structure predictions show an excellent correspondence between the GyrA/ParC C-terminal domain and RCC1 families.

The consensus fold recognition method of Fischer that combines sequence, structural, and evolutionary information[22] was applied to several topoisomerase II sequences. Seven- or six-bladed β-propellers were consistently found as the top scoring proteins. For instance, the top three fold recognition hits for gyrase gi|121882 are: a theoretical model of human nidogen ywtd β-propeller domain (PDB entry 1NDX, score 17.8); C-terminal WD40 domain of tup1 (PDB entry 1ERJ, score 17.4); and phytase from *Bacillus amyloliquefaciens* (PDB entry 1CVM, score 13.0). There is a substantial gap in the consensus scores between the top three hits and the fourth one with the score of 5.7, which suggests that no other known fold "fits" the gyrase sequence well. In the results from 3D-PSSM, 6- or 7-bladed β-propellers were the top scoring protein folds with 0.05–0.5 PSSM E-values and 90–50% certainty. Furthermore, query gi|1346229 found RCC1 at PSSM E-value of 0.533, with 50% certainty. The results from FFAS also showed 7- or 6-bladed propeller as top hits. gi|68494 found RCC1 as the second hit with E-value of 13.4 and Z-score of 6.02. gi|121882 found RCC1 as the third hits with E-value of 31.2 and Z-score 5.76. FUGUE also found 7- or 4-bladed β-propellers as top hits, but failed to find RCC1. Sausage found β-propellers and antiparallel β-sheet proteins as top hits for the majority of the query sequences. For gi|1346229, it found RCC1 as the top hit with a score of 3.31. TOPITS and GenTHREADER did not find β-propellers; other mostly β-sheet proteins were the top hits with marginal statistics.

Multiple Sequence Alignment

PSI-BLAST searches demonstrated that sequence repeats in GyrA/ParC are more similar to each other than to repeats in other proteins. Thus, GyrA/ParC repeats should be more easy to align with each other. RCC1 family was the only group that displayed statistically supported sequence similarity (PSI-BLAST E-value of 0.017, 12–29% identity) to GyrA/ParC repeats. Therefore, we selected RCC1 for more detailed analysis. To probe further potential homology between the GyrA/ParC C-terminal domain and RCC1, a multiple sequence alignment was constructed (Fig. 2). The alignment confirmed the presence of 6 repeats in GyrA/ParC sequences. Each repeat was predicted to contain 4 β-strands (A to D). Loops were relatively short (2–6 residues) between all but two β-strands. Only between β-strands C and D loops were longer (typically about 15 residues). The alignment revealed conservation of hydrophobic residues in β-strands, conserved positively charged residues in β-strand C, and a pair of conserved small residues (typically glycines) in each repeat (Fig. 2).

TABLE I. Average Hydrophobicity of β Strands in GyrA/ParC and RCC1

Strand	GyrA/ParC	RCC1
A	0.21	0.256
B	0.37	0.34
C	0.17	0.29
D	0.085	−0.13

The alignment of the RCC1 family was constructed independently and showed 7 sequence repeats with 4 predicted β-strands in each repeat in agreement with the crystal structure of human RCC1. The alignments of GyrA/ParC and RCC1 were merged on the basis of PSI-BLAST local alignments that superimposed the long loops between the strands C and D (Fig. 2). Such alignment results in a different placement of the Velcro of the propeller in GyrA/ParC and RCC1. In RCC1, Velcro is between the strands B and C. GyrA/ParC are predicted to have a Velcro between A and B. To obtain additional support for the register of β-strands between GyrA/ParC and RCC1, average hydrophobicities were calculated for each β-strand in GyrA/ParC and RCC1 (Table I). Comparison of the 4 hydrophobicity values confirms the alignment of β-strands and thus Velcro placement in GyrA/ParC.

DISCUSSION

Validity of the Fold Prediction

The results of PSI-BLAST searches, secondary structure predictions, fold recognition, and multiple alignment analysis allow us to deduce the fold of the GyrA/ParC C-terminal fragment. The presence of 6 sequence repeats with 4 predicted β-strands each (Fig. 2), the PSI-BLAST hit to RCC1, and the detection of propeller folds with threading method strongly argue that the GyrA/ParC domain adopts the 6-bladed β-propeller structure.

