Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 544–548, January 1992
Cell Biology

Molecular structure of the human desmoplakin I and II amino terminus

(desmosomes/intercellular junctions/alternative splicing)

MARIA LUISA A. VIRATA*, RITA M. WAGNER*, DAVID A. D. PARRY†, AND KATHLEEN J. GREEN*‡

*Department of Pathology and the Cancer Center, Northwestern University Medical School, Chicago, IL 60611; and †Department of Physics and Biophysics, Massey University, Palmerston North, New Zealand

Communicated by Irving M. Klotz, September 27, 1991 (received for review July 25, 1991)

ABSTRACT **Desmoplakins (DPs) I and II are closely related proteins found in the innermost region of the desmosomal plaque, which serves as a cell surface attachment site for cytoplasmic intermediate filaments. Overlapping cDNA clones comprising 9.2 kilobases of DP-I, predicted to encode a full-length 310-kDa polypeptide (2677 amino acid residues), have now been identified. Here we report the predicted protein sequence and structural analysis of the N terminus of DP, extending our previous study of the rod and carboxyl domains. The N terminus contains groups of heptad repeats that are predicted to form at least two major α-helical-rich bundles. Unlike the rod and carboxyl domains, the N terminus did not display a periodic distribution of charged residues. Northern blot mapping and genomic sequence analysis were also undertaken to examine the organization of the DP mRNAs. A 1-kilobase intron was located at the 3′ boundary of a DP-I-specific region; however, instead of an intron at the 5′ junction, a possible splice donor site was observed within a potential coding sequence, suggesting alternative RNA splicing from an internal donor site.**

Desmosomes are major sites of intercellular contact found in a variety of cells—i.e., epithelial cells, cardiac myocytes, arachnoidal cells of the meninges, and dendritic reticulum cells of germinal centers in lymph nodes (1–3). The desmosomal plaques serve as specific attachment points for intermediate filaments (IFs), which form a structural framework for the cytoplasm. Of the proteins comprising the desmosomal plaques, two closely related major components, desmoplakins (DPs) I and II [reported sizes vary from 240 to 285 kDa and from 210 to 225 kDa, respectively (4–6)], are among the candidate linkers for mediating IF anchorage to the cell surface. These two proteins have been localized to the fibrous, less electron-dense portion of the plaque through which IFs appear to loop and are associated with the IF cytoskeleton during desmosome assembly (7).

Until recently (8, 9), relatively little was known about the molecular structure of DP-I and -II (DNA or protein level), the regulatory mechanisms governing DP expression, the functional roles of the DPs in desmosome assembly, or their relationships with other desmosomal components and cytoskeletal elements, such as IFs. In a previous study (9), computer-aided analysis carried out on a partial predicted DP-I amino acid sequence identified two major structural domains within DP-I: a rod-like central domain and a globular C terminus. Here we report the sequence and predicted structure of the DP N terminus.§ Using this information, we propose a model for the entire DP-I polypeptide that is consistent with rotary shadowed electron microscope images showing DP-I to be a dumbbell-shaped molecule (6). We have also examined in more detail the mechanism by which the highly related DP-I and -II mRNAs are generated. Northern blot mapping and sequence analysis of a DP genomic clone were performed to confirm the boundaries of the DP-I-specific region and to investigate the origin of the two messages. The current evidence suggests that the mature mRNAs are generated by a mechanism of alternative splicing in which one of two alternative donor sites is selected for a single acceptor site. Of the two donor sites, the more 5′ site is an internal one contained within the potential coding sequence for DP-I.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

cDNA and Genomic Library Screening. The human keratinocyte λgt11 cDNA library (10) was obtained from John Stanley (National Cancer Institute, National Institutes of Health), while the human Charon 4A genomic library (11) was a gift from Richard Scarpulla (Northwestern University Medical School). Previously characterized DP cDNAs (9) were used as probes for library screening (12). The 5′ limits of the DP cDNA clones containing additional 5′ sequence were verified by PCR using λgt11 primers [18 base pairs (bp)] in combination with internal DP primers (15–30 bp).

