JOURNAL OF BACTERIOLOGY, Apr. 1996, p. 1881–1894 Vol. 178, No. 7
0021-9193/96/$04.00+0
Copyright © 1996, American Society for Microbiology

Evolutionary Conservation of RecA Genes in Relation to Protein Structure and Function

SAMUEL KARLIN* AND LUCIANO BROCCHIERI

Department of Mathematics, Stanford University, Stanford, California 94305-2125

Received 19 October 1995/Accepted 12 January 1996

Functional and structural regions inferred from the *Escherichia coli* RecA protein crystal structure and mutation studies are evaluated in terms of evolutionary conservation across 63 RecA eubacterial sequences. Two paramount segments invariant in specific amino acids correspond to the ATP-binding A site and the functionally unassigned segment from residues 145 to 149 immediately carboxyl to the ATP hydrolysis B site. Not only are residues 145 to 149 conserved individually, but also all three-dimensional structural neighbors of these residues are invariant, strongly attesting to the functional or structural importance of this segment. The conservation of charged residues at the monomer-monomer interface, emphasizing basic residues on one surface and acidic residues on the other, suggests that RecA monomer polymerization is substantially mediated by electrostatic interactions. Different patterns of conservation also allow determination of regions proposed to interact with DNA, of LexA binding sites, and of filament-filament contact regions. Amino acid conservation is also compared with activities and properties of certain RecA protein mutants. Arginine 243 and its strongly cationic structural environment are proposed as the major site of competition for DNA and LexA binding to RecA. The conserved acidic and glycine residues of the disordered loop L1 and its proximity to the RecA acidic monomer interface suggest its involvement in monomer-monomer interactions rather than DNA binding. The conservation of various RecA positions and regions suggests a model for RecA–double-stranded DNA interaction and other functional and structural assignments.

The RecA protein promotes and participates in many functions, including an essential role in homologous recombination, DNA strand exchange, DNA repair, and coprotease activity in response to DNA damage resulting in the SOS response, prophage induction, and/or mutagenesis subsequent to LexA cleavage (for reviews, see references 8, 25, 26, and 36). RecA is substantially conserved across bacterial organisms. Sixty-three RecA sequences are currently available, including 37 proteobacterial (Proteo) sequences, 11 gram-positive bacterial [Gram(+)] sequences, 3 mycoplasma sequences, 3 cyanobacterial sequences, 3 *Deinococcus-Thermus* sequences, and 6 singular cases. Similarity among RecA sequences was previously assessed by the significant segment pair alignment method (20). The extent of similarity among RecA sequences ranges from 43 to 100% at the amino acid level.

The X-ray crystal structure of the *Escherichia coli* RecA polymer complexed to ADP has been determined (38, 39). Physical and chemical analyses of RecA mutants (more than 250 amino acid substitutions) provide information pertinent to RecA primary sequence and function (23–28, 30, 36, 37, 43, 44). On the basis of structural, biochemical, and mutation studies, several key function and structure regions have been proposed, including positions for binding and hydrolyzing an ATP cofactor, DNA binding domains, monomer-monomer (M-M) polymerization surfaces, places of filament-filament contacts, and binding sites associated with the target proteins LexA and UmuD (29, 33, 38, 39). Another source of information for correlating sequence and function derives from the analysis of conservation among RecA sequences across the eubacterial kingdom. The alignments of 63 RecA eubacterial protein sequences coupled to information on conservation and variation of corresponding structural neighboring residues provide a challenging opportunity to integrate sequence (evolutionary relationships), structure, function, and mutagenesis information now available.

Evolution provides a virtual laboratory for mutation experiments. In this context, we investigate which residue positions and which positions of their structural environments are totally invariant, which positions show only similar amino acid changes, which positions are highly variable, and which positions show restricted variability. How do the conserved and variable residues relate to known or putative structural and functional aspects of the protein? These same inquiries apply with respect to various bacterial subclasses. For example, it is of interest to highlight and interpret strong group identities and strong group differences between the Proteo and Gram(+) sequences. Our observations on sequence conservation for various RecA positions and regions suggest in some cases (see Discussion) alternative functional and structural assignments and models.

MATERIALS AND METHODS

Alignment. The significant segment pair alignment method (20) effectively provides global alignments across all RecA sequences (Table 1). The explicit RecA species groupings developed in reference 20 are indicated in Table 1 (the leftmost column). Consonant with the high conservation among eubacterial RecA sequences, for all 63 sequences the alignments apply to positions 14 to 312 (referring to *E. coli* coordinates, numbering starting with 0 for the initiator methionine) with only a few scattered gaps. At each position of Table 1 a consensus amino acid is indicated, identifying the most frequent residue in that position. Consensus sequences are also determined separately for the 33 most classical A+B+C (Table 1, first column) Proteo sequences and for the 11 Gram(+) sequences. Differences among these consensus sequences and the specific *E. coli* sequence are displayed in Fig. 2.

Assessments of conservation for each RecA aligned position. A useful statistic is the degree of conservation at each position in the alignment. To this end, let $s(i, j)$ denote a measure of similarity between amino acid i and amino acid j. For example, the conservative Dayhoff evolutionary PAM120 substitutability matrix components are suitable (1, 10). A more sanguine similarity scoring matrix is BLOSUM-62 (15) (consult reference 17 for other determinations). We use PAM120, but the results are qualitatively analogous using BLOSUM-62.

The similarity assignments are normalized to $s^*(i, j) = s(i, j)/\sqrt{s(i, i)s(j, j)}$ so

* Corresponding author.

TABLE 1. RecA species sequence alignment[a]

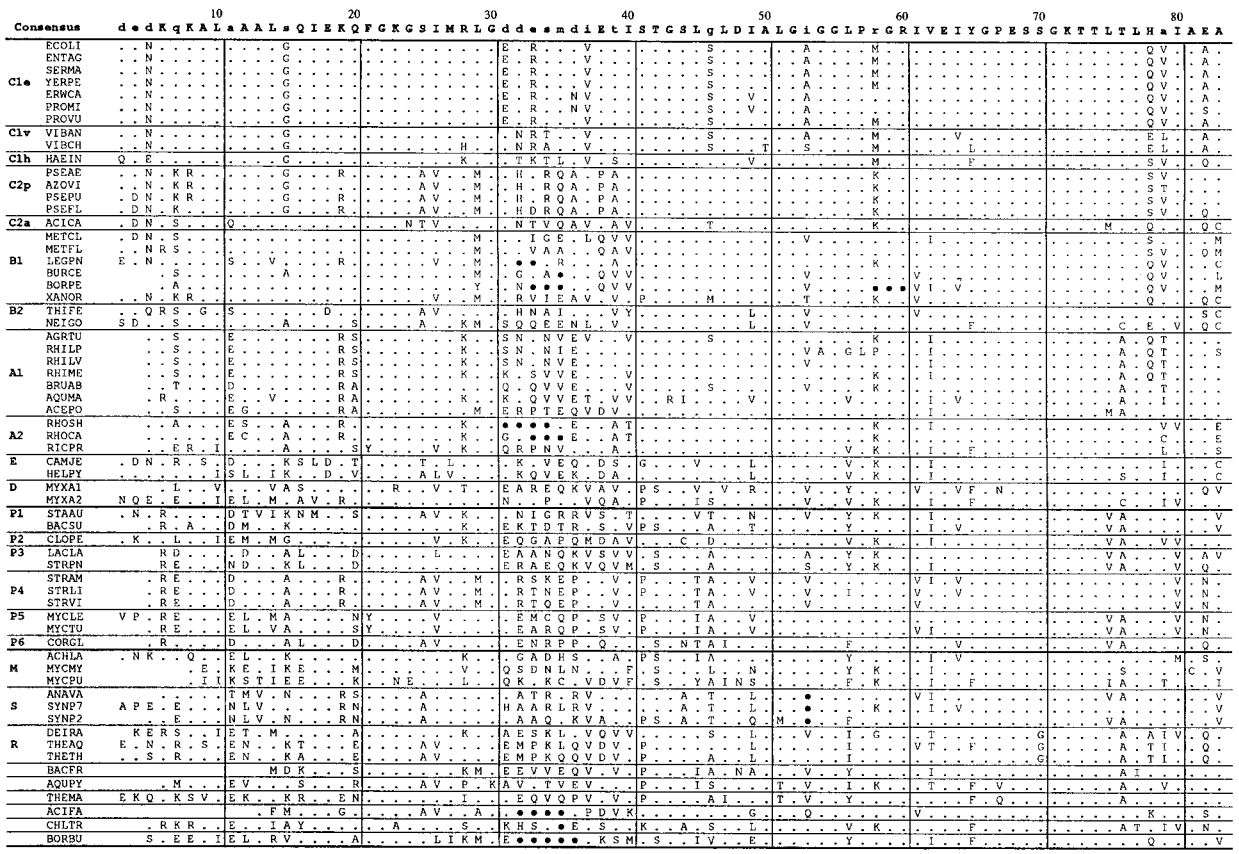
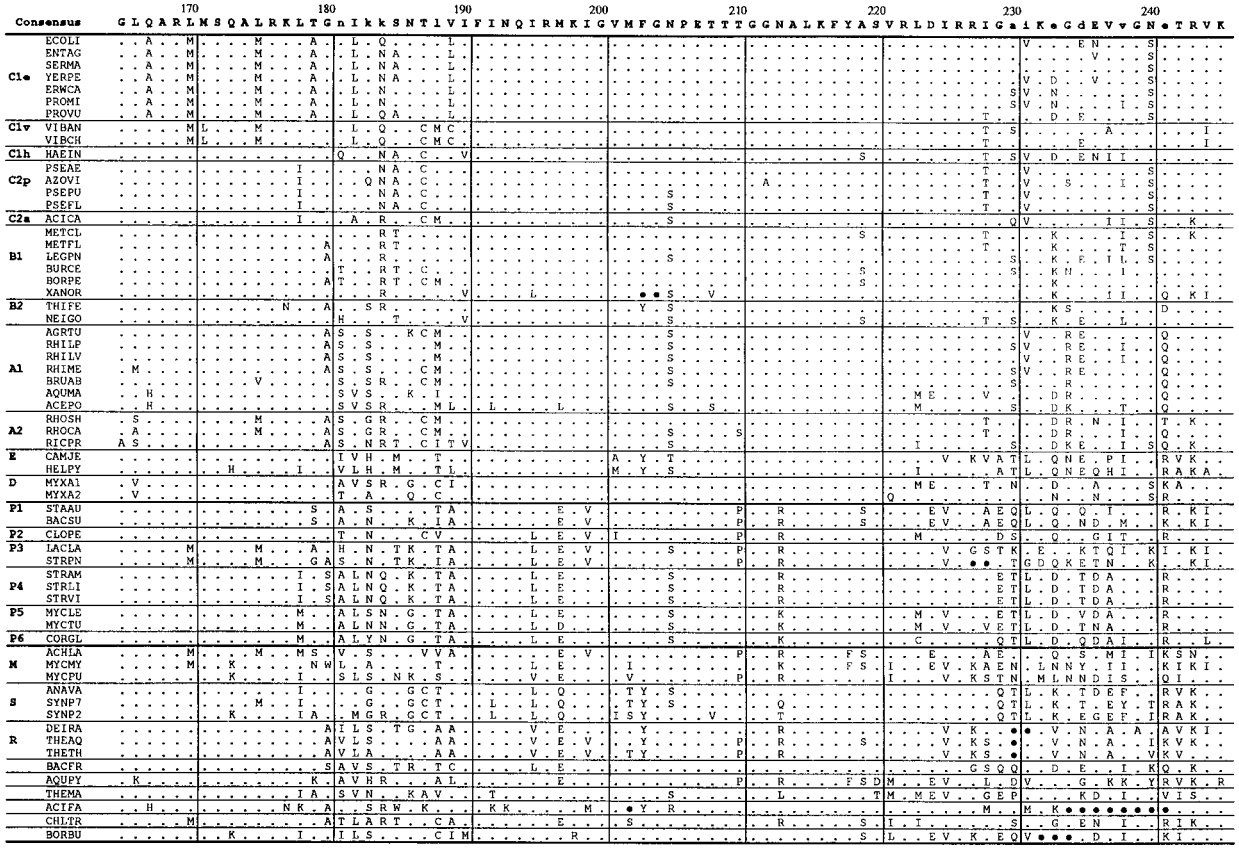
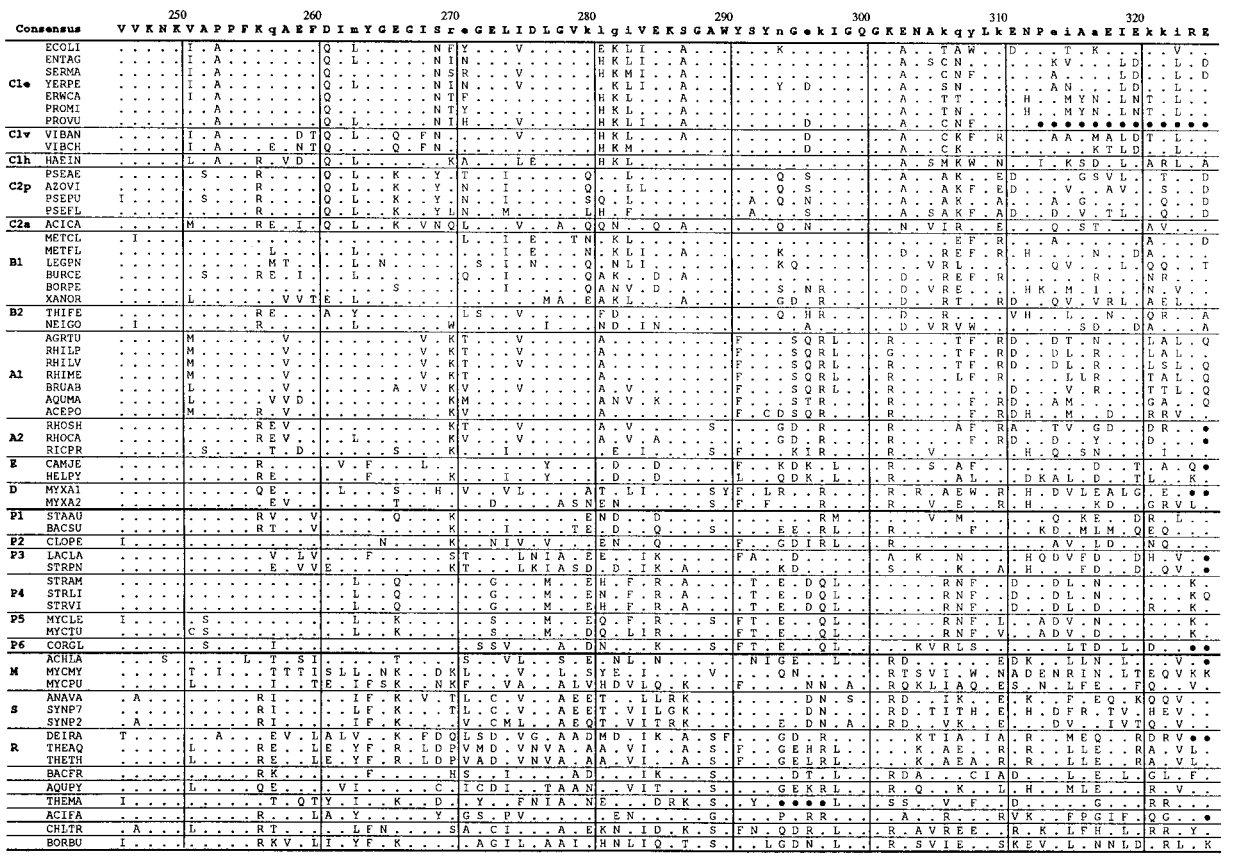