Proper alignment of the GyrA/ParC sequences with the RCC1 structure is challenging because of the low level of sequence similarity. Most importantly, corresponding β-strands in GyrA/ParC and RCC1 should be found and correctly aligned. Due to repetitive sequences in GyrA/ParC and the hydrophobic character of β-strands, it is potentially possible to miss the register of β-strands and to

align a β-strand in GyrA/ParC to a non-equivalent β-strand in RCC1. For instance, the inner β-strand of the propeller blade may be incorrectly aligned with the outer β-strand.

Three lines of evidence support the alignment presented in Figure 2. First, it matches pairwise alignments produced by automatic tools such as PSI-BLAST and the fold recognition method of Fischer. Second, the longest loop between the strands (C and D) in GyrA/ParC is aligned with the longest loop between the strands in RCC1. Third, and most importantly, hydrophobicity analysis of β-strands reveals correspondence in patterns between GyrA/ParC and RCC1 (Table I). Each blade of the propeller is composed of 4 β-strands (A, B, C, D). Since these β-strands are placed at non-equivalent positions in the overall circular structure of the propeller [Fig. 3(a)], average hydrophobicities of these 4 β-strands differ. The β-strand D is the outermost strand, and it is the most exposed. Thus, the β-strand D is expected to be the most hydrophilic. The β-strand A is the innermost strand located along the central shaft of the propeller. The shaft of the propeller contains water molecules and thus the β-strand A is not expected to be the most hydrophobic. The β-strand B is the one with the highest hydrophobicity (Table I). Excellent fulfillment of these tendencies in GyrA/ParC and RCC1 families strongly supports the alignment on Figure 2.

Structural Differences Between GyrA/ParC and RCC1

Typically, homology-based predictions can deduce only similarities between the query and its homologue with experimentally determined structure. The differences are more challenging to predict. Some differences may be wrongly missed and similarities falsely predicted instead. Such bias is more likely to occur at very low sequence similarity levels when homology is remote. This is the case with the GyrA/ParC-RCC1 superfamily. Here, we argue that the two most important differences between GyrA/ParC and RCC1 can be predicted.

First, GyrA/ParC should fold as a 6-stranded propeller rather than a 7-stranded propeller as RCC1. This simply follows from the fact that only 6 sequence repeats can be detected in GyrA/ParC sequences. The sequences outside the 6-repeat fragment either belong to the domain of determined structure (N-terminal to the first repeat) or lack clearly predicted β-strands (the extreme C-terminal region). Additionally, the fragment of GyrA that corresponds exactly to the 6 repeats is naturally expressed in *Borrelia burgdorferi* (see discussion below).[31] Homology between propellers that display a different number of blades have been reported before[32] and, therefore, is not surprising.

Second, the Velcro position should differ between GyrA/ParC and RCC1 propellers (Fig. 2). In RCC1, the first blade starts from β-strand C and the last blade ends with the β-strand B. Thus, one half of the first blade is made from the N-terminal β-strands of the protein, and the other half is made from the C-terminal β-strands (2+2 Velcro). Such an assembly is favorable for the stabilization of the circular arrangement of blades. Since the first repeat of GyrA/ParC starts from the β-strand B and the last repeat ends with the β-strand A, the stabilization of the propeller circular arrangement is probably achieved by a 1+3 rather than 2+2 combination of β-strands. This 1+3 Velcro is known for other propellers such as methylamine dehydrogenase (PDB entry 2BBK), nitrite reductase (PDB entry 1NIR), and tachylectin-2 (PDB entry 1TL2); however, 2+2 Velcro of RCC1 is apparently unique.[15]

Functional Implications

The GyrA/ParC C-terminal domain remains the longest sequence segment of topoisomerase II without available structural information. Therefore, the function of this domain is not fully understood despite some effort in this direction. The structure prediction presented here and homology of the GyrA/ParC domain with the RCC1 protein have several functional implications. The RCC1 molecule functions as a protein-binding and a DNA-binding module. One side of the propeller accommodates a protein (Ran) binding site, and the Ran-RCC1 complex structure is available.[30] It is believed that the opposite side of the propeller is involved in interactions with DNA. Available experimental information about GyrA/ParC C-terminal domain suggests similar properties. Being expressed separately, the GyrA domain can associate with the rest of the A subunit, thus possessing a protein binding site. GyrA domain lacks catalytic activity, but binds DNA in a non-sequence-specific manner. Therefore, it should have a nucleic acid binding site.