DNA and Predicted Amino Acid Sequence Analysis. DP cDNA and genomic inserts from phage clones and PCR products were subcloned into Bluescript vectors (Stratagene) and were sequenced on both strands as described (9). The DNA sequence from PCR-generated inserts was confirmed by comparison with overlapping DP cDNAs and other PCR products. Sequence analysis was performed by using the AASAP and PC Gene software package (IntelliGenetics). A sequence homology search was done using the Protein Identification Resource protein data base (Release 27.0, December 1990). Fast Fourier transform (FFT) analyses to determine the periodic distribution of amino acid residues (charged or apolar) were carried out as described (13, 14).

RNA/DNA Blotting and Primer Extension. Total RNA was isolated from normal human epidermal keratinocytes by a single-step procedure (15). Northern and Southern blot analyses were carried out as described (9). Primer extension was carried out as described (16) using 1×10^5 cpm of a ^{32}P-end-labeled oligonucleotide primer (30 bp long) and 10–20 μg of total keratinocyte RNA. The primer/RNA mixture was hybridized overnight at either 30°C, 37°C, or 42°C. Extension reaction mixtures using 50 units of a modified Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories) were incubated at 42°C for 1.5 hr. Primer-extension products were then run on a 6% polyacrylamide/urea gel at 2000 V.

Abbreviations: DP, desmoplakin; IF, intermediate filament; FFT, fast Fourier transform.

‡To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M77830).

RESULTS

Origin of the DP-II mRNA. Overlapping DP cDNA clones encoding the N terminus were isolated from the keratinocyte cDNA library and verified by PCR. Radiolabeled cDNA fragments or PCR products derived from the clones were used to probe total RNA from normal human keratinocytes on Northern blots to determine whether these hybridized to the 9.5-kilobase (kb) DP-I message and/or the 7.5-kb DP-II message (Fig. 1*A*). All of the N-terminal fragments tested hybridized to both DP-I and -II bands of the expected message sizes. The previously identified 1797-bp DP-I-specific region (9) was also verified by RNA blot analysis. In Fig. 1*A*, probes p36 and p46 hybridized to the 9.5-kb DP-I message only, while probes derived from the regions flanking the 1797-bp sequence recognized both DP-I and -II messages.

To examine in more detail the mechanism for generating the DP-I and -II mRNAs, we isolated and characterized an ≈14-kb DP genomic clone from a human Charon 4A library. DNA sequence analysis of this genomic clone uncovered a 1-kb intron situated precisely at the 3′ boundary of the DP-I-specific region (Fig. 1*B*). However, there was no intron at the 5′ end of the DP-I-specific 1797-bp region. Instead, closer inspection of the 5′ junction revealed the nucleotide sequence AAG/GTAAGA, which conformed closely to the consensus sequence for exon splice donor sites, AAG/GTAAGT (17, 18).

Characterization of the DP N-Terminal Sequence. Sequence analysis of the overlapping 5′ end clones brought the total sequence of the DP-I cDNA to 9.2-kb of the total 9.5 kb expected by Northern blot analysis (Fig. 2). Primer-extension experiments were carried out by using a 30-bp oligomer derived from the 5′ end (location of the primer is underlined in Fig. 2). As shown in Fig. 3, a primer extension product of 343 bp showed up consistently in three samples that were incubated overnight at various hybridization temperatures. Taking into account the size and location of the primer, the transcriptional start site is estimated to be 295 bp upstream of the 5′ end of the sequence shown in Fig. 2 and brings the total lengths of the DP-I and DP-II transcripts to 9.5 and 7.7 kb, respectively. These values correspond well with the message sizes of 9.5 and 7.5 kb previously predicted by Northern blot analysis (9). Faint bands of larger molecular mass were also seen in all three samples, suggesting the presence of minor transcriptional initiation sites in the DP gene.