TABLE 1—*Continued*





Continued on following page

that for amino acid identity ($i = j$) the similarity score is 1 independent of i. At each position we calculate the conservation index (CI) as $[2/N(N-1)]\Sigma s^*(a,b)$, where the sum extends over all pairs of sequences for the given position in the alignment and N is the number of sequences. The CI value at a perfectly conserved position is 1.0. At a position where only similar amino acids appear (e.g., D or E or I, V, or L), a positive CI, generally >0.4, would be attained. A negative CI value reflects many nonsimilar amino acid substitutions or several drastic replacements at the given position. Table 2 reports CI values for three groups: all 63 sequences, the Proteo sequences, and the Gram(+) sequences. A position is declared strongly conserved if a residue is conserved in at least 60 of all 63 sequences, 31 of 33 Proteo sequences, and 10 of 11 Gram(+) sequences.

RESULTS

RecA sequence conservation in relation to *E. coli* RecA structure and function. It is commonly accepted that families of related (and generally similar) proteins generate essentially the same structure (5). On the basis of the known crystal structure and of biochemical and mutational studies of *E. coli* the key function and structure regions have been assigned as displayed in Fig. 1 and Table 2 (25, 26, 38, 39). However, on the basis of the evolutionary alignment, we consider some alternative assignments.

(i) Nucleotide binding (NB) domain, A site (NBA). The segment from positions 66 to 73, GPESSGKT (one-letter amino acid code used), is almost totally conserved relative to the aggregate of 63 sequences. The two exceptions are S-70 (serine at position 70), identical in 60 sequences but changing to G in the *Deinococcus-Thermus* group, and P-67, identical in 60 sequences but changing to N in MYXA1 (see Table 1, footnote *a*, for complete bacterial names), V in AQUPY, and Q in THEMA (Table 1; compare with the ATP-binding motif GPXGXGKSTL [Walker box] in many *E. coli* proteins [e.g., see references 3 and 42]).

The segment from positions 47 to 65, immediately amino terminal to the A site, mostly buried in the structure (Fig. 1), is largely conserved among the Proteo and the Gram(+) classes, including two invariant glycine positions and several conserved hydrophobic residues (Table 2). The segment from positions 74 to 88, immediately carboxyl to the ATP-binding A site, is similarly buried but less conserved. This hydrophobic environment presumably enhances structural stability at the A site. The invariant residue E at position 63 is mutually closest to the basic residue R-222 (invariant and involved in M-M interactions), with the oppositely charged side chain atoms of E-63 and R-222 forming a salt bridge which may add stability to the A-site region. Experimental mutations of R-222 produce only defective phenotypes (37), but since R-222 is also part of M-M interaction regions (Fig. 1), mutational analysis of E-63 could be helpful in elucidating the importance of its salt bridge interaction with R-222.

(ii) NB (hydrolysis) region, B site (NBB). The B motif of the Walker box comprises four successive hydrophobic residues followed by aspartate at positions 140 to 144 (e.g., see references 3 and 42). The key residue of the B site is D-144 (38, 39). The secondary structure of this segment is a β-strand that can anchor residue D-144 with a dense hydrophobic milieu. Intriguingly, the immediately carboxyl segment from positions 145 to 149 (SVAAL) is invariant across all RecA sequences.

(iii) Other NB positions (96, 100, 103, and 265). Residue E-96 is invariant over all sequences except for the single β-proteobacterium METCL, which substitutes the similar acidic residue D. Residue D-100 twice allows the replacement E. Residue Y-103 is maintained in 62 of 63 sequences and is substituted by the similar aromatic F in the sequence ACIFA. Residue 265 features G in 58 sequences, N in 4 sequences, and S in 1 sequence.

(iv) DNA binding domains. Despite the availability of a crystal structure (38, 39), the sites of RecA responsible for DNA binding remain unclear. The disordered loops L1 (residues 156 to 165) and L2 (194 to 210) are proposed (39) as DNA binding domains. The RecA sequences are strongly conserved only over residues 156 to 161 of L1, featuring three acidic residues and two glycine residues. The L2 stretch is substantially conserved. In particular, the boundary residue Q-194 is invariant, as are R-196, G-200, P-206, E-207, and T-209. Apart from these, the Proteo sequences maintain precisely the segment from residues 198 to 202 and the Gram(+) sequences conserve the segment from residues 202 to 209 except for position 205. Position 197 sharply contrasts in that the Proteo RecA sequences have only M, except for one hydrophobic L, whereas the Gram(+) sequences show only acidic residues. The L2 carboxyl position 210 is quite variable among the Gram(+) sequences but maintains a small hydroxyl residue among the Proteo sequences.

On the basis of UV irradiation promoting cross-linking of the RecA protein with specific DNA poly(dT), Rehrauer and

[a] The global alignment of 63 RecA sequences from 62 bacterial species was derived by combining high-scoring segment pairs (20). Position numbers refer to the *E. coli* sequence (ECOLI). The reported alignment traverses positions 3 to 325 of ECOLI, since other regions are poorly aligned. In the N terminus (positions 0 to 2) the strength of alignments relative to *E. coli* is inconclusive, involving only 12 to 31 sequences. Many of the RecA sequences appear to begin at position 2, where isoleucine (I) of *E. coli*, maintained in the enterobacteria, becomes the methionine (M) initiation amino acid in 17 other sequences. Residues 8 to 13 can be aligned for 61 of the 63 sequences, and all sequences are aligned traversing positions 14 to 312. The carboxyl 30 positions of the RecA sequences do not align well (i.e., generally do not participate in high-scoring segment pairs). "Consensus" refers to the most frequent amino acid at each position in the alignment. An uppercase letter designates an amino acid conserved in >50% of the 63 sequences. A lowercase designation indicates an amino acid conserved in a plurality of the 63 sequences but in fewer than 50%. In each position, dots signify agreement with the consensus amino acid. Dark circles (e.g., positions 32 and 33 of the LEGPN sequence) correspond to residues in the sequence not participating in a high-scoring segment pair with ECOLI. For example, in the pairing of ECOLI versus LEGPN, the high-scoring segment pairs associate positions 2 to 31 of ECOLI with 0 to 29 of LEGPN and 34 to 341 of ECOLI with 33 to 340 of LEGPN. The rarity of dark circles in the table attests to the generally strong RecA sequence conservation. The groupings of the species sequences were derived in reference 20 (concerning Proteo classifications, see also reference 11). Sequence names correspond to the following species. C1e: ECOLI, *E. coli*; ENTAG, *Enterobacter agglomerans*; SERMA, *Serratia marcescens*; YERPE, *Yersinia pestis*; ERWCA, *Erwinia carotovora*; PROMI, *Proteus mirabilis*; PROVU, *Proteus vulgaris*. C1v: VIBAN, *Vibrio anguillarum*; VIBCH, *Vibrio cholerae*. C1h: HAEIN, *Haemophilus influenzae*. C2p: PSEAE, *Pseudomonas aeruginosa*; AZOVI, *Azotobacter vinelandii*; PSEPU, *Pseudomonas putida*; PSEFL, *Pseudomonas fluorescens*. C2a: ACICA, *Acinetobacter calcoaceticus*. B1: METCL, *Methylomonas clara*; METFL, *Methylobacillus flagellatum*; LEGPN, *Legionella pneumophila*; BURCE, *Burkholderia cepacia* (*Pseudomonas cepacia*); BORPE, *Bordetella pertussis*; XANOR, *Xanthomonas oryzae*. B2: THIFE, *Thiobacillus ferrooxidans*; NEIGO, *Neisseria gonorrhoeae*. A1: AGRTU, *Agrobacterium tumefaciens*; RHILP, *Rhizobium leguminosarum* subsp. *phaseoli*; RHILV, *Rhizobium leguminosarum* subsp. *viciae*; RHIME, *Rhizobium meliloti*; BRUAB, *Brucella abortus*; AQUMA, *Aquaspirillum magnetotacticum*; ACEPO, *Acetobacter polyoxogenes*. A2: RHOSH, *Rhodobacter sphaeroides* (*Rhodopseudomonas sphaeroides*); RHOCA, *Rhodobacter capsulatus* (*Rhodopseudomonas capsulata*); RICPR, *Rickettsia prowazekii*. E: CAMJE, *Campylobacter jejuni*; HELPY, *Helicobacter pylori* (*Campylobacter pylori*). D: MYXA1, *Myxococcus xanthus*; MYXA2, *Myxococcus xanthus* (also see reference 32). P1: STAAU, *Staphylococcus aureus*; BACSU, *Bacillus subtilis*. P2: CLOPE, *Clostridium perfringens*. P3: LACLA, *Lactococcus lactis* subsp. *lactis* (*Streptococcus lactis*); STRPN, *Streptococcus pneumoniae*; P4: STRAM, *Streptomyces ambofaciens*; STRLI, *Streptomyces lividans*; STRVI, *Streptomyces vinaceus*. P5: MYCLE, *Mycobacterium leprae*; MYCTU, *Mycobacterium tuberculosis*. P6: CORGL, *Corynebacterium glutamicum*. M: ACHLA, *Acholeplasma laidlawii*; MYCMY, *Mycoplasma mycoides*; MYCPU, *Mycoplasma pulmonis*. S: ANAVA, *Anabaena variabilis*; SYNP7, *Synechococcus* sp. (strain pcc 7942); SYNP2, *Synechococcus* sp. (strain pcc 7002). R: DEIRA, *Deinococcus radiodurans*; THEAQ, *Thermus aquaticus*; THETH, *Thermus thermophilus*. Others: BACFR, *Bacteroides fragilis*; AQUPY, *Aquifex pyrophilus*; THEMA, *Thermotoga maritima*; ACIFA, *Acidiphilium facilis*; CHLTR, *Chlamydia trachomatis*; BORBU, *Borrelia burgdorferi.*

Kowalczykowski (33) propose that regions from residues 61 to 72, 178 to 183 (especially position 183), and 233 to 243, quite distinct from the L1 and L2 loops, are DNA binding domains. Appropriately, residues 183, 184, and 243 in the alignments are principally cationic. The same experimental protocol applied in reference 29 identified only positions 178 to 183 and the NB position 103 as sites of DNA binding. Surprisingly, the regions from positions 178 to 183 and 233 to 243 are greatly variable (Fig. 1 and Table 2). Another region proposed to interact with DNA is contained in the segment from positions 1 to 33 (16, 22, 48). This region, also involved in M-M interactions, includes five highly conserved cationic residues. Three of these (positions 8, 19, and 23), although exposed, do not contribute to intermonomer contacts.

(v) Regions of M-M interaction. On the basis of the RecA three-dimensional (3D) polymeric structure seven segments, M-M1 to M-M7, are assigned to M-M interaction regions (Fig. 1; Table 2). Segments M-M1, M-M6, and M-M7 form a surface on one monomer that contacts a corresponding surface on the facing monomer composed from segments M-M2 to M-M5. The bulk of residues found in M-M surfaces are conserved across all RecA sequences, possibly attesting to the importance of the precise polymerization geometry, with the variable positions mostly confined to the boundaries of the M-M segments.

(vi) M-M1, M-M6, and M-M7. The M-M1 region involves five conserved basic residues (K or R) and one acidic residue (60 sequences with E and 3 with D). Interspersed are nine aliphatic residues and one aromatic residue (represented by 60 sequences with F and 3 with Y). Three conserved glycine residues may provide flexibility for within and between monomer assembly. The M-M6 sequence, highly conserved, features an invariant K (position 216) and an invariant R (position 222). There are no acidic residues in this region. Notably, the boundary residue 213 represents N in all Proteo sequences but switches to a cationic residue (eight with R and three with K) in all Gram(+) sequences. The M-M7 segment is substantially conserved, highlighting two invariant K residues and an additional basic position involving both K and R. There are no acidic residues.

(vii) M-M2 to M-M5. The M-M2 region features three highly conserved acidic and no basic residues. Also, three alanine positions which can provide hinge capabilities to the M-M2 interface for M-M engagements are present. The composition of M-M3 emphasizes a central hydrophobic core and an acidic charge at both its boundaries. The four residues 110 to 113 of M-M3 are the most variable of all M-M surfaces. The aggregate net charge of M-M4 is again negative, with three acidic residues and one arginine residue. Except for position 150, the M-M5 interface is strongly conserved. It contains two acidic residues and one basic residue, again providing a net negative charge. There are several small hydrophobic residues which can enhance stability for M-M interactions.

(viii) Intermonomer contacts. In the RecA structure of *E. coli*, a single intermonomer salt bridge is formed from K-6 (of M-M1) and D-139 (of M-M6). Both residues are conserved in the alignment. However, a mutational study (29) showed that the interaction is not essential for protein function. Certain conserved charged residues establish intermonomer hydrogen bonds with polar residues (R-222 with H-97; E-156 with Y-218), whereas other interface charges (K at positions 8, 19, 23, 152, and 256; D at positions 94, 100, and 130) show no intermonomer contacts (two residues are considered to be structural contact neighbors if the minimum distance between their side chain atoms is $\leq$4.5 Å [$\leq$0.45 nm]). Nevertheless, RecA polymerization appears to be facilitated by electrostatic interactions that help to orient and localize monomers (see Discussion).