It has been demonstrated that *Borrelia burgdorferi* expresses a 34-kDa fragment translated from an abundant transcript initiated within the GyrA coding region.[31] This fragment corresponds exactly to the 6 blades of the predicted β-propeller structure, starting from the strand B and ending with the strand A. *Borrelia burgdorferi* gives a unique example, for prokaryotes, of constitutive expression of two proteins, one being a fragment of another, from the same open reading frame. It has been shown that a naturally synthesized transcript abundant in *B. burgdorferi* corresponding to the predicted β-propeller functions as a non-specific DNA-binding protein, forming higher-order nucleoprotein complexes.[31]

Our prediction allows researchers to visualize the distribution of residues in space for the C-terminal domain of GyrA/ParC, despite its unsolved structure. The structural diagram of the C-terminal domain of RCC1 is shown in Figure 3(a). We predict protein- and DNA-binding surfaces in GyrA/ParC to be similar to the ones in RCC1 [Fig. 3(b,c)]. One way to visualize sequence properties on a structure is to use conservation mapping.[33] The conservation in the blade-to-blade alignments of all available sequences of GyrA/ParC and RCC1 is mapped onto the structure of the 3rd blade in RCC1 [Fig. 3(b,c)]. Similarities in conservation between GyrA/ParC and RCC1 include mainly small residues (C,A,P,S,T) in loops. These residues bear potential structural importance. The most pronounced difference in conservation patterns of GyrA/ParC and RCC1 is due to the presence of a conserved residue stretch closer to the N-terminus of the β-strand C in

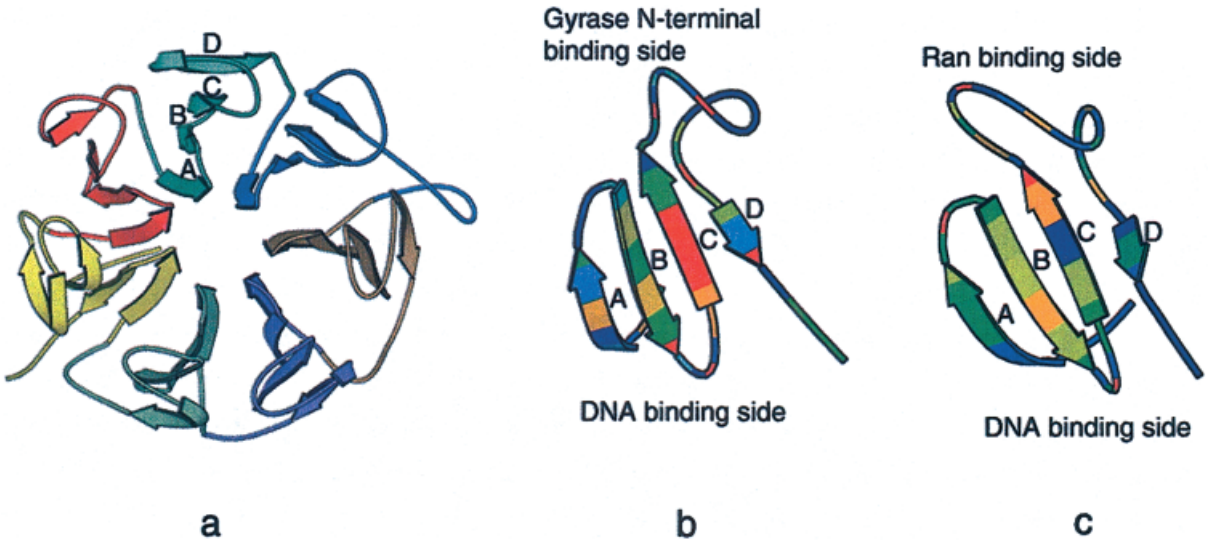

Fig. 3. **(a)** The structural diagram of RCC1, PDB entry 1A12 chain A. Each blade is shown in a different color and β-strands in the third blade are labeled. Sequence conservation in (**b**) GyrA/ParC and (**c**) RCC1 mapped onto the structure of the third blade in RCC1 are rainbow colored from low conservation (dark blue) to high conservation (red).

GyrA/ParC. These conserved residues are mainly positively charged (shown in blue in Fig. 2) and could potentially contribute to a DNA-binding site in GyrA/ParC.

In summary, using extensive sequence and structure analysis of the GyrA/ParC C-terminal domain and RCC1, we infer homology between these proteins and therefore predict the fold of the GyrA/ParC C-terminal domain, which remains the largest topoisomerase fragment without available structural information.