From the first in-frame ATG triplet (methionine), an open reading frame of 8031 bp is predicted to encode a 310-kDa polypeptide of 2677 amino acid residues. The areas surrounding the second and third ATG codons also conformed closely to the consensus sequence for ribosome binding sites (19) and thus are also marked as possible translational start sites (Fig. 2). Two independent clones were sequenced to confirm the sequence of this region containing the putative translational start sites. Based on the designation of the DP-I-specific region, a 238-kDa DP-II protein is expected from an open reading frame of 6234 bp (2078 amino acid residues). From the predicted protein sequences, the pI values of DP-I and -II were estimated to be 6.3 and 6.4, respectively, as compared to pI 6.8–7.2 previously reported for the DPs (4). On the other hand, the percentages in amino acid composition (data not shown) compared favorably with results from actual amino acid compositional analysis (6).

Predicted Structure of the DP N Terminus. Heptad repeat substructure, which is characteristic of α-fibrous proteins (20) and underlies the predicted formation of the DP rod

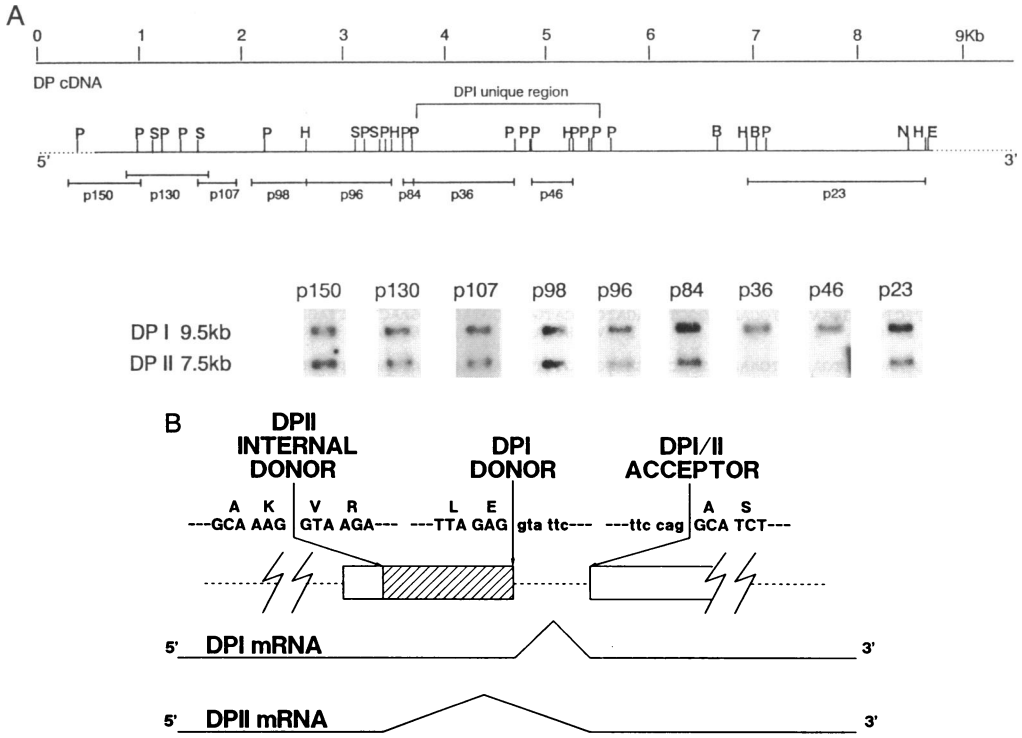

FIG. 1. Boundaries of the DP-I-specific region. (*A*) Northern blot mapping of a DP-I-specific region and the DP common regions. Replica blots of total normal human keratinocyte RNA were probed with radiolabeled cDNA fragments (p150, p130, p107, p98, p96, p84, p36, p46, p23) derived from overlapping λgt11 DP phage clones. The relative sizes and positions of the probes used are indicated below the restriction map of the entire DP-I cDNA. Message sizes for DP-I and -II are 9.5 and 7.5 kb, respectively. P, *Pst* I; S, *Sac* I; H, *Hin*dIII; B, *Bam*HI; N, *Nco* I; E, *Eco*RI. (*B*) Diagram depicting the location of the intron/exon borders within the region thought to be involved in generating the alternatively spliced DP-I and -II mRNAs. Here, exons are denoted as rectangles (the 1797-bp DP-I-specific region is denoted by a hatched box), while introns are indicated as dashed lines. Sequences for the alternative donor sites (one of which is internal) and a single acceptor site are also indicated. Intron/exon borders were based on DNA sequence analysis of an ≈14-kb Charon 4A genomic clone and previously characterized DP-I and DP-II cDNA clones (9).