There are 14 conserved major and minor hydrophobic positions in the M-M1–M-M6–M-M7 interface and 15 in the M-M2–to–M-M5 interface. Ten of the residues from the first face (positions 10, 17, 21, 26, 27, 29, 214, 217, and 255) establish hydrophobic contacts with nine conserved residues of the opposite surface (positions 98, 99, 114, 116, 128, 132, 138, 148, and 155). The remaining hydrophobic positions are mostly buried, which can help to stabilize the interface surfaces. Are there distinctive residues in interchain interactions? Indeed, F-255 in M-M7 and Q-118 in M-M3 have many intermonomer contacts. Both these residues are attractive candidates for mutational studies.

(ix) Filament-filament contacts. Positions 37 to 39 and 298 to 301 are identified as possible filament-filament contact positions on the basis of mutational analyses (43). Whereas positions 37 and especially 38 are highly variable, position 39 is an invariant hydrophobic buried residue. Positions 299 to 301 are conserved in all sequences except for replacements for G at 299 (Table 1). Position 298, mostly buried, is hydrophobic in all sequences. A 3D contact residue (having closest side chain atoms within 4.5 Å) of Q-300 is W-290, exposed and totally conserved among Proteo and Gram(+) sequences. The two residues appear to establish a hydrogen bond between their side chains, bonding the oxygen of Q and the imino nitrogen of W, separated by about 3.3 Å (0.33 nm). Moreover, W-290 is also a 3D contact neighbor of G-299. The only 3D contact neighbor of G-301 (invariant) is Y-291 or F-291. These tertiary relations and conservation suggest that the invariant aromatic residues W-290 and Y-291 or F-291 play an important structural or functional role with respect to the filament-filament contacts.

An extended set of filament-filament contact sites can be deduced from the analysis of the *E. coli* filament structure. The interfilament contacts within 4.5 Å include the 13 residues at positions 12, 15, 16, 18, 19, 23, 33, 35 to 38, 60, and 183 with the 12 residues at positions 290, 294, 296 to 299, 308, 311 to 312, 314, 315, and 318 (cf. with results reported in reference 27). The average conservation of these positions is relatively low (CI = 0.23), including some of the most variable positions of the RecA alignment. Virtually all contact pairings entail at least one highly variable position. The least variable interfilament contacts involve residue N or H at position 312 on one surface with residues E-18 and K-23 on the opposite surface. The substantial variability in the alignments of the filament-filament contacts makes their biological significance problematical.

(x) Target protein binding (TPB) sites 229 and 243. TPB sites 229 and 243 are hypothesized as active sites for LexA binding (e.g., see references 25 and 39). G-229 is invariant across Proteo sequences but quite variable with respect to non-Proteo sequences (Tables 1 and 2). This position is in close contact with the conserved basic residue at position 227. Position 243 is a strongly conserved basic amino acid with K and R about equally represented. Its only conserved contact neighbor is again a cationic residue, K-245, which is in turn closest to R-226, totally conserved. The extra cationic residues in the structural environment of both sites suggest that a positive-charge preponderance is important for TPB function. A more intriguing hypothesis is that this region is a site for DNA binding (see Discussion). The L1 region has also been proposed for LexA contacts (46).

Conservation assessments relative to function and structure domains. The average CI (see Materials and Methods) over distinct regions of the RecA protein structure is summarized in

TABLE 2. Conserved residues in RecA sequences[a]

RESIDUE	GLOBAL CONSERVATION Residue type	CI	PROTEO CONSERVATION Residue type	CI	GRAM(+) CONSERVATION Residue type	CI	FUNCT.	S.A.* S.S.
0 m		0.03		0.05		0.00		
1 a		0.02		0.05		0.00		
2 m		0.04		0.15		−0.04		
3 d		0.09		0.24		−0.02		81.8
4 e		−0.01		0.25		−0.05		51.7
5 d		0.06		0.24		0.66		73.3 H
6 K		0.54	**Basic**	0.83	*Arg*	0.82	M-M1	94.5 H
7 q		−0.38		−0.35		0.03	M-M1	79.4 H
8 K		0.57	**Basic**	0.74	*Lys*	0.62	M-M1	83.6 H
9 A		0.66		0.91	Ala	1.00	M-M1	77.7 H
10 L	*Leu-Ile*	0.63	*Leu*	0.92	*Leu*	0.75	M-M1	58.6 H
11 a		−0.24		0.05		0.09	M-M1	68.2 H
12 A		−0.14	*Ala-Cys*	0.74		−0.63	M-M1	72.9 H
13 A	*Ala-Val*	0.76	**Ala**	1.00	*Ala*	0.74	M-M1	65.8 H
14 L	*Major hydro*	0.44	*Leu*	0.85	**Major hydro**	0.30	M-M1	39.5 H
15 s		−0.38	**Small**	0.00		−0.08	M-M1	74.0 H
16 Q		0.26	**Gln**	1.00		0.00	M-M1	64.2 H
17 I	*Ile*	0.90	**Ile**	1.00	*Ile*	0.80	M-M1	39.9 H
18 E	*Glu*	0.91	*Glu*	0.94	**Glu**	1.00	M-M1	23.2 H
19 K	*Basic*	0.47	**Basic**	0.51	**Basic**	0.56	M-M1	77.7 H
20 Q		−0.22		0.07		−0.29	M-M1	66.7 H
21 F	*Phe*	0.92	*Phe*	0.95	**Phe-Tyr**	0.71	M-M1	65.6
22 G	**Gly**	1.00	**Gly**	1.00	**Gly**	1.00	M-M1	93.9
23 K	*Lys*	0.86	**Lys**	1.00	**Lys**	1.00	M-M1	44.0
24 G	*Gly*	0.88	*Gly*	0.88	**Gly**	1.00	M-M1	12.7
25 S	*Ser-Ala*	0.40	*Ser-Ala*	0.55	**Ala-Ser**	0.30	M-M1	16.3
26 I	*Ile-Val*	0.56	**Ile-Val**	0.70	**Aliphatic**	0.55	M-M1	47.2
27 M	*Met*	0.89	**Met**	1.00	**Met**	1.00	M-M1	12.7
28 R		0.23	**Basic**	0.50	**Basic**	0.56	M-M1	75.4
29 L	*Leu-Met*	0.50	*Leu-Met*	0.45	**Leu-Met**	0.56	M-M1	62.2
30 G	*Gly*	0.94	**Gly**	1.00	**Gly**	1.00	M-M1	96.3
31 d		−0.04		−0.10	**Acidic**	0.49		33.8
32 d		−0.45		−0.16		−0.43		115.9
33 e		−0.59		−0.48		−0.42		76.3
34 s		−0.56		−0.42		−0.64		24.8
35 m		−0.42		−0.31		−0.46		32.1
36 d		−0.54		−0.07		−0.34		122.7
37 i	*Aliphatic*	0.43	*Aliphatic*	0.45	*Ile-Val*	0.51	F-F1	16.3
38 E		−0.13		0.20		−0.40	F-F1	77.7
39 t		0.02		0.05		0.20		13.2 E
40 I		0.39		0.47	*Major hydro*	0.31		0.6 E
41 S		0.19	*Ser*	0.88		−0.05		42.2
42 T	**Small hydrxl**	0.69	**Thr**	1.00	**Small hydrxl**	0.42		0.0
43 G	*Gly*	0.95	*Gly*	0.91	**Gly**	1.00		41.9
44 S	*Ser-Ala*	0.71	*Ser*	0.88		0.55		0.0
45 L		0.37	**Leu**	1.00		−0.21		0.0 H
46 g		−0.22	*Gly-Ser*	0.03		0.43		24.9 H
47 L	*Leu-Ile*	0.83	**Leu**	1.00	*Leu*	0.75		0.0 H
48 D	*Asp*	0.92	**Asp**	1.00	**Asp**	1.00		0.1 H
49 I		−0.00	**Aliphatic**	0.61		0.12		18.2 H
50 A	*Ala*	0.96	*Ala*	0.92	**Ala**	1.00		5.9 H
51 L	*Leu*	0.85	**Leu**	1.00	**Leu**	1.00		0.3
52 G	**Gly**	1.00	**Gly**	1.00	**Gly**	1.00		28.4
53 i		0.11	*Aliphatic*	0.03	*Aliphatic*	0.09		10.2
54 G	*Gly*	0.96	*Gly*	0.91	**Gly**	1.00		10.5
55 G	**Gly**	1.00	**Gly**	1.00	**Gly**	1.00		0.0
56 L		−0.01	*Leu-Val*	0.71		−0.28		0.1 E
57 P	*Pro*	0.92	*Pro*	0.85	**Pro**	1.00		0.2 E
58 r		0.14		−0.03	**Basic**	0.49		2.2
59 G	*Gly*	0.97	*Gly*	0.94	**Gly**	1.00		3.5
60 R	*Arg*	0.97	*Arg*	0.94	**Arg**	1.00		21.4
61 I	**Ile-Val**	0.77	**Ile-Val**	0.84	**Ile-Val**	0.63		1.5 E
62 V	*Ile-Val*	0.54	**Ile-Val**	0.68	**Ile-Val**	0.60		0.0 E
63 E	**Glu**	1.00	**Glu**	1.00	**Glu**	1.00		4.3 E
64 I	**Ile-Val**	0.80	**Ile-Val**	0.88	**Ile-Val**	0.63		0.0 E
65 Y	*Phe-Tyr*	0.68	*Phe-Tyr*	0.74	**Tyr**	1.00		27.1 E
66 G	**Gly**	1.00	**Gly**	1.00	**Gly**	1.00	NBA	11.5
67 P	*Pro*	0.79	**Pro**	1.00	**Pro**	1.00	NBA	58.4
68 E	**Glu**	1.00	**Glu**	1.00	**Glu**	1.00	NBA	91.5
69 S	**Ser**	1.00	**Ser**	1.00	**Ser**	1.00	NBA	91.6
70 S	*Ser-Gly*	0.78	*Ser*	0.88	**Ser**	1.00	NBA	1.7
71 G	**Gly**	1.00	**Gly**	1.00	**Gly**	1.00	NBA	32.5
72 K	**Lys**	1.00	**Lys**	1.00	**Lys**	1.00	NBA	12.4 H
73 T	**Thr**	1.00	**Thr**	1.00	**Thr**	1.00	NBA	78.0 H
74 T	**Thr**	1.00	**Thr**	1.00	**Thr**	1.00		6.2 H
75 L	*Leu-Val*	0.57	*Leu*	0.88	**Val-Leu**	0.44		0.0 H
76 T		0.26		0.47		0.44		3.7 H
77 L	*Leu*	0.89	**Leu**	1.00	**Leu**	1.00		3.3 H
78 H		−0.05		−0.17	**His**	1.00		17.9 H
79 a		−0.08		−0.06	*Ala*	0.74		0.0 H
80 I	*Ile-Val*	0.72	*Ile*	0.91	**Val-Ile**	0.68		0.2 H
81 A	*Ala*	0.95	**Ala**	1.00	**Ala**	1.00		1.8 H
82 E		−0.12		−0.10		−0.23		23.5 H
83 A		−0.01		−0.07	**Ala-Val**	0.38		0.0 H
84 Q	**Gln**	1.00	**Gln**	1.00	**Gln**	1.00		9.7 H
85 K		0.34	*Basic*	0.46		−0.11		70.9 H
86 a		−0.61		−0.58		−0.48		58.4
87 G	*Gly*	0.91	**Gly**	1.00	*Gly*	0.82		06.0
88 G		0.26		−0.06	*Gly*	0.72		25.7
89 t		0.00		0.31		−0.45	M-M2	47.1
90 c	*Hydrophobic*	−0.03	*Cys-Ala*	0.47	*Ala-Val*	0.17	M-M2	0.0 E
91 A	*Ala*	0.87	*Ala*	0.91	**Ala**	1.00	M-M2	0.8 E
92 F	*Phe*	0.92	**Phe**	1.00	*Phe*	0.84	M-M2	4.3 E
93 I	*Ile-Val*	0.63	**Ile-Val**	0.63	**Ile-Val**	0.68	M-M2	1.0 E
94 D	**Asp**	1.00	**Asp**	1.00	**Asp**	1.00	M-M2	3.1 E
95 A	*Ala*	0.91	*Ala*	0.92	**Ala**	1.00	M-M2	16.3
96 E	*Glu*	0.97	*Glu*	0.94	**Glu**	1.00	M-M2, NB	40.0
97 H	*His*	0.88	*His*	0.90	*His*	0.80	M-M2	66.5
98 A	*Ala*	0.92	**Ala**	1.00	**Ala**	1.00	M-M2	72.8
99 L	*Leu*	0.86	*Leu*	0.91	**Leu**	1.00	M-M2	26.3
100 D	*Asp*	0.94	**Asp**	1.00	*Asp*	0.82	M-M2, NB	65.2
101 P		0.52		0.72	**Pro**	1.00	M-M2	46.5 H
102 v		−0.71		−0.60		−0.16	M-M2	83.7 H
103 Y	*Tyr*	0.97	**Tyr**	1.00	**Tyr**	1.00	NB	21.6 H
104 A	*Ala*	0.91	**Ala**	1.00	**Ala**	1.00		0.0 H
105 k		−0.25		−0.30		0.04		54.0 H
106 K		0.21	**Lys**	1.00		−0.04		56.4 H
107 L	*Leu*	0.92	**Leu**	1.00	**Leu**	1.00		3.7
108 G	**Gly**	1.00	**Gly**	1.00	**Gly**	1.00		48.3
109 V	*Aliphatic*	0.88	**Val**	0.94	**Val**	1.00		0.1
110 D		0.15		0.12		0.36		60.4
111 i		0.11	**Aliphatic**	0.23		0.39	M-M3	18.2
112 D		0.01		−0.07	*Asp*	0.82	M-M3	84.0
113 n		−0.11		−0.01		−0.33	M-M3	83.9
114 L	**Leu**	1.00	**Leu**	1.00	**Leu**	1.00	M-M3	3.4
115 L		0.39	*Leu-Ile*	0.72	*Aliphatic*	0.02	M-M3	62.3 E
116 i	**Hydrophobic**	0.05	**Hydrophobic**	0.01	**Leu-Val**	0.35	M-M3	39.8 E
117 S	*Ser*	0.91	**Ser**	1.00	**Ser**	1.00	M-M3	31.4 E
118 Q	*Gln*	0.96	*Gln*	0.93	**Gln**	1.00	M-M3	58.8
119 P	**Pro**	1.00	**Pro**	1.00	**Pro**	1.00	M-M3	3.6
120 D	*Asp*	0.91	*Asp*	0.83	**Asp**	1.00	M-M3	86.3
121 T		0.28		0.69		−0.05		19.1
122 G	**Gly**	1.00	**Gly**	1.00	**Gly**	1.00		0.3 H
123 E	**Glu**	1.00	**Glu**	1.00	**Glu**	1.00		42.2 H
124 Q		0.77	**Gln**	1.00	**Gln**	1.00		44.3 H
125 A	*Ala*	0.87	**Ala**	1.00	*Ala-Gly*	0.38		0.9 H
126 L	**Leu**	1.00	**Leu**	1.00	**Leu**	1.00		4.4 H
127 E	*Acidic*	0.83	**Glu**	1.00	*Glu*	0.80	M-M4	84.9 H
128 I	**Ile**	1.00	**Ile**	1.00	**Ile**	1.00	M-M4	35.3 H
129 a		−0.21		−0.14		0.21	M-M4	0.5 H
130 D	*Acidic*	0.56	**Asp**	1.00	*Acidic*	0.27	M-M4	11.5 H
131 m		−0.33		−0.05		−0.18	M-M4	52.0 H
132 L	*Leu*	0.87	**Leu**	1.00	*Leu*	0.73	M-M4	26.7 H
133 V		0.19		0.26	**Val-Ile**	0.63	M-M4	0.0 H
134 R		0.64	**Arg**	1.00		0.29	M-M4	52.8 H
135 S	*Ser*	0.94	*Ser*	0.88	**Ser**	1.00	M-M4	53.0 H
136 G		0.47		0.42	**Gly**	1.00	M-M4	55.9
137 A	*Small*	0.45	*Small*	0.40	**Ala**	1.00	M-M4	84.9
138 V	**Aliphatic**	0.53	**Ile-Val**	0.70	**Aliphatic**	0.17	M-M4	4.6
139 D	*Asp*	0.90	*Asp*	0.82	**Asp**	1.00	M-M4	35.6
140 v	**Major hydro**	0.23	**Major hydro**	0.38	*Leu-Ile*	0.18	NBB	0.0 E
141 I	**Aliphatic**	0.43	**Aliphatic**	0.36	*Val-Ile*	0.44	NBB	0.0 E
142 V	**Ile-Val**	0.80	**Ile-Val**	0.68	**Val**	1.00	NBB	0.0 E
143 V	*Ile-Val*	0.60	**Ile-Val**	0.70	**Ile-Val**	0.63	NBB	0.7 E
144 D	**Asp**	1.00	**Asp**	1.00	**Asp**	1.00	NBB	22.7 E
145 S	**Ser**	1.00	**Ser**	1.00	**Ser**	1.00		9.7
146 V	**Val**	1.00	**Val**	1.00	**Val**	1.00		0.1
147 A	**Ala**	1.00	**Ala**	1.00	**Ala**	1.00		36.4
148 A	**Ala**	1.00	**Ala**	1.00	**Ala**	1.00	M-M5	29.4
149 L	**Leu**	1.00	**Leu**	1.00	**Leu**	1.00	M-M5	2.2
150 V		0.28		0.35		0.53	M-M5	38.3
151 P	**Pro**	1.00	**Pro**	1.00	**Pro**	1.00	M-M5	5.6
152 K	*Basic*	0.45	**Basic**	0.62	**Basic**	0.49	M-M5	64.8 H
153 A	*Small*	0.67	*Small*	0.77	**Ala**	1.00	M-M5	90.0 H
154 E	**Glu**	1.00	**Glu**	1.00	**Glu**	1.00	M-M5	58.3 H
155 I	**Ile-Leu**	0.65	*Ile*	0.84	**Ile-Leu**	0.54	M-M5	41.2 H
156 E	*Acidic*	0.74	**Glu**	1.00	**Acidic**	0.67	M-M5, L1	72.7
157 G	**Gly**	1.00	**Gly**	1.00	**Gly**	1.00	L1	
158 E	*Acidic*	0.62	**Acidic**	0.74	**Acidic**	0.67	L1	
159 M	**Met-Ile**	0.70	**Met-Ile**	0.62	**Met-Ile**	0.64	L1	
160 G	*Gly*	0.83	**Gly**	1.00	**Gly**	1.00	L1	
161 D	*Asp*	0.91	**Asp**	1.00	**Asp**	1.00	L1	
162 S		−0.10		0.15	*Ser*	0.79	L1	
163 H		0.14		0.21		0.46	L1	
164 V		−0.09		−0.30	**Val**	1.00	L1	
165 G	*Gly*	0.96	*Gly*	0.91	**Gly**	1.00	L1	121.3
166 L		0.60		0.54	**Leu**	1.00		58.8 H
167 Q		0.49		0.18	**Gln**	1.00		38.5 H
168 A	**Ala**	1.00	**Ala**	1.00	**Ala**	1.00		41.9 H
169 R	**Arg**	1.00	**Arg**	1.00	**Lys**	1.00		65.7 H
170 L	**Leu-Met**	0.65	**Leu-Met**	0.59	**Leu-Met**	0.67		2.4 H
171 M	*Met*	0.94	*Met*	0.88	**Met**	1.00		0.4 H
172 S	**Ser**	1.00	**Ser**	1.00	**Ser**	1.00		58.1 H
173 Q		0.81	**Gln**	1.00	**Gln**	1.00		59.9 H
174 A	**Ala**	1.00	**Ala**	1.00	**Ala**	1.00		32.1 H
175 L	*Leu-Met*	0.60	*Leu-Met*	0.48	**Leu-Met**	0.67		0.0 H
176 R	**Arg**	1.00	**Arg**	1.00	**Arg**	1.00		88.4 H
177 K	*Lys*	0.91	**Lys**	0.92	**Lys**	1.00		70.2 H
178 L	*Major hydro*	0.34	**Leu-Ile**	0.63	**Major hydro**	0.16		0.5 H
179 T		0.43		0.56		0.12		29.7 H
180 G	*Small*	0.23	*Gly-Ala*	0.35	*Gly-Ser*	0.09		90.6 H
181 n		−0.22		0.28		0.28		21.9 H
182 I	*Aliphatic*	0.21	*Aliphatic*	0.28	**Leu-Ile**	0.24		0.2 H
183 k		−0.34		−0.04		0.36		62.3 H
184 k		0.05		−0.06		0.04		86.2 H
185 S		0.22		0.24	**Small hydrxl**	0.63		17.6 H
186 N		0.18	*Asn*	0.84		−0.15		49.0
187 T		0.15		0.12	*Thr*	0.69		0.1
188 l		−0.30	*Leu-Met*	0.42		0.36		0.0 E
189 V	*Aliphatic*	−0.03	*Aliphatic*	0.24	*Ala*	0.74		0.0 E
190 I	*Ile-Val*	0.88	**Ile-Val**	0.84	**Ile**	1.00		0.0 E
191 F	**Phe**	1.00	**Phe**	1.00	**Phe**	1.00		0.0 E
192 I	*Leu-Ile*	0.76	*Ile*	0.92	**Ile**	1.00		5.1 E
193 N	*Asn*	0.96	**Asn**	1.00	**Asn**	1.00		2.1 E
194 Q	**Gln**	1.00	**Gln**	1.00	**Gln**	1.00	L2	59.8
195 I	**Aliphatic**	0.42	*Ile*	0.92	**Leu-Ile**	0.54	L2	
196 R	**Arg**	1.00	**Arg**	1.00	**Arg**	1.00	L2	
197 M		−0.03	*Met*	0.94	*Glu*	0.82	L2	
198 K	*Lys*	0.97	**Lys**	1.00	**Lys**	1.00	L2	
199 I	*Ile-Val*	0.80	**Ile**	1.00	**Ile-Val**	0.60	L2	