REFERENCES

1. Wang JC. DNA topoisomerases. Annu Rev Biochem 1996;65:635–692.
2. Caron PR, Wang JC. Appendix. II: Alignment of primary sequences of DNA topoisomerases. Adv Pharmacol 1994:271–297.
3. Maxwell A. The molecular basis of quinolone action. J Antimicrob Chemother 1992;30:409–414.
4. Hiasa H, Yousef DO, Marians KJ. DNA strand cleavage is required for replication fork arrest by a frozen topoisomerase-quinolone-DNA ternary complex. J Biol Chem 1996;271:26424–26429.
5. Zechiedrich EL, Khodursky AB, Bachellier S, Schneider R, Chen D, Lilley DM, Cozzarelli NR. Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. J Biol Chem 2000;275:8103–8113.
6. Deibler RW, Rahmati S, Zechiedrich EL. Topoisomerase IV, alone, unknots DNA in E. coli. Genes Dev 2001;15:748–761.
7. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
8. Lo Conte L, Ailey B, Hubbard TJ, Brenner SE, Murzin AG, Chothia C. SCOP: a structural classification of proteins database. Nucleic Acids Res 2000;28:257–259.
9. Aravind L, Leipe DD, Koonin EV. Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Nucleic Acids Res 1998;26:4205–4213.
10. Berger JM, Fass D, Wang JC, Harrison SC. Structural similarities between topoisomerases that cleave one or both DNA strands. Proc Natl Acad Sci USA 1998;95:7876–7881.
11. Morais Cabral JH, Jackson AP, Smith CV, Shikotra N, Maxwell A, Liddington RC. Crystal structure of the breakage-reunion domain of DNA gyrase. Nature 1997;388:903–906.
12. Reece RJ, Maxwell A. The C-terminal domain of the Escherichia coli DNA gyrase A subunit is a DNA-binding protein. Nucleic Acids Res 1991;19:1399–1405.
13. Renault L, Nassar N, Vetter I, Becker J, Klebe C, Roth M, Wittinghofer A. The 1.7 A crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven-bladed propeller. Nature 1998;392:97–101.
14. Nemergut M, Mizzen CA, Stukenberg T, Allis CD, Macara IG. Chromatin docking and exchange activity enhancement of RCC1 by histones H2A and H2B. Science 2001;292:1540–1543.
15. Paoli M. Protein folds propelled by diversity. Prog Biophys Mol Biol 2001;76:103–130.
16. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
17. Walker DR, Koonin EV. SEALS: a system for easy analysis of lots of sequences. Ismb 1997;5:333–339.
18. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000;302:205–217.
19. Palliser CC, Parry DA. Quantitative comparison of the ability of hydropathy scales to recognize surface beta-strands in proteins. Proteins 2001;42:243–55.
20. Cuff JA, Barton GJ. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins 2000;40:502–511.
21. Karplus K, Barrett C, Hughey R. Hidden Markov models for detecting remote protein homologies. Bioinformatics 1998;14:846–856.
22. Fischer D. Hybrid Fold Recognition: combining sequence derived properties with evolutionary information. Pacific Symp Biocomputing, Hawaii, January 2002. World Scientific. p 119–130.
23. Kelley LA, MacCallum RM, Sternberg MJ. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol 2000;299:499–520.
24. Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol 1999;287:797–815.
25. Huber T, Russell AJ, Ayers D, Torda AE. Sausage: protein threading with flexible force fields. Bioinformatics 1999;15:1064–1065.

26. Rost B. TOPITS: threading one-dimensional predictions into three-dimensional structures. Proc Int Conf Intell Syst Mol Biol 1995;3:314–321.
27. Rost B, Schneider R, Sander C. Protein fold recognition by prediction-based threading. J Mol Biol 1997;270:471–480.
28. Rychlewski L, Jaroszewski L, Li W, Godzik A. Comparison of sequence profiles. Strategies for structural predictions using sequence information. Protein Sci 2000;9:232–241.
29. Shi J, Blundell TL, Mizuguchi K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J Mol Biol 2001;310: 243–257.
30. Renault L, Kuhlmann J, Henkel A, Wittinghofer A. Structural basis for guanine nucleotide exchange on Ran by the regulator of chromosome condensation (RCC1). Cell 2001;105:245–255.
31. Knight SW, Samuels DS. Natural synthesis of a DNA-binding protein from the C-terminal domain of DNA gyrase A in Borrelia burgdorferi. Embo J 1999;18:4875–4881.
32. Wolf YI, Brenner SE, Bash PA, Koonin EV. Distribution of protein folds in the three superkingdoms of life. Genome Res 1999;9:17–26.
33. Pei J, Grishin NV. AL2CO: calculation of positional conservation in a protein sequence alignment. Bioinformatics 2001;17:700–712.