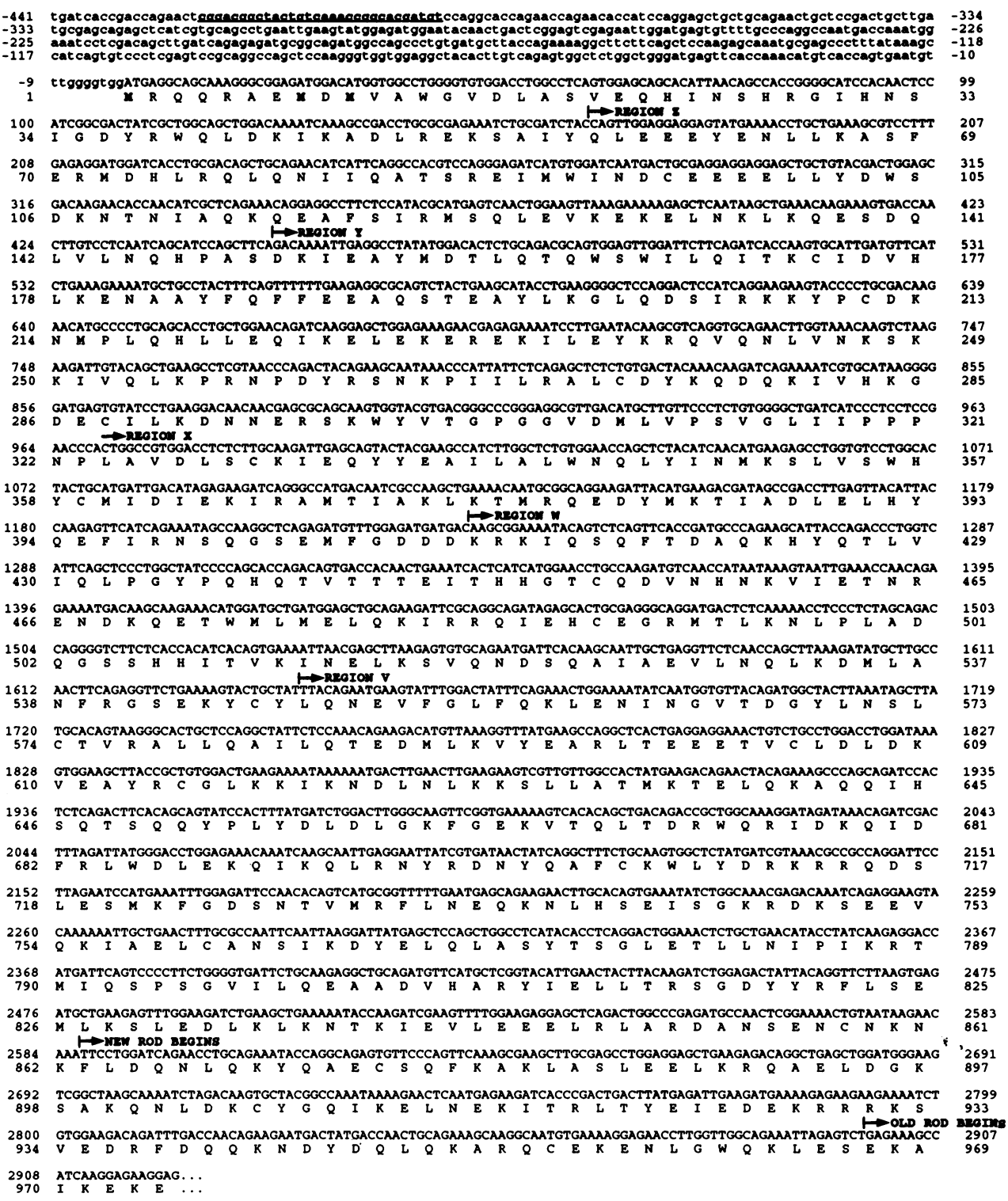

FIG. 2. Nucleotide and predicted amino acid sequence of DP-I. New and old designations (9) for the rod domain, as well as the 5′ boundaries of the N-terminal regions V, W, X, Y, and Z are indicated by arrows. The first three methionine (M) residues are marked in boldface to denote the location of putative translation initiation sites. The sequence complementary to the 30-bp oligonucleotide used in primer extension is underlined.

domain (9), is also prevalent in much of the N terminus. In contrast, however, to the extensive series of heptads in the rod, the stretches of heptads in the N terminus are shorter (Fig. 4). Within the N-terminal domain, two major heptad-containing regions have been predicted to be largely α-helical and these have been designated as V and Y (each with five strands of heptads but with 55 and 35 amino acid residues per strand, respectively). Three minor ones are designated W (three strands, each $\approx$30 amino acids long), X and Z (each with two strands $\approx$20 amino acids long) (Fig. 5). As shown in Table 1, the overall charged/apolar ratio is relatively low—0.72—for these segments, and this is consistent with the formation of a compact bundle rather than an extended rope (21), in which the ratio is generally much higher—$\approx$1.3. Calculations also show that there would have been very few interchain ionic interactions stabilizing a two-stranded rope if