TABLE 2—*Continued*

RESIDUE	GLOBAL CONSERVATION Residue type	CI	PROTEO CONSERVATION Residue type	CI	GRAM(+) CONSERVATION Residue type	CI	FUNCT.	S.A. S.S.
200 G	Gly	1.00	Gly	1.00	Gly	1.00	L2	
201 V	*Val-Ile*	0.87	Val	1.00	*Val*	0.87	L2	
202 M		0.63	Met	1.00	Met	1.00	L2	
203 F	*Phe-Tyr*	0.73	*Phe*	0.89	Phe	1.00	L2	
204 G	*Gly*	0.97	*Gly*	0.94	Gly	1.00	L2	
205 N		0.38		0.42		0.42	L2	
206 P	Pro	1.00	Pro	1.00	Pro	1.00	L2	
207 E	Glu	1.00	Glu	1.00	Glu	1.00	L2	
208 T	*Thr*	0.88	*Thr*	0.85	Thr	1.00	L2	
209 T	Thr	1.00	Thr	1.00	Thr	1.00	L2	
210 T		0.45	*Thr*	0.93		−0.05	L2	80.3
211 G	Gly	1.00	Gly	1.00	Gly	1.00		0.0
212 G	*Gly*	0.97	*Gly*	0.91	Gly	1.00		71.9
213 N		0.11	Asn	1.00	Basic	0.56	M-M6	86.5 H
214 A	Ala	1.00	Ala	1.00	Ala	1.00	M-M6	10.2 H
215 L	Leu	1.00	Leu	1.00	Leu	1.00	M-M6	1.1 H
216 K	Lys	1.00	Lys	1.00	Lys	1.00	M-M6	49.9 H
217 F	Phe	1.00	Phe	1.00	Phe	1.00	M-M6	70.0 H
218 Y	*Tyr*	0.92	Tyr	1.00	Tyr	1.00	M-M6	39.3 H
219 A	Ala-Ser	0.57	Ala-Ser	0.66	Ala-Ser	0.58	M-M6	0.0
220 S	*Ser*	0.91	Ser	1.00	Ser	1.00	M-M6	10.5
221 V	*Aliphatic*	0.75	Val	1.00	Val	1.00	M-M6	0.0 E
222 R	Arg	1.00	Arg	1.00	Arg	1.00	M-M6	20.3 E
223 L	*Major hydro*	0.62	Major hydro	0.80	*Leu-Met*	0.24		0.1 E
224 D	Acidic	0.75	*Asp*	0.94	Acidic	0.67		10.8 E
225 I	Ile-Val	0.73	Ile	1.00	Ile-Val	0.60		0.0 E
226 R	Arg	1.00	Arg	1.00	Arg	1.00		52.8 E
227 R	*Basic*	0.68	Arg	1.00		0.57		38.0 E
228 I		−0.02		0.28		−0.05		24.7 E
229 G		0.03	Gly	1.00		0.02	TPB	61.9 E
230 a		−0.13	*Ala-Ser*	0.29		0.00		27.1 E
231 i	*Aliphatic*	0.24	Ile-Val	0.63	*Leu-Ile*	0.16		17.0 E
232 K		0.72	Lys	1.00		0.35		38.7 E
233 e		−0.29		−0.07		0.10		77.0 E
234 G		0.16		0.05	*Gly*	0.65		130.5
235 d		−0.17	Acidic	0.54		−0.51		87.8
236 E		0.08		0.44		−0.22		100.8 E
237 V		−0.03	*Val-Ile*	0.75		−0.29		42.7 E
238 v		0.22		0.40	*Major hydro*	0.34		40.7 E
239 G	*Gly*	0.92	Gly	1.00	Gly	1.00		0.0 E
240 N		0.13		0.42		0.55		16.1 E
241 e		−0.23		0.36		0.06		32.9 E
242 T		0.41	Thr	1.00	Thr	1.00		0.0 E
243 R	*Basic*	0.47	Basic	0.74	Basic	0.49	TPB	29.6 E
244 V	*Val-Ile*	0.74	Val-Ile	0.88	*Val-Ile*	0.44		0.0 E
245 K	*Lys*	0.97	Lys	1.00	Lys	1.00		25.8 E
246 V	*Val-Ile*	0.85	*Val*	0.96	Val-Ile	0.76		0.0 E
247 V	*Val-Ile*	0.82	*Val*	0.91	Val	1.00	M-M7	27.7 E
248 K	Lys	1.00	Lys	1.00	Lys	1.00	M-M7	26.7 E
249 N	*Asn*	0.96	Asn	1.00	Asn	1.00	M-M7	0.0 E
250 K	Lys	1.00	Lys	1.00	Lys	1.00	M-M7	45.3
251 V	*Major hydro*	0.30	Major hydro	0.23	*Val*	0.62	M-M7	0.2
252 A	Ala-Ser	0.74	Ala-Ser	0.72	Ala-Ser	0.44	M-M7	6.1
253 P		0.39		0.23	Pro	1.00	M-M7	0.5
254 P	Pro	1.00	Pro	1.00	Pro	1.00	M-M7	16.1
255 F	*Phe*	0.95	Phe	1.00	Phe	1.00	M-M7	67.3
256 K	*Basic*	0.46	Basic	0.52	Basic	0.67	M-M7	49.6
257 q		−0.20		0.27		−0.34	M-M7	54.9 E
258 A	*Ala-Val*	0.44	*Ala-Val*	0.29	Ala	1.00		4.1 E
259 E		0.34		0.29		0.23		49.6 E
260 F		0.23		0.59	Major hydro	0.17		4.2 E
261 D		0.02		0.08	*Asp*	0.82		26.7 E
262 I	Aliphatic	0.83	Ile	1.00	Ile	1.00		8.4 E
263 m		0.17	*Leu-Met*	0.40	Leu-Met	0.46		22.9 E
264 Y	Tyr-Phe	0.72	Tyr	1.00	*Tyr*	0.84		61.0
265 G		0.72	*Gly*	0.88	*Gly*	0.65	NB	57.3
266 E		−0.09		0.14		0.09		52.3
267 G	Gly	1.00	Gly	1.00	Gly	1.00		1.0 E
268 I	*Aliphatic*	0.61	*Ile-Val*	0.62	Ile	1.00		0.1 E
269 S		0.15		0.18	Ser	1.00		14.0
270 r		−0.20		−0.30	*Basic*	0.22		27.2 H
271 e		−0.67		−0.68		0.32		27.2 H
272 G	*Small*	0.47	*Gly*	0.80	*Gly*	0.69		2.1 H
273 E		0.23	Glu	1.00		−0.46		8.2 H
274 L	*Aliphatic*	0.26	*Leu-Ile*	0.39	*Leu-Ile*	0.38		0.0 H
275 I	*Aliphatic*	0.37	*Ile-Val*	0.58	*Ile-Leu*	0.44		0.0 H
276 D		0.62	*Acidic*	0.76		0.42		62.5 H
277 L	*Hydrophobic*	0.36	*Leu*	0.86	Major hydro	0.15		14.2 H
278 G	*Gly-Ala*	0.33	*Gly*	0.83	Gly-Ala	0.38		0.0 H
279 V		0.42	*Val*	0.91		0.40		13.9 H
280 k		−0.29		0.19	*Acidic*	0.28		71.6 H
281 l		−0.55		−0.44		−0.46		18.6
282 g		−0.27		−0.24		0.10		57.3
283 l	*Major hydro*	0.06	Major hydro	0.13	Major hydro	0.10		22.8
284 V	*Val-Ile*	0.56	*Val-Ile*	0.63	Val-Ile	0.68		0.0
285 E		−0.14		0.52		−0.01		74.9 E
286 K	*Basic*	0.85	Lys	1.00	Lys	1.00		42.1 E
287 S		0.14	Ser-Ala	0.39	Ser-Ala	0.35		85.2 E
288 G	Gly	1.00	Gly	1.00	Gly	1.00		129.0
289 A	*Ala-Ser*	0.45	*Ala*	0.85	Ala-Ser	0.44		64.0
290 W	*Trp*	0.90	Trp	1.00	Trp	1.00		41.8 E
291 Y	*Tyr-Phe*	0.56	Tyr-Phe	0.66	Phe-Tyr	0.51		7.4 E
292 S		0.58	Ser-Ala	0.85	*Small hydrxl*	0.27		31.0 E
293 Y		0.75	*Tyr*	0.90	Tyr	1.00		6.0 E
294 n		−0.33		−0.04		−0.09		73.9
295 G		−0.06		0.02		0.13		111.2
296 e		−0.32		−0.31	*Acidic*	0.13		94.8 E
297 k		0.24	Basic	0.51		0.21		61.2 E
298 I	*Ile-Leu*	0.32	Ile-Leu	0.63	*Leu-Ile*	0.41	F-F2	2.0 E
299 G	*Gly*	0.86	Gly	1.00	Gly	1.00	F-F2	25.9 E
300 Q	Gln	1.00	Gln	1.00	Gln	1.00	F-F2	58.0 E
301 G	Gly	1.00	Gly	1.00	Gly	1.00	F-F2	80.1 H
302 K		0.33	*Basic*	0.50		0.05		38.4 H
303 E		−0.02		−0.12	Glu	1.00		91.9 H
304 N		0.51	Asn	1.00		0.55		74.1 H
305 A		0.32		0.42	Ala-Val	0.53		0.0 H
306 k		−0.43		−0.39	Basic	0.46		19.8 H
307 q		−0.44		−0.48		−0.18		67.6 H
308 y		0.07	Aromatic	0.34	Phe-Tyr	0.25		33.9 H
309 L	*Leu*	0.91	Leu	1.00	Leu	1.00		0.0 H
310 k		−0.21		0.03		−0.09		61.9 H
311 E		0.23	*Acidic*	0.33	Acidic	0.56		81.4 H
312 N		0.28		0.62		0.63		45.1
313 P		0.25		0.60		−0.19		53.8 H
314 e		0.02		−0.12		0.15		89.7 H
315 i		−0.12		−0.09	Aliphatic	0.16		26.9 H
316 A		−0.27		0.11		−0.44		2.6 H
317 a		−0.58		−0.52		−0.41		66.4 H
318 E		0.12		0.14	*Acidic*	0.45		46.9 H
319 I	*Aliphatic*	0.31	*Ile-Leu*	0.28	*Ile*	0.75		0.0 H
320 E		0.05		0.44	*Acidic*	0.39		26.8 H
321 k		−0.50		−0.52		−0.23		65.0 H
322 k		−0.22		−0.34		0.46		35.2 H
323 i	*Aliphatic*	0.28	*Aliphatic*	0.22	Aliphatic	0.54		0.0 H
324 R		0.31	*Arg*	0.94	*Basic*	0.36		28.4 H
325 E		0.14		−0.01		0.37		67.7 H
326 k		−0.15		−0.14		0.27		51.2 H
327 l		0.13		0.19		0.27		24.9 H
328 l		−0.00		0.13		0.00		16.8
329 n		−0.03		−0.07		0.00		
330 n		−0.03		−0.06		0.00		
331 p		−0.03		−0.04		0.00		
332 n		−0.02		−0.03		0.00		
333 s		−0.02		−0.01		0.00		
334 a		−0.01		−0.01		0.00		
335 l		−0.02		−0.02		0.00		
336 a		−0.01		−0.02		0.00		
337 a		−0.01		−0.01		0.00		
338 s		−0.01		−0.00		0.00		
339 i		−0.01		−0.02		0.00		
340 d		−0.00		−0.00		0.00		
341 d		−0.00		0.01		0.00		
342 s		−0.01		−0.01		0.00		
343 e		−0.00		−0.00		0.00		
344 a		−0.00		−0.00		0.00		
345 v		−0.00		−0.01		0.00		
346 a		0.00		0.00		0.00		
347 e		0.00		0.00		0.00		
348 s		0.00		0.00		0.00		
349 e		0.00		−0.00		0.00		
350 e		−0.00		−0.00		0.00		
351 d		0.00		0.00		0.00		
352 f		0.00		0.00		0.00		

