Hypothesis

α-Helical solenoid model for the human involucrin

A.V. Kajava*

Center for Molecular Modeling, CIT, National Institutes of Health, Bldg 12A, Bethesda, MD 20892, USA

Received 9 February 2000; received in revised form 10 April 2000

Edited by Gunnar von Heijne

Abstract Involucrin is a key component of the cross-linked envelope of terminally differentiated keratinocytes. The human molecule largely consists of 10 residue repeats and forms a thin 460 Å long rod. Summarized experimental data and a detailed stereochemical analysis made with computer modeling resulted in a structural model for the involucrin molecule. The suggested structure is a left-handed α-helical solenoid built of a tandem array of helix-turn-helix folds. The structure enables us to explain the whole set of experimental data and residue conservations within the repeats. It is ideally suited to serve as a scaffold for cell envelope assembly and proposes a possible mode of the intermolecular interactions of involucrin during cell cornification.

© 2000 Federation of European Biochemical Societies.

Key words: Modeling; Repetitive sequence; Skin; Structural prediction; Three-dimensional

1. Introduction

The surface of the skin consists of a blanket of interlinked corneocytes that function to protect the organism from harmful environmental influences. The protein composition of the cell envelopes (CE) varies between epithelia, however, involucrin seems to be an ubiquitous component of most if not all CE [1,2]. Several lines of evidence suggest that involucrin binds to CE lipid membrane and serves as an initial scaffold component of CE structure onto which other proteins are later added to achieve final stabilization of CE [2]. The main central segment of 68 kDa human involucrin is composed of 39 tandem repeats, each consisting of 10 amino acids [3]. The consensus sequence of the major repeat is Q_1-E_2-G_3-Q_4-L_5-K_6-H_7-L_8-E_9-Q_{10}. This repeating pattern is conserved in involucrins from all higher primates, although the number of tandem repeats varies [4]. The central repetitive segment (397 residues) is flanked by 147 amino acid N-terminal and 41 amino acid C-terminal segments that lack this repeat pattern [3]. In spite of considerable progress in demonstrating that involucrin is an essential molecule of CE [2,5], its three-dimensional (3D) structure is not known. Hydrodynamic and electrophoretic studies have shown that involucrin behaves like an elongated rod in solution [6,7]. Electron microscopic visualization of negatively stained involucrin preparation reveals thin, flexible rods having a contour length of 460 ± 70 Å and an approximate width of 15 Å [8]. Circular dichroism (CD) measurement indicates that 50% of the molecule is α-helical [8]. In addition, there is independent evidence suggesting that these rod-like molecules are monomers [6,7]. To explain these data, Yaffe et al. [8] proposed that most of the molecule consists of α-helices of different length, joined at various angles by flexible segments of random coil. However, this arrangement is open to several questions. Indeed, in this model, clusters of non-polar leucines are exposed to the solvent. In contrast, in the known protein structures, such 'leucine surfaces' usually interact with each other and form a hydrophobic core. Furthermore, the α-helices with flexible connections cannot maintain the rod-like structure; rather they will collapse into a globular structure. In the other suggested arrangement, the central rod-like domain has a mixture of α-helical and 3_{10}-helical conformations [9]. An attractive feature of such an $\alpha/3_{10}$-helix is the clustering of its glutamates on one face and its glutamines on the opposite face, which can interact with CE lipids and proteins, respectively. However, the proposed high helicity of involucrin originated from a CD study of its fragments in a non-native organic solvent [9]. Moreover, this model does not explain the dimensions of the molecule established by electron microscopy [8]. These shortcomings of the previous models encouraged us to undertake another attempt to predict the 3D structure of involucrin.

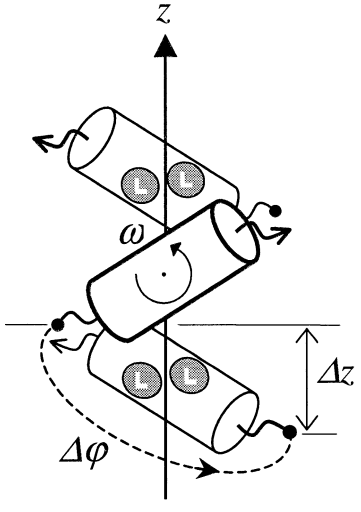


Fig. 1. Scheme of the repetitive multiplication of the 10 residue unit during search of the initial template for the involucrin structure. α-Helical segments are represented by cylinders. The crossing angle ω was used to vary the orientation of the α-helical segments; Δz and $\Delta\phi$ are components of the screw translation. Circles with 'L' on the surface of the α-helices indicate location of leucines 5 and 8.

*Fax: (1)-301-4022867.
E-mail: kajava@helix.nih.gov

Abbreviations: CE, cell envelope; 3D, three-dimensional

0014-5793/00/$20.00 © 2000 Federation