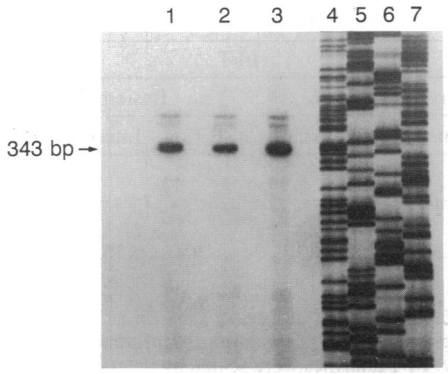

FIG. 3. Primer extension. Autoradiogram of a 6% polyacrylamide/urea gel of primer extension products hybridized overnight at different temperatures before reverse transcription at 42°C for 1.5 hr. Lanes: 1–3, primer extension products hybridized at 30°C (lane 1), 37°C (lane 2), or 42°C (lane 3); 4–7, dideoxynucleotide chain termination reactions (A, C, G, T) used as a molecular size ladder. A 343-bp labeled product generated under all three conditions is indicated by the arrow. The faint larger molecular mass bands that appear consistently in lanes 1–3 may represent minor transcriptional start sites.

that structure had been adopted in preference to bundles. For these reasons, we believe that the conformation adopted is that of several α-helical bundles consisting largely of antiparallel strands. A FFT analysis was performed to detect periodic distributions of acidic and basic residues in various parts of the N terminus; however, no significant long-range period could be detected (Table 1). This is in contrast to the rod and C domains, which both contained significant periodicities of 10.4 and 9.5, respectively (9).