[a] For each position, the CI (see Materials and Methods) is evaluated with respect to the global alignment of all 63 sequences and with respect to the alignment of the 33 A+B+C-Proteo sequences and of the 11 Gram(+) sequences. Conservation at different positions is assessed for individual amino acids or amino acid types such as acidic (E and D), basic (K and R), major hydrophobic (Major hydro; L, I, V, M, and F), aromatic (F, Y, and W), small hydroxyl (Small hydrxl; S and T), amide (N and Q), major and minor hydrophobics (L, I, V, F, A, and C), aliphatic (L, I, V, and A), and small (G, A, and S). A residue (individual or type) is marked in boldface when it is invariant for all sequences of the designated group and in italics (referred to as "strongly conserved") when it occurs in 60 of all 63 sequences, 31 of 33 Proteo sequences, and 10 of 11 Gram(+) sequences. S.A., percent surface accessibility; S.S., secondary structure (H, an α-helix position; E, a β-strand position). Functional regions (FUNCT.) are indicated as follows: L1 and L2, DNA binding loops 1 and 2; F-F1 and F-F2, filament-filament interaction regions.

Table 3. Each monomer interface section is substantially conserved (CI ≥ 0.5) with respect to the Proteo and Gram(+) sequence collections. Among the different monomer surfaces, M-M6 is almost perfectly conserved and the least conserved are M-M3 in Proteo sequences and M-M1 in Gram(+) sequences.

NBA is invariant in the Gram(+) sequences (actually covering positions 65 to 74) and strongly conserved for all sequences. The nucleotide functional hydrolysis B segment assigned to position 140 to 144 (the active site being D-144 [36]) maintains positions 140 to 143 as hydrophobic and D-144 as invariant. This is a consensus form of the B site in many *E. coli* ATP-binding proteins (e.g., see references 3 and 42). However, the immediate carboxyl segment, from positions 145 to 149, is perfectly conserved (amino acid specific) over all RecA se-

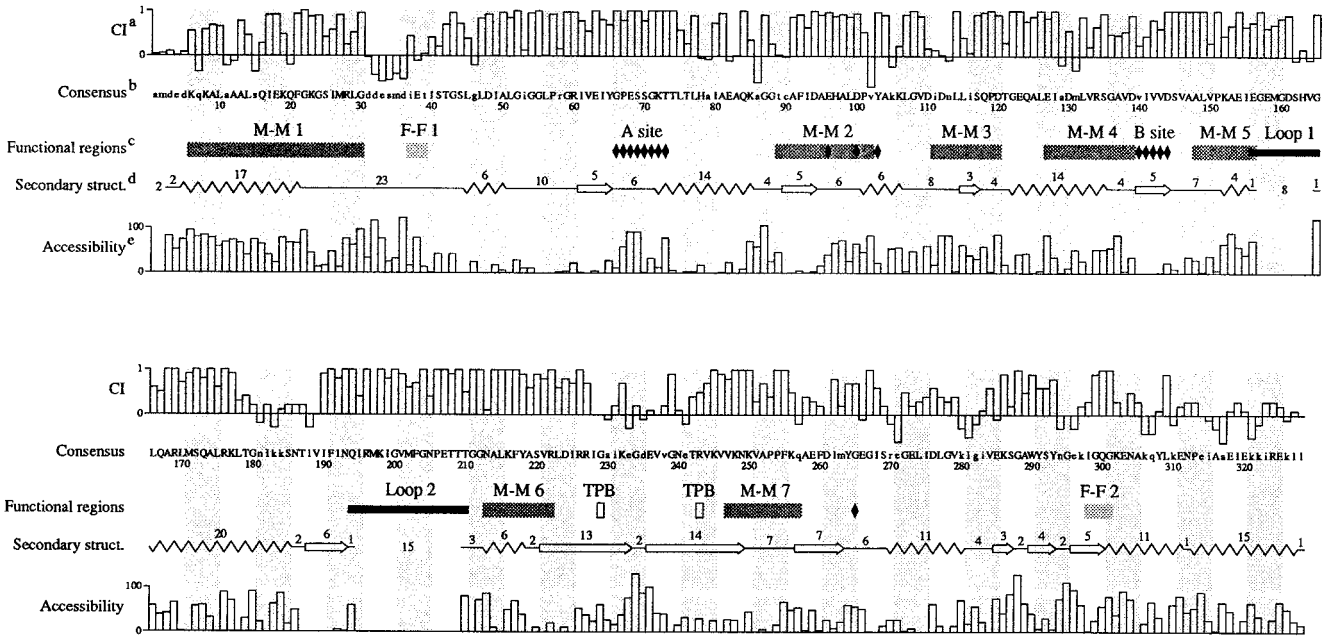


FIG. 1. Consensus of 63 RecA proteins mounted on the RecA structure of *E. coli*. (a) See Materials and Methods for the definition and calculation of CI values. (b) See Table 1, footnote *a*, concerning the determination of the consensus sequence. (c) Functional regions are assigned according to reference 25. Diamonds indicate nucleotide binding or hydrolysis. F-F, filament-filament interactions. (d) Secondary-structure determinations are derived with the DSSP program of Kabsch and Sander (18). Zigzag lines indicate α-helices, arrows reflect β-strands, straight lines represent coil regions, and missing lines correspond to regions not seen in the crystal structure. (e) Solvent accessibilities are evaluated according to the method of Richmond and Richards (35) and refer to the monomer.

quences, and on this basis we hypothesize that positions 144 to 149 constitute a key functional NBB region.

The CI evaluations for the loop DNA binding domains L1 and L2 indicate that the latter is significantly more conserved than the former (see Discussion).

For conservation and solvent accessibility, consistent with common knowledge, Table 3 reveals that buried residues have a significantly higher average CI (0.62) than do partially buried residues (average CI = 0.45) while the most exposed positions are still less conserved (CI = 0.38).

Variable regions. The open coil region from positions 31 to 41, with the exception of enterobacterial sequences, is substantially variable; for example, position 38 shows the amino acids D, E, P, Q, S, H, and K. Within this region the only conserved position is 37. Other variable regions include positions 110 to 113 (the amino end of the M-M3 surface), 180 to 188, 228 to 241, 270 to 271, and 279 to 287 and the 30 residues at the carboxyl terminus of RecA proteins. All of these positions mostly correspond to exposed positions and presumably are less critical functionally, or some part of this variability may reflect species-specific differentiated regions (see Discussion). The region from positions 228 to 241, which contains the proposed TPB site 229, is also substantially variable.

(i) Positions of strong contrasts between Proteo and Gram(+) RecA sequences. A position is considered deviant between the Proteo and Gram(+) sequence sets if the CI value exceeds 0.5 in one set but is nonpositive or about zero in the other set. Conservation differences between Proteo and Gram(+) sequences are generally reflected in differences between the respective consensus residues (Fig. 2).

(i) Residues conserved among Proteo sequences but variable among Gram(+) sequences (Table 2) include Q-16, in the M-M1 surface; T-210, at the carboxyl boundary of L2; and TBP site G-229. Other positions include A-12 or C-12, S-41, L-45,

TABLE 3. Average CI in different regions of RecA

Region	Position (amino acids)	Length (amino acids)	CI		
			Global	Proteo	Gram(+)
M-M1	6–30	25	0.48	0.69	0.54
M-M2	89–102	14	0.62	0.72	0.69
M-M3	111–120	10	0.52	0.56	0.62
M-M4	127–139	13	0.53	0.64	0.61
M-M5	148–156	9	0.75	0.84	0.80
M-M6	213–222	10	0.83	0.97	0.91
M-M7	247–257	11	0.68	0.72	0.76
Avg		92	0.60	0.72	0.67
Filament-filament 1	37–38	2	0.15	0.33	0.05
Filament-filament 2	298–301	4	0.80	0.91	0.85
Avg		6	0.58	0.71	0.59
NBA	66–73	8	0.95	0.99	1.00
NBB	140–144	5	0.61	0.63	0.65
Avg		13	0.82	0.85	0.87
DNA binding loop 1	156–165	10	0.57	0.63	0.82
DNA binding loop 2	194–210	17	0.77	0.93	0.84
Avg		27	0.70	0.82	0.83
α-Helix		124	0.40	0.55	0.52
β-Strand		76	0.51	0.70	0.59
Coil		153	0.45	0.53	0.55
Buried (≤7%)[a]		91	0.62	0.73	0.70
Intermediate (>7 and ≤40%)[a]		87	0.45	0.57	0.57
Exposed (>40%)[a]		125	0.38	0.55	0.46
Overall avg		303	0.47	0.61	0.56

[a] Values in parentheses show surface accessibility.

FIG. 2. RecA residue conservation between evolutionary groups. The global consensus sequence and differences with the Proteo consensus, the Gram(+) consensus, and the explicit *E. coli* sequences are displayed.

L-56 or V-56, K-106, T-121, N-186, E-235 or D-235, V-237 or I-237, E-273, and P-313.