DISCUSSION

Here we report the cDNA and predicted amino acid sequence of the DP N terminus, and we investigate further the mechanism by which the related DP-I and -II mRNAs are generated. Early studies have emphasized the close biochemical similarities shared by both DP-I and -II. We have previously shown that the DPs are encoded by separate messages derived from a single precursor transcript, possibly by alternative splicing (8, 9). Here we did not detect any sequences unique to DP-I or DP-II, other than the previously described DP-I-specific region (Fig. 1*A*). Our analysis of the splice sites flanking the DP-I unique region revealed a 1-kb intron at the 3′ end but none at the 5′ junction. In lieu of an intron at the 5′ end of the DP-I-specific region, we found a nucleotide sequence AAG/GTAAGA, which matched almost perfectly the consensus sequence for exon donor splice sites. This observation suggests that optional removal of the DP-I-specific region occurs via splicing at an internal donor site.

Residues:	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*a*	*b*	*c*	*d*	*e*	*f*	*g*
652-665	Y	P	L	Y	D	L	D	L	G	K	F	G	E	K
666-679	V	T	Q	L	T	D	R	W	Q	R	I	D	K	Q
680-693	I	D	F	R	L	W	D	L	E	K	Q	I	K	Q
694-707	L	R	N	Y	R	D	N	Y	Q	A	F	C	K	W
708-728	L	Y	[DRKKRQDSLESMKFGDSNT]											
729-742	V	M	R	F	L	N	E	Q	K	N	L	H	S	E
743-756	I	S	G	K	R	D	K	S	E	E	V	Q	K	I
757-770	A	E	L	C	A	N	S	I	K	D	Y	E	L	Q
771-784	L	A	S	Y	T	S	G	L	E	T	L	L	N	I

FIG. 4. Heptapeptide repeats of the form $(a,b,c,d,e,f,g)_n$ within bundle V of the DP-I N terminus (residues 652–784), wherein positions a and d are usually occupied by apolar residues. Conserved apolar residues are indicated in shaded boxes. Nonheptad sequence is in brackets.

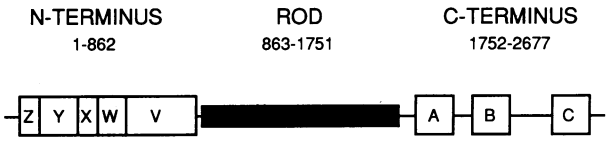

FIG. 5. Diagram of the model of the DP-I polypeptide. The heptad-containing N-terminal subregions are grouped into two major bundles (V, Y) and three minor ones (W, X, Z), while the C-terminal subregions containing the 38-residue motif are designated as bundles A, B, and C. At the central rod domain, two adjacent DP molecules are predicted to dimerize into an α-helical coiled-coil of 132 nm, which is flanked by globular end domains.

This type of mechanism usually involves at least two alternative donor sites competing for a single acceptor site during pre-mRNA splicing. Alternative RNA splicing via an internal donor site has been previously demonstrated with several genes—i.e., human fibronectin, human granulocyte colony-stimulating factor, mouse immunoglobulin heavy chains, chicken ovomucoid, and porcine plasminogen activator (22).

On the basis of the predicted amino acid sequence for the DP N terminus, we can now propose a model for the entire DP-I polypeptide (Fig. 5). This model, comprising a central rod domain and globular end domains, is consistent with the appearance of rotary shadowed electron microscope images of purified DP-I (6). In a previous report (9), the rod domain was predicted to be capable of self-aggregation and assembly into higher-ordered filamentous structures, and the C terminus was proposed to contain a possible binding site for IF or IF-like proteins. Here we focused on the structural analysis of the DP N terminus, which was based on predicted amino acid sequence derived from sequencing DP 5′ cDNA clones. The additional N-terminal cDNA sequence brought the open reading frames containing potential coding sequence to 8031 bp for DP-I and 6234 bp for DP-II. Based on the predicted protein sequence, we calculated the molecular masses of DP-I and -II to be 310 and 238 kDa, respectively, which far exceed the values previously deduced by SDS/PAGE (4–6). The difference in size is likely to be due to the known aberrant migration of fibrous proteins on SDS/polyacrylamide gels, as in the cases of dynein and plectin (23). The first three ATG codons that occurred in-frame were all marked as potential translational initiation sites (Fig. 2). N-terminal peptide sequencing using automated Edman degradation (Northwestern University Biotechnology Facility) was attempted to identify the initiating methionine but proved unsuccessful because of blockage of the N terminus.

The basic structural unit for much of the DP N terminus is the heptad repeat. Unlike the rod, though, the heptad series in the N terminus is often interrupted by relatively short nonheptad sequences, which may allow folding of the α-helical stretches of heptads onto each other. The resulting antiparallel bundles could be stabilized primarily by apolar interactions in a quasi-coiled-coil-like manner (21). To confirm that a bundle rather than a rod-like structure is present, we have calculated the charged/apolar ratio for the heptad-containing sequences as well as the maximum number of interchain ionic interactions that could stabilize a rod structure. In both cases, the relevant values are low (0.72 and 0.095 interaction per heptad pair, respectively) compared to those expected of a coiled-coil rod (typically $\approx$1.3 and 0.5 interactions per heptad pair, respectively) (Table 1). This predicted conformation of the V, W, X, Y, and Z subdomains is reminiscent of the C-terminal subdomains A, B, and C, whose structural unit is based on a 38-residue repeat. This also adopts an α-helical conformation capped by a β-turn, although it lacks the heptad substructure. In neither of the end domains, however, is the relative spatial disposition of the subdomains known.

Table 1. Summary of structural analysis of DP-I domains

Domain	Mass, kDa	Vol, nm^3	Length, nm	Charged/apolar		FFT					
						Acidic residues (D, E)			Basic residues (K, R)		
				Domain	Heptads	Period	Intensity	Probability	Period	Intensity	Probability
N terminus (residues 1–862)	101.2	125.8	—	0.81	0.72	—	—	—	—	—	—
Rod domain (residues 863–1751)	106.8*	130.8	132	1.52	1.27	10.4	36.36	1.62×10^{-16}	10.4	16.67	5.76×10^{-8}
C terminus (residues 1752–2677)	101.8	126.6	—	0.78	—	9.5	7.66	4.7×10^{-4}	9.5	8.26	2.6×10^{-4}

Molecular mass, volume, and length predictions are based on the predicted amino acid sequence. The numbers of interchain ionic interactions per heptad pair for the DP-I rod domain and N terminus are 0.54 and 0.095, respectively. FFT analyses were scaled by the methods of McLachlan and Stewart (13) and Parry (14). FFT values for the C terminus are the average values of the three homologous regions A, B, and C (9). *Taking into account the DP-I-specific region, the DP-II rod domain is estimated to be 35.2 kDa.

With the completion of DP-I sequencing, we have reevaluated the previous predictions for the rod boundaries (9) by using FFT analysis on the entire protein sequence. After comparing the FFT intensities for several possible rod lengths, we found that amino acid residues 863–1751 comprise the best defined rod, thereby moving the 5′ and 3′ rod boundaries upstream of their former designations 104 and 74 bp, respectively. These new assignments were based on maintenance of a significant periodicity of charged residues (with somewhat increased intensity) and the addition of six extra interhelix interactions at the N terminus (in contrast to zero interactions at the 3′ end of the old rod), which resulted in a more stable α-helical coiled-coil dimer. With this new boundary assignment, the rod was calculated to be 132 nm long (Table 1), which still compared favorably with the observed length of 130 nm of purified DP-I dimers (6).

Unlike the rod domain and C terminus, FFT analysis on the N-terminal domain did not reveal any extensive periodic distribution of charged residues (Table 1). Such a feature would be expected only under two conditions: (*i*) when there is a well-defined structural repeat, such as the 38-residue repeats in the C-terminal subdomains A, B, and C, or (*ii*) when the domain has an interacting rod structure. Neither of these situations occurs in the N domain.