(ii) Residues conserved in Gram(+) sequences but variable in Proteo sequences include H-78, A-79, G-88, acidic residues at position 112 (in the M-M3 surface), S-162, V-164 (in the L1 domain), D-261, S-269, E-271, E-303, and basic residues at position 306. For example, position 303 is significantly variable among Proteo sequences (CI = −0.119) but is maintained as E across all Gram(+) sequences.

(iii) Residues conserved in Proteo sequences and conserved in Gram(+) sequences but represented by quite distinct amino acid types include position 197 of the L2 region, where M is predominant among the Proteo sequences and E is predominant among the Gram(+) sequences; position 213 of the M-M6 surface, where the Gram(+) sequences maintain a basic residue but N is conserved in the Proteo sequences; and position 58, featuring the relatively large hydrophobic M strictly among the enterobacteria but a cationic residue, K or R, in almost all other RecA sequences.

(ii) *E. coli* sequence versus consensus Proteo sequence. Since the *E. coli* RecA protein has been used as a model RecA protein for functional and structural studies, it may be interesting to examine positions where the *E. coli* RecA sequence differs from the consensus of the Proteo sequences. There are 43 differences between the *E. coli* and the consensus Proteo RecA sequences traversing positions 3 to 325 (Fig. 2). Nineteen replacements are of similar residues (according to the similarity scoring matrix BLOSUM-62 [15], such as I↔V, L↔M, D↔E, R↔K, and Y↔F). However, potent physical and chemical differences are manifest at positions 88 (G↔K, a large steric change in the M-M2 interface) and 129 (C↔T, signifying a change from a hydrophobic to a polar residue). Less severe but negatively scoring changes occur at positions 53 (A↔I), 58 (M↔K), 82 (A↔E), 116 (in M-M3, C↔I), 167 (A↔Q), and 253 (A↔P). All these positions are highly conserved among enterobacterial sequences. There are no differences of any kind in the monomer interface regions of M-M1, M-M5, or M-M6. Comparing the *E. coli* sequence with a consensus enterobacterial RecA sequence (covering coordinates 3 to 310), we find 13 differences. These substitutions mostly involve similar amino acids. There are no differences from positions 3 through 104.

(iii) Variability in carboxyl regions. The carboxyl section among the RecA proteins is the most variable. This region has been proposed in *E. coli* to have regulatory roles in that deletion of portions of the C-terminal region leads to enhanced double-stranded DNA binding (2) and constitutive induction of SOS activities (41). We ascertained amino acid usages for the 30 carboxyl positions averaged over all 63 sequences and correspondingly for the Proteo and the Gram(+) sequences (Table 2). The following comparisons emerge. Acidic residue frequencies over the whole RecA protein, on average 13.6% (for all), 13.5% (for Proteo sequences), and 14.6% [for Gram(+) sequences], increase to 23.2, 24.0, and 23.9%, respectively, in the 30 carboxyl positions. Exceptions include MYXA2 and SYNP7 species sequences, which are devoid of acidic residues at the carboxyl end. Generally, for other amino acid types there is a balancing small reduction from their overall frequencies at the carboxyl positions. Since the carboxyl positions tend to be more exposed, the major hydrophobic residue frequencies (L, I, V, M, and F) are diminished from about 28 to 20% and glycine is cut from about 10 to 5%.

Examples of 3D structural conservation and RecA mutations in relation to alignment. In this section the degree of conservation of a position is evaluated in the context of its structural contact neighbors. Contact neighbors are defined to be residues whose minimum distance between their side chain atoms (designated d_m distance) is below a prescribed threshold, 4.5 Å. Contact neighbors for all positions are ascertained from the *E. coli* crystal structure coordinates. Here we illustrate a few examples of how functionally important positions are usually not only strongly conserved but also surrounded by conserved 3D contact neighbors (Tables 4 and 5). Some mutation examples in relation to RecA alignment are also discussed. A thorough analysis of RecA mutation results and evolutionary conservation will be presented elsewhere.

The d_m distance of 3.03 Å (0.303 nm) between the side chain atoms of R-60 and S-220, both highly conserved (Table 4), suggests a hydrogen bond. The solvent accessibility values of the structural neighborhood of R-60 depicts a predominantly buried ambience. The mutant R-60-C (R-60 replaced by C-60) produced a null allele (7). Is the H bond between R-60 and S-220 disrupted by the substitution C-60 with deleterious consequences to the protein structure?

TABLE 4. Examples of structural contact neighbor conservation[a]

Residue no. and neighbor	d_m distance (Å)[b]	Closest atoms	Consensus (CI)			Functional region[c]	% Solvent accessibility
			Global	Proteo	Gram(+)		
R-60			R (0.97)	R (0.94)	R (1.00)	—	21.4
S-220	3.03	N_η, O_γ	S (0.91)	S (1.00)	S (1.00)	M-M6	10.5
I-251	3.25	N_η, C_δ	ϕ[d] (0.30)	ϕ (0.23)	V (0.62)	M-M7	32.1
K-183	3.66	N_ε, N_ζ	K (−0.34)	K (−0.04)	N (0.36)	—	0.2
P-57	4.02	C_γ, C_γ	P (0.92)	P (0.85)	P (1.00)	—	2.3
M-35	4.25	N_η, C_ε	M (−0.42)	M (−0.31)	Q (−0.46)	—	0.2
E-96			E (0.97)	E (0.94)	E (1.00)	M-M2	40.0
S-145	2.60	O_ε, O_γ	S (1.00)	S (1.00)	S (1.00)	—	9.7
A-148	3.27	O_ε, C_β	A (1.00)	A (1.00)	A (1.00)	M-M5	29.4
D-144	3.59	C_γ, O_δ	D (1.00)	D (1.00)	D (1.00)	NBB	22.7
D-94	3.75	C_β, O_δ	D (1.00)	D (1.00)	D (1.00)	M-M2	3.1
A-98	4.13	C_β, C_β	A (0.92)	A (1.00)	A (1.00)	M-M2	72.8
Y-103			Y (0.97)	Y (1.00)	Y (1.00)	NB	21.6
T-74	2.90	O_η, O_γ	T (1.00)	T (1.00)	T (1.00)	—	6.2
D-100	3.50	C_β, C_δ	D (0.94)	D (1.00)	D (0.82)	M-M2–NB	65.2
G-267	3.72	O_η, C_α	G (1.00)	G (1.00)	G (1.00)	—	1.0
L-107	3.77	C_ε, C_δ	L (0.92)	L (1.00)	L (1.00)	—	3.7
G-265	3.80	C_ε, C_α	G (0.72)	G (0.88)	G (0.65)	NB	57.3
L-77	3.88	C_δ, C_δ	L (0.89)	L (1.00)	L (1.00)	—	3.3
I-262	4.03	O_η, C_γ	I (0.83)	I (1.00)	I (1.00)	—	8.4
K-152			Basic (0.45)	Basic (0.62)	Basic (0.49)	M-M5	64.8
E-156	2.80	N_ζ, O_ε	E (0.74)	E (1.00)	E (0.67)	M-M5–L1	72.7
T-121	3.69	C_γ, C_γ	T (0.28)	T (0.69)	T (−0.05)	—	19.1
E-123	3.83	C_β, O_ε	E (1.00)	E (1.00)	E (1.00)	—	42.2

[a] Contact neighbors for a reference residue (boldface) are defined as all residues of the tertiary structure whose closest side chain atoms are within 4.5Å (based on the RecA ECOLI structure) of at least one side chain atom of the reference residue (cf. 21). For the examples, the following characteristics of the reference residue and its contact neighbors are given: distance, closest atoms, CI with respect to all sequences (Global), CI to Proteo sequences, CI to Gram(+) sequences, the functional region, and percent solvent accessibility.

[b] d_m distance, minimum distance between side chain atoms. 1 Å = 0.1 nm.

[c] —, no functional region assigned.

[d] ϕ, major hydrophobic.

All the contact neighbors of E-96 among Proteo and Gram(+) proteins are invariant (except for changes of A-98 to S in two *Mycoplasma* sequences) (Table 4). Enigmatically, the similar-charge amino acid mutation E-96-D produces a null allele (4a). The residue E-96 is off the carboxyl end of a β-strand, and substituting by the residue D could be detrimental to the requisite β-strand secondary structure. The difference in size of E and D side chains might also engender charge repulsion in proximity to D-144 and D-94.

Residue Y-103 is a key residue contributing to nucleotide binding. The contact neighbors which surround Y-103, about half buried and half exposed, are clearly strongly conserved (Table 4).

On the basis of the close atoms for the pairing K-152 and E-156, a salt bridge formation is strongly suggested. A rather comprehensive mutational analysis of position 152 has been implemented (30) (Table 4). The mutant K-152-E leads to a phenotype with defective repair, and the mutant E-156-K leads to a coprotease constitutive phenotype. Presumably the replacement K-152-E (E-156-K) in 3D juxtaposition to E-156 (K-152) creates a disruptive repelling-charge ambience. Surprisingly, the mutation E-156-R has no negative effects on RecA functionality. It is possible that the delocalized charge of R and the hydrophobic methylene groups of K allow for a compatible interaction. The defective hydrophobic mutant construct K-152-I runs counter to the exposed character of the 3D environment of K-152. From this perspective, the hydrophilic substitutions of K-152 by A, N, Q, S, T, R, and H should not and indeed do not generate adverse RecA phenotypes.

The residues of the fundamental ATP binding domain from residues 66 to 73 (A site) in different combinations include as contact neighbors, apart from linear sequence neighbors, the hydrophobic residues I-225 or V-225, I-262, I-192, Q-194, and L-77, each invariant among the Proteo and Gram(+) sequences. Thus, the residues of the ATP binding domain in conjunction with their structural neighbors are almost all invariant (Table 2). Table 5 displays all residues in the segment from positions 140 to 149 and their side chain contact neighbors. The individual residues 144 to 149 are themselves invariant, and their structural contact neighbors are also highly conserved to about the same extent as the ATP A site. This contrasts markedly with the proposed B site, residues 140 to 144 (38), whose positions and contact neighbors are variable although restricted to hydrophobic amino acids.

With respect to structural contact neighbors, the two TPB sites G-229 and R-243, although moderately conserved across Proteo sequences, are quite variable in the Gram(+) sequences, with the interesting exception of the cationic conglomerate R-226, R-227, and K-245 (data not shown).

There are no disulfide bonds in the RecA protein structure, and cysteine in the RecA alignments behaves as a hydrophobic residue. No cysteine residues are highly conserved, but positions 90 (in M-M2), 116 (in M-M3), 129 (in M-M4), and 187 show many cysteines. Positions 90, 129, and 187 are totally buried. The contact neighbors of C-129 in the *E. coli* structure are exclusively the aliphatic amino acids V-143, I-93, L-178, I-141, and L-189. Nonhydrophobic mutations at C-129, as might be expected, tend to produce defective phenotypes (44).

TABLE 5. Reference residue and structural contact neighbor conservation in the region from residues 140 to 149[a]

Residue and neighbors	Distance/ Å	Global Conservation	Proteo Conservation	Gram(+) Conservation	Functional Region	% Solvent Acc.	Secondary Structure
V 140		0.23	0.38	0.18	NB B	0.0	E
C 90	3.51	-0.03	0.47	0.17	M-M 2	0.0	E
A 83	3.55	-0.01	-0.07	0.38	-	0.0	H
V 142	4.07	0.80	0.68	1.00	NB B	0.0	E
D 139	4.23	0.90	0.82	1.00	M-M 4	35.6	
L 188	4.25	-0.30	0.42	0.36	-	0.0	E
I 141		0.43	0.36	0.44	NB B	0.0	E
T 187	3.71	0.15	0.12	0.69	-	0.1	
V 138	3.85	0.53	0.70	0.17	M-M 4	4.6	
I 93	3.86	0.63	0.63	0.68	M-M 2	1.0	E
L 132	3.89	0.87	1.00	0.73	M-M 4	26.7	H
L 189	3.98	–0.03	0.24	0.74	-	0.0	E
C 129	4.05	–0.21	–0.14	0.21	M-M 4	0.5	H
V 143	4.33	0.60	0.70	0.63	NB B	0.7	E
A 91	4.34	0.87	0.91	1.00	M-M 2	0.8	E
L 182	4.42	0.21	0.28	0.24	-	0.2	H
V 142		0.80	0.68	1.00	NB B	0.0	E
T 76	3.33	0.26	0.47	0.44	-	3.7	H
I 192	3.84	0.76	0.92	1.00	-	5.1	E
I 80	3.88	0.72	0.91	0.68	-	0.2	H
V 140	4.07	0.23	0.38	0.18	NB B	0.0	E
F 92	4.17	0.92	1.00	0.84	M-M 2	4.3	E
I 190	4.32	0.88	0.84	1.00	-	0.0	E
V 143		0.60	0.70	0.63	NB B	0.7	E
C 129	3.56	–0.21	–0.14	0.21	M-M 4	0.5	H
F 191	3.69	1.00	1.00	1.00	-	0.0	E
L 189	3.86	–0.03	0.24	0.74	-	0.0	E
L 149	4.12	1.00	1.00	1.00	M-M 5	2.2	
I 93	4.22	0.63	0.63	0.68	M-M 2	1.0	E
I 141	4.33	0.43	0.36	0.44	NB B	0.0	E
D 144		1.00	1.00	1.00	NB B	22.7	E
S 145	3.54	1.00	1.00	1.00	-	9.7	
E 96	3.59	0.97	0.94	1.00	M-M 2, NB	40.0	
I 192	3.81	0.76	0.92	1.00	-	5.1	E
F 92	3.94	0.92	1.00	0.84	M-M 2	4.3	E
S 145		1.00	1.00	1.00	-	9.7	
E 96	2.60	0.97	0.94	1.00	M-M 2, NB	40.0	
D 144	3.54	1.00	1.00	1.00	NB B	22.7	E
A 147	3.62	1.00	1.00	1.00	-	36.4	
A 148	3.72	1.00	1.00	1.00	M-M 5	29.4	
I 192	4.49	0.76	0.92	1.00	-	5.1	E
V 146		1.00	1.00	1.00	-	0.1	
F 191	3.52	1.00	1.00	1.00	-	0.0	E
M 171	3.54	0.94	0.88	1.00	-	0.4	H
N 193	3.80	0.96	1.00	1.00	-	2.1	E
L 215	3.86	1.00	1.00	1.00	M-M 6	1.1	H
G 211	4.32	1.00	1.00	1.00	-	0.0	
A 147		1.00	1.00	1.00	-	36.4	
S 145	3.62	1.00	1.00	1.00	-	9.7	
N 193	4.36	0.96	1.00	1.00	-	2.1	E
Q 194	4.42	1.00	1.00	1.00	L2	59.8	
A 148		1.00	1.00	1.00	M-M 5	29.4	
E 96	3.27	0.97	0.94	1.00	M-M 2, NB	40.0	
A 95	3.60	0.91	0.92	1.00	M-M 2	16.3	
S 145	3.72	1.00	1.00	1.00	-	9.7	
L 149		1.00	1.00	1.00	M-M 5	2.2	
M 171	3.64	0.94	0.88	1.00	-	0.4	H
A 95	3.72	0.91	0.92	1.00	M-M 2	16.3	
G 122	3.91	1.00	1.00	1.00	-	0.3	H
A 125	4.11	0.87	1.00	0.38	-	0.9	H
V 143	4.12	0.60	0.70	0.63	NBB	0.7	E
M 170	4.32	0.65	0.59	0.67	-	2.4	H

[a] For the proposed extended B-site region 140 to 149, the CI of each position and of its contact neighbors with respect to all sequences, to Proteo sequences, and to Gram(+) sequences are determined as in footnote *a* of Table 4. Secondary structure (see footnote *a* of Table 2 for symbols) and percent solvent accessibility are also shown.

A wide range of mutant *recA* alleles in *E. coli* have been described, and a number of corresponding phenotypes have been characterized biochemically (e.g., see references 13, 22 to 28, 30, 31, 36, 37, 43, and 44). Biochemical mutation studies cover a broad span (associated with more than 280 mutant phenotypes), including conservative replacements such as I→V, L→M, K→R, and E→D, presumably maintaining hydrophobic, electrostatic, or chemical characteristics, or replacements of a more drastic nature such as E→K, significantly changing the charge ambience, or such as S→F, significantly changing volume and chemistry.

Some mutant alleles show enhanced activity relative to that of the wild type. Others are recombination defective or show differential repressor and UmuD cleavage specificities or produce toxic and null alleles. For example, the mutant alleles for V-37-M (V-37 is mutated to M) and E-38-K display abnormal levels of function (see references 25 and 36). In fact, E-38-K expresses constitutive coprotease activity in vivo and enhanced ATP and single-stranded DNA binding in vitro. Position 38 is considerably variable on an evolutionary scale. For example, the switch E-38-K is actually found in the BORBU RecA sequence. Also, in all C2p Proteo sequences E-38 is altered to P. For many of the B-group sequences (Table 1), the replacement E-38-Q is extant and viable. For many of the Gram(+) sequences, E is replaced by S and secondarily by Q. The *E. coli* structural neighbors of residue E-38 are residues N-186 and M-58. N-186 is maintained in almost all Proteo sequences. The Gram(+) N-186 is replaced by K or G. The consensus amino acid at positions 58 is a basic residue. Thus, the mutant E-38-K introduces a basic residue that already occurs at the 3D contact position 58. The additional positive charge of the mutant allele at position 38 may induce many RecA functional activities above their normal (optimal) needs.

The *E. coli* RecA mutant Q-184-K binds ATP and single-stranded DNA more efficiently than the wild type and exhibits enhanced repressor cleavage activity (43). Among most enterobacteria and *Pseudomonas* strains, Q-184 is replaced by the smaller amide N. The occurrence of K (the mutant form) or R at position 184 is actually seen in 43 distinct RecA proteins (Table 1); that is, the principal residue at position 184 is a positively charged amino acid.

The RecA mutants G-157-C or -D, R-169-C, and G-301-D all result in constitutive coprotease activity (30, 43). From the sequence comparisons, these positions are totally conserved, suggesting that these identities are required for correct function. The positions K-216, F-217, and R-222, all invariant, are apparently essential for RecA function and tolerate no substitutions (37). By contrast, the proximal positions N-213 and Y-218 can be broadly mutated with hardly any or no adverse effects (37). Consistently, these positions are variable in sequence comparisons. In *E. coli*, the I-225-V mutation produces a highly defective phenotype (13). Paradoxically, in Gram(+) species sequences the residue V-225 is predominant.

DISCUSSION

What do patterns of evolutionary conservation and variation tell us about the structural and/or functional importance of RecA sequence regions? Sequences of a protein family can typically be divided into subregions of three types: (i) conserved regions that generally reflect a common function or structure maintained by functional constraints (negative selection); (ii) freely varying regions, generally nonfunctional, changing by random drift (selectively neutral); and (iii) differentiated regions functionally adapted in part to new roles in different species (positive selection).

A common premise asserts that sequences essential to the functioning of proteins are conserved for species over a broad evolutionary range. In free regions among homologous proteins, amino acid replacements are basically randomly generated, conforming to the neutral theory of molecular evolution. A prototype example of freely varying regions occurs with the extended globin family (19). In other protein families there occur among protein members significant (nonrandom) sequence variations that reflect structural and functional differences in response to species environment, to variant mechanisms of oligomerization and regulation, and to other factors. In some respects different species do things differently, and we

expect that this also applies to RecA activities. For example, *Bacillus subtilis* RecA does not hydrolyze ATP as does *E. coli* RecA, but it does stringently require dATP (25, 26).

In the context of evolutionary alignments the following questions arise. (i) Consider a highly conserved position (or region) in the sequence. There are cases (relatively few in RecA) in which mutations in these positions produce wild-type phenotypes. Why are these substitutions not seen in the alignments? (ii) How can one account for a RecA mutation leading to null alleles in *E. coli* whereas the same amino acid occurs in a different bacterial sequence? (iii) Under what conditions do variable regions in the sequence correspond to neutral changes or rather to species-dependent positive selection reflecting significantly differentiated regions?

The conservation of positions that can be experimentally mutated to a limited extent while retaining wild-type behavior may be explained by the small probability associated with a restricted set of possible substitutions in the sampled evolution and in this way can account for their not being observed in the alignment. Alternatively, tests of functionality performed on the mutant proteins may not be sensitive enough to discern small reductions in functionality that would be selective on an evolutionary scale. A standard answer to question ii above refers to intragenic second-site suppressor effects of a compensatory nature, e.g., correlated changes at several positions. The phenomenon of significantly differentiated regions among members of a protein family reflecting species-specific structural and functional changes is pertinent to question iii.

In the following discussion we venture some interpretations and models of RecA structure and function based on sequence and structural conservation in the alignment. The most evolutionarily conserved segments in the sequence correspond to the ATP-binding domain A site (positions 66 to 73) and to the segment from positions 144 to 149. We propose the latter segment to be an important part of the B-site hydrolytic region. Another striking observation is the emphatic opposite charges on contiguous RecA units at their interface, a finding suggesting that electrostatic interactions mediate attainment of RecA polymerization. Other issues relate to some uncertainty in determining the RecA DNA binding domains, the LexA target sites, and the filament-filament contacts.

Interactions at M-M interfaces: some hypotheses. Monomer interfaces are substantially conserved, to about the same extent as buried residues. This argues for the importance of the residue makeup of the M-M interfaces. The 3D RecA structure shows that M-M1, M-M6, and M-M7 aggregate on one side and M-M2 to M-M5 aggregate on the opposite side of the monomer. There are several conserved charged residues broadly distributed in each monomer interface. However, M-M1, M-M6, and M-M7 emphasize a basic charge ambience, whereas M-M2 to M-M5 favor acidic residues. Specifically, monomer surfaces M-M1, M-M6, and M-M7 show an average charge of +8.2, ranging from +4 to +10, whereas the surfaces M-M2 to M-M5 show an average charge of −7.2, ranging from −4 to −9. We hypothesize that charge plays a decisive role in correctly localizing and orienting RecA monomers for polymerization. Interchain interaction analyses mostly center on the stabilizing effect of hydrophobic and aromatic interactions, while electrostatic interactions are generally believed to play a secondary role (e.g., see references 6, 14, and 45). However, interchain stabilization in quaternary protein structures and/or protein complexes via electrostatic interactions has many precedents (e.g., glutathione *S*-transferase [1gst, Brookhaven Protein Data Bank reference code], catalase [8act], coat viral protein [Southern bean mosaic virus] [4sbv], nitrogenase molybdenum-iron protein [1min] and fructose-1,6-bisphosphate aldolase [1fbc] [4]). Moreover, it was recently observed that the most frequent interchain specific residue contacts involve charged residues of the opposite sign (4). The preponderant opposite charge of the M-M1–M-M6–M-M7 surface versus the M-M2–to–M-M5 surface generates potent opposite-sign charge potentials (calculated by the software program DelPhi; Biosym Technologies, San Diego, Calif.). We suggest that this potential can produce a long-range electrostatic attraction which helps orient the surfaces. Close hydrophobic interactions and hydrogen bonding connections subsequently act to complete the polymerization process. In this respect, a number of the conserved residues that are involved in intermonomer hydrogen bonding (positions 97, 156, 218, and 222) or hydrophobic interactions (positions 116, 155, and 217) have been experimentally mutated, mostly leading to defective phenotypes (30, 31, 37, 44). Each monomer interface also contains several highly conserved small residues (G, A, or S), which may augment flexibility for precision in the M-M contacts (see, e.g., references 34 [page 43] and 9 [page 7]). Residues F-255 and Q-118 protrude from the M-M surfaces and contact many residues of the neighboring monomer. These residues can significantly contribute to the stability of the intermonomer contact and are thus attractive candidates for mutation experiments.

A number of pairs of charged residues in these monomer segments form intrachain salt bridges. These include the pairings D-130 with R-134 of M-M4, K-152 with E-156 of M-M5, and E-18 with K-23 of M-M1. These, together with hydrophobic interactions, undoubtedly increase stability within the monomer surface. Only one interchain salt bridge is observed (K-6 with D-139). With the prevalent opposite charge of M-M1, M-M6, and M-M7 versus M-M2 to M-M5, how can one reconcile the intermonomer shortage of salt bridges with our hypothesis on the importance of intermonomer electrostatic interactions? It is possible that specific charged residues not contacting in the ADP crystal do interact in the RecA active structural conformation. Alternatively, we propose that the charge interactions are nonspecific and mainly help to localize and orient contiguous monomer surfaces during polymerization. Moreover, in passing from the active to the inactive form, loose interactions may be desirable since the restraints of multiple salt bridges may render the RecA polymer overly stable.

Conservation of the ATP binding and hydrolysis sites. Nucleotide cofactors in the form of DNA-dependent ATP hydrolysis activity allosterically modulate the stability and conformation of RecA protein-DNA complexes. These NB sites tend to be highly conserved. Notably, the A site from residues 71 to 74 and the segment from residues 144 to 149 are among the longest invariant amino acid-specific segments over all 63 sequences. Not only the sequence but the complete 3D-structural environments of the A site and of the segment from positions 145 to 149 are highly conserved (see "Examples of 3D contact neighbor conservation and RecA mutations in relation to alignment" in Results and Table 4). By contrast, apart from the invariant residue D-144, the segment from residues 140 to 144, designated the ATP hydrolysis B site (25, 38), is predominantly hydrophobic but not amino acid specific. On this basis we hypothesize that the interval from residues 144 to 149 is structurally and/or functionally essential for ATP hydrolysis. The stringent requirements of this region are also emphasized by the unusual *cis*-peptide bond between D-144 and S-145 (38). Not surprisingly, the mutation of S-145-F produces a nonfunctional phenotype (47).

DNA binding domains. DNA is not included in the available crystal structure of RecA, and several regions in the RecA sequence have been proposed to interact with DNA. These include the regions from residues 157 to 165 (L1) and 194 to

210 (L2) (39). Only the amino half of L1 is conserved. This region contains three acidic and two glycine residues. Its proximity to the acidic M-M2–to–M-M5 surface and its conserved acidic residues suggest that L1 participates in M-M interactions. The absence of any cationic residue makes the hypothesis of DNA contacts less likely. Conceivably, L1 may be stabilized in the active ATP structure interacting with the neighboring monomer, or it may become stable on contact with LexA. L2 is overall strongly conserved, including two cationic residues which can facilitate RecA interaction with DNA.

On the basis of the results of photochemical (UV-induced) cross-linking procedures, the segments from residues 61 to 72, 233 to 243, and 178 to 183 and residue 103 are also proposed as DNA binding domains (29, 33). Residues 61 to 72 and 103 overlap the highly conserved NB sites. The substantial evolutionary variability of the regions from residues 233 to 243 and 178 to 183 is puzzling. However, the impressive conservation of positive charges in the structural environment of R-243 (R-226, R-227, and K-245) is consistent with binding to the negatively charged phosphate backbone of DNA, and this may argue that the key residue for binding DNA among positions 233 to 243 is 243 or its immediate structural environment. Position 183 emphasizes lysine among the C- and B-Proteo groups (consult Table 1) and mostly hydrophilic residues otherwise, and notably position 184 for most sequences presents a cationic residue (R or K) (see also reference 33). The segment from positions 1 to 33, also proposed to interact with DNA (16, 22, 48), includes three conserved cationic residues which do not participate in intermonomer contacts.

The segment from residues 1 to 33 of RecA and the two positions 183 and 184 are located at the border of the groove formed by the helical filament of RecA. Other cationic residues in a similar position include R-60, R-176, R-169, K-152, and R-134. These produce a substantial positive-charge potential (measured by the software program DelPhi; Biosym Technologies) that can be important for interacting with DNA. Accordingly, we speculate that this pervasive positive charge facilitates recognition of the negatively charged double-stranded DNA molecules. The interaction between these cationic residues and the DNA phosphates could be a first step in attracting the DNA molecule toward the axis of the filament, whereas the 3D region about R-243 may be a primary binding site for double-stranded DNA, ultimately leading to homologous recombination. The cationic residues in the groove may give different contributions to generating the positive-charge potential and/or to interacting with double-stranded DNA. This fact could account for some mutations not detectably altering RecA wild-type behavior (e.g., K-8-A, K-19-A, K-23-A [29], and K-152-A, -N, -Q, -S, -T, -H, or -R [30]) versus other mutations producing defective phenotypes (R-60-C [36] and R-169-C [43]). Interestingly, the mutation Q-184-K in *E. coli* augments the positive charge in this region and enhances DNA binding (43). An analogous scenario may explain the enhanced DNA binding activity of the mutation E-38-K (43).