A sequence homology search of the Protein Identification Resource data base revealed no additional significant homologies other than those previously reported (9, 10), in which the C-terminal region was shown to be similar to that of the 230-kDa bullous pemphigoid antigen, a known hemidesmosomal plaque component. Recently, though, a large IF-associated protein called plectin has been reported to contain the 38-residue repeating motif found in this region of DP and the 230-kDa bullous pemphigoid antigen (23). This observation is consistent with the suggestion that these molecules are encoded by a gene family whose protein products play a role in the organization of IF networks (10).

We thank Dr. John Stanley for the human keratinocyte λgt11 cDNA library; Dr. Richard Scarpulla for the human Charon 4A genomic library; Dr. Rex Chisholm, Dr. Ka-Leung Ngai, Dr. M. Kathleen Rundell, and Bruce Ostrow for helpful discussions; and Laura Nilles for technical assistance. This work was aided by grants from the National Institutes of Health (Grant HD24430), The Council for Tobacco Research, USA, Inc. (Grant 2432), a March of Dimes Basil O'Connor Starter Scholar Award, and an American Cancer Society Junior Faculty Research Award to K.J.G.

1. Steinberg, M. S., Shida, H., Giudice, G. J., Shida, M., Patel, N. H. & Blaschuk, O. W. (1986) *Ciba Found. Symp.* **125,** 3–25.
2. Garrod, D. R., Parrish, E. P., Mattey, D. L., Marston, J. E., Measures, H. R. & Vilela, M. J. (1990) in *Morphoregulatory Molecules*, eds. Edelman, G. M., Cunningham, B. A. & Thiery, J. P. (Wiley, New York), pp. 315–339.
3. Schwarz, M. A., Owaribe, K., Kartenbeck, J. & Franke, W. W. (1990) *Annu. Rev. Cell Biol.* **6,** 461–491.
4. Mueller, H. & Franke, W. W. (1983) *J. Mol. Biol.* **163,** 647–671.
5. Jones, S. M., Jones, J. C. R. & Goldman, R. D. (1988) *J. Cell Biochem.* **36,** 223–236.
6. O'Keefe, E. J., Erickson, H. P. & Bennett, V. (1989) *J. Biol. Chem.* **264,** 8310–8318.
7. Green, K. J. & Jones, J. C. R. (1990) in *Cellular and Molecular Biology of Intermediate Filaments*, eds. Goldman, R. D. & Steinert, P. M. (Plenum, New York), pp. 147–171.
8. Green, K. J., Goldman, R. D. & Chisholm, R. L. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2613–2617.
9. Green, K. J., Parry, D. A. D., Steinert, P. M., Virata, M. L. A., Wagner, R. M., Angst, B. D. & Nilles, L. A. (1990) *J. Biol. Chem.* **265,** 2603–2612.
10. Tanaka, T., Parry, D. A. D., Klaus-Kovtun, V., Steinert, P. M. & Stanley, J. R. (1991) *J. Biol. Chem.* **266,** 12555–12559.
11. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) *Cell* **15,** 1157–1174.
12. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
13. McLachlan, A. D. & Stewart, M. (1982) *J. Mol. Biol.* **162,** 693–698.
14. Parry, D. A. D. (1975) *J. Mol. Biol.* **98,** 519–535.
15. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
16. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1989) *Current Protocols in Molecular Biology* (Greene and Wiley, New York).
17. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472.
18. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
19. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8132.
20. Conway, J. F. & Parry, D. A. D. (1990) *Int. J. Biol. Macromol.* **12,** 328–334.
21. Cohen, C. & Parry, D. A. D. (1990) *Proteins: Struct. Funct. Genet.* **7,** 1–15.
22. Breitbart, R. E., Andreadis, A. & Nadal-Ginard, B. (1987) *Annu. Rev. Biochem.* **56,** 467–495.
23. Wiche, G., Becker, B., Luber, K., Weitzer, G., Castañon, M. J., Hauptmann, R., Stratowa, C. & Stewart, M. (1991) *J. Cell Biol.* **114,** 83–99.