LexA binding regions. Another interesting region for mutational studies pertains to the surroundings of the TPB sites G-229 and R-243. G-229 is conserved in Proteo sequences and mostly substituted by hydrophilic residues in other sequences. This position belongs to an elongated exposed β-strand and structurally could be substituted by a host of hydrophilic residue types in *E. coli* and presumably all other Proteo strains. It seems possible that constraints on G-229 are imposed on Proteo sequences by its interactions with LexA residues, as suggested by the reduced coprotease activity of the mutant G-229-S. Changes in the protein sequence of LexA could engender the variability of position 229 observed in non-Proteo RecA sequences. The proposed DNA binding region L1 (30, 39) has also been indicated as a LexA binding site (46), consistent with experimental evidence that excess DNA inhibits RecA coprotease activity on LexA (12, 40). In view of our hypothesis that the cationic concentration around R-243 serves DNA-binding activity and of the observation that L1 composition is not typical of DNA binding sequences, we propose that R-243 and its structural environment constitute the actual sites at which DNA and LexA compete for binding to RecA.

Filament-filament contacts. In the RecA crystal structure 25 residues participate in filament-filament structural contacts (see Results). Among these, positions 37 to 39 and 301 have been mutated, inducing coprotease constitutive phenotypes. On this basis, Story et al. (39) have proposed these sites as regulating the formation of filament-filament bundles. In their model, disruption of these bundles frees RecA monomers to engage in RecA–single-stranded DNA polymerization in conjunction with LexA cleavage. Virtually no interactions between conserved residues are explicit between filaments. Also, mutations at positions distinct from filament-filament contacts show constitutive coprotease activity (30). The low average conservation of the residues associated with most filament-filament contacts raises questions as to their functional role. However, there are a few highly conserved residues among the filament-filament sites, for example, W-290. This residue and several of its structural contact neighbors (K-286, G-299, Q-300, Y-291 or F-291, and G-288) are highly conserved. Tryptophan residues are generally buried in the protein core, but here W-290 is exposed and may be important for some functional activity.

Conclusions and perspectives. This paper attempts to integrate a robust alignment of RecA sequences with structure, function, and mutational information. Evolutionary conservation is evaluated not only with respect to individual sequence positions but also with respect to their structural contact neighbors. In this context, functionally important positions are usually conserved, as are their 3D contact neighbors. In this summary section it seems useful to highlight several new insights, experiments, and observations pertinent to RecA structure and function consistent with the sequence alignments.

The specific residues at positions 144 to 149 immediately carboxyl to the ATP hydrolysis B site are invariant across the alignment, and their structural contact neighbors are also highly conserved. Although many combinatorial mutagenesis experiments have been conducted on the residues of the ATP-binding A site (e.g., see references 23, 24, and 28), no mutant phenotypes (except S-145-F [47]) have been characterized relative to the extended B-site residues 140 to 149.

The conservation of charged residues at the M-M interface emphasizes basic residues on the surface of M-M1, M-M6, and M-M7 versus acidic residues on the opposite surface of M-M2 to M-M5. This suggests that electrostatic interactions play an intermediary role in RecA M-M polymerization. The importance of the charge contacts between M-M1, M-M6, and M-M7 and M-M2 to M-M5, apart from mutational manipulations, might be experimentally tested in vitro by varying the pH ambience or by subjecting the *E. coli* filament to a gradient of salt concentrations.

The amino half of the disordered L1 loop region is conserved, emphasizing anionic and glycine residues. Its composition and its proximity to the RecA acidic monomer surface suggest that L1 is involved in M-M interactions rather than DNA binding. On the other hand, there is evidence that positions 183, 184, and 243, predominantly cationic, may participate in protein-DNA interactions (29, 33). We propose that the strongly cationic structural environment of arginine 243 is the site of competition for DNA and LexA binding.

The large variability in the sequence alignment of currently identified residues associated with filament-filament contacts seems to indicate that these structural determinations are somewhat uncertain.

There are many strong contrasts in amino acid usages between the Proteo and Gram(+) sequences, which may relate to species-specific differences in RecA function. For example, position 197 in Proteo sequences uses mainly the hydrophobic amino acid M but in Gram(+) sequences uses mainly the acidic amino acid E. Invariant positions among Proteo sequences which are quite variable among Gram(+) sequences include L-45 and E-273, and positions invariant in Gram(+) sequences which are variable in Proteo sequences include H-78 and acidic residues at position 112. These can be attractive sites for biochemical and genetic studies.

In view of the alignment a natural program of mutation experiments with *E. coli* RecA is to make one or several substitutions of the kind revealed in comparing the *E. coli* RecA gene with other RecA bacterial sequences. Combinations of mutations at 3D contact neighbors of the positions under study can also be informative.

ACKNOWLEDGMENTS

We are happy to acknowledge valuable discussions with B. E. Blaisdell, V. Brendel, A. M. Campbell, J. Eisen, S. C. Kowalczykowski, W. M. Rehrauer, G. Weinstock, and E. Zaitsev on various aspects of the manuscript.

S.K. is supported in part by NIH grants 5R01GM10452-31 and 5R01HG00335-07 and NSF grant DMS 9403553.

REFERENCES

1. **Altschul, S. F.** 1993. A protein alignment scoring system sensitive at all evolutionary distances. J. Mol. Evol. **36:**290–300.
2. **Benedict, R. C., and S. Kowalczykowski.** 1988. Increase of the DNA strand assimilation activity of the recA protein by removal of the C-terminus and structure-function studies of the resulting fragment. J. Biol. Chem. **263:**15513–15520.
3. **Blaisdell, B. E., K. E. Rudd, A. Matin, and S. Karlin.** 1993. Significant dispersed recurrent DNA sequences in the *Escherichia coli* genome: several new groups. J. Mol. Biol. **229:**833–848.
4. **Brocchieri, L., and S. Karlin.** 1995. How are close residues of protein structures distributed in primary sequences? Proc. Natl. Acad. Sci. USA **92:**12136–12140.

4a. **Campbell, M.** Personal communication.

5. **Chothia, C., and A. Lesk.** 1986. The relation between the divergence of sequence and structure in proteins. EMBO J. **5:**823–826.
6. **Clackson, T., and J. A. Wells.** 1995. A hot spot of binding energy in a hormone-receptor interface. Science **267:**383–386.
7. **Clark, A. J.** 1973. Recombination-deficient mutants of *E. coli* and other bacteria. Annu. Rev. Genet. **7:**67–86.
8. **Clark, A. J., and S. J. Sandler.** 1994. Homologous genetic recombination: The pieces begin to fall into place. Crit. Rev. Microbiol. **20:**125–142.
9. **Creighton, T. E.** 1993. Proteins. Structures and molecular properties. Freeman, New York.
10. **Dayhoff, M. O., R. M. Swartz, and B. D. Orcutt.** 1978. A model of evolutionary change in proteins, p. 345–352. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biomedical Research Foundation, Washington, D.C.
11. **de Ley, J.** 1992. Introduction to the proteobacteria, p. 2110–2140. *In* H. G. Trüper, M. Dworkin, W. Harder, and K. H. Schleifer (ed.), The prokaryotes. Springer-Verlag, Berlin.
12. **DiCapua, E., M. Cuillel, E. Hewat, M. Schnarr, P. A. Timmins, and R. W. H. Ruigrok.** 1992. Activation of RecA protein. The open helix model for LexA cleavage. J. Mol. Biol. **226:**707–719.
13. **Dutriex, M., P. L. Moreau, A. Bailone, F. Galibert, J. R. Battista, G. C. Walker, and R. Devoret.** 1989. New RecA mutations that dissociate the various RecA protein activities in *Escherichia coli* provide evidence for an additional role for RecA protein in UV mutagenesis. J. Bacteriol. **171:**2415–2423.
14. **Freitag, N., and K. McEntee.** 1988. Affinity chromatography of RecA protein and RecA nucleoprotein complexes on RecA protein-agarose columns. J. Biol. Chem. **263:**19525–19534.
15. **Henikoff, S., and J. G. Henikoff.** 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA **89:**10915–10919.
16. **Horii, T., N. Ogawa, and H. Ogawa.** 1992. Inhibitory effects of N- and C-terminal truncated *Escherichia coli recA* gene products on functions of the wild-type *recA* gene. J. Mol. Biol. **223:**105–114.
17. **Johnson, M. S., and J. P. Overington.** 1993. A structural basis for sequence comparisons: an evaluation of scoring methodologies. J. Mol. Biol. **233:**716–738.
18. **Kabsch, W., and C. Sander.** 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers **22:**2577–2637.
19. **Karlin, S., V. Brendel, and P. Bucher.** 1992. Significant similarity and dissimilarity in homologous proteins. Mol. Biol. Evol. **9:**152–167.
20. **Karlin, S., G. Weinstock, and V. Brendel.** 1995. Bacterial classifications derived from RecA protein sequence comparisons. J. Bacteriol. **177:**6881–6893.
21. **Karlin, S., M. Zuker, and L. Brocchieri.** 1994. Measuring residue associations in protein structures. J. Mol. Biol. **239:**227–248.
22. **Kawashima, H., T. Horii, T. Ogawa, and H. Ogawa.** 1984. Functional domains of *Escherichia coli* RecA protein deduced from the mutational sites in the gene. Mol. Gen. Genet. **193:**288–292.
23. **Konola, J. T., K. M. Logan, and K. L. Knight.** 1994. Functional characterization of residues in the P-loop motif of the RecA protein ATP binding site. J. Mol. Biol. **237:**20–34.
24. **Konola, J. T., H. G. Nastri, K. M. Logan, and K. L. Knight.** 1995. Mutations at Pro^{67} in the RecA protein P-loop motif differentially modify coprotease function and separate coprotease from recombination activities. J. Biol. Chem. **270:**8411–8419.
25. **Kowalczykowski, S. C., D. A. Dixon, A. K. Eggleston, S. D. Lauder, and W. M. Rehrauer.** 1994. Biochemistry of homologous recombination in *Escherichia coli*. Microbiol. Rev. **58:**401–465.
26. **Kowalczykowski, S. C., and A. K. Eggleston.** 1994. Homologous pairing and DNA strand-exchange proteins. Annu. Rev. Biochem. **63:**991–1043.
27. **Liu, S.-K., J. A. Eisen, P. C. Hanawalt, and I. Tessman.** 1993. *recA* mutations that reduce the constitutive coprotease activity of the RecA1202(Prt^c) protein: possible involvement of interfilament association in proteolytic and recombination activities. J. Bacteriol. **175:**6518–6529.
28. **Logan, K. M., and K. L. Knight.** 1993. Mutagenesis of the P-loop motif in the ATP binding site of the RecA protein from *Escherichia coli*. J. Mol. Biol. **232:**1048–1059.
29. **Morimatsu, K., and T. Horii.** 1995. Analysis of the DNA binding site of *Escherichia coli* RecA protein. Adv. Biophys. **31:**23–48.
30. **Nastri, H. G., and K. L. Knight.** 1994. Identification of residues in the L1 region of the RecA protein which are important to recombination or coprotease activities. J. Biol. Chem. **269:**26311–26322.
31. **Nguyen, T. T., K. A. Muench, and F. R. Bryant.** 1993. Inactivation of the RecA protein by mutation of histidine 97 or lysine 248 at the subunit interface. J. Biol. Chem. **268:**3107–3113.
32. **Norioka, N., M.-Y. Hsu, I. Sumiko, and M. Inouye.** 1995. Two RecA genes in *Myxococcus xanthus*. J. Bacteriol. **177:**4179–4182.
33. **Rehrauer, W. M., and S. C. Kowalczykowski.** The DNA binding site(s) of the *Escherichia coli* RecA protein. J. Biol. Chem., in press.
34. **Richardson, J. S., and D. C. Richardson.** 1989. Principle and patterns of protein conformation, p. 1–98. *In* G. D. Fasman (ed.), Prediction of protein structure and principles of protein conformation. Plenum Press, New York.
35. **Richmond, T. J., and F. M. Richards.** 1978. Packing of alpha-helices: geometrical constraints and contact areas. J. Mol. Biol. **119:**537–555.
36. **Roca, A. I., and M. M. Cox.** 1990. The RecA protein: structure and function. Crit. Rev. Biochem. Mol. Biol. **25:**415–456.
37. **Skiba, M. C., and K. L. Knight.** 1994. Functionally important residues at a subunit interface site in the RecA protein from *Escherichia coli*. J. Biol. Chem. **269:**3823–3828.
38. **Story, R. M., and T. A. Steitz.** 1992. Structure of the RecA protein-ADP complex. Nature (London) **355:**374–376.
39. **Story, R. M., I. T. Weber, and T. A. Steitz.** 1992. The structure of the *E. coli RecA* protein monomer and polymer. Nature (London) **355:**318–325.
40. **Takahashi, M., and M. Schnarr.** 1989. Investigation of RecA-polynucleotide interactions from the measurement of LexA repressor cleavage kinetics. Eur. J. Biochem. **183:**617–622.
41. **Tateishi, S., T. Horii, T. Ogawa, and H. Ogawa.** 1992. C-terminal truncated *Escherichia coli* RecA protein RecA5327 has enhanced binding affinities to single- and double-stranded DNAs. J. Mol. Biol. **223:**115–129.
42. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1:**945–951.
43. **Wang, W.-B., and E. S. Tessman.** 1986. Location of functional regions of the *Escherichia coli* RecA protein by DNA sequence analysis of RecA protease-constitutive mutants. J. Bacteriol. **168:**901–910.
44. **Weisemann, J. M., and G. M. Weinstock.** 1988. Mutations at the cysteine codons of the recA gene of *Escherichia coli*. DNA **7:**389–398.
45. **Wells, J. A.** 1991. Systematic mutational analyses of protein-protein interfaces. Methods Enzymol. **202:**390–411.
46. **Yu, X., and E. H. Egelman.** 1993. The LexA repressor binds within the deep helical groove of the activated RecA filament. J. Mol. Biol. **231:**29–40.
47. **Zaitsev, E., A. Alexseyev, V. Lanzov, L. Satin, and A. J. Clark.** 1994. Nucleotide sequence between *recA* and *alaSp* in *E. coli* K12 and the sequence change in four recA mutations. Mutat. Res. **323:**173–177.
48. **Zlotnick, A., and S. L. Brenner.** 1988. An alpha-helical peptide model for electrostatic interactions of proteins with DNA, the N terminus of RecA. J. Mol. Biol. **209:**447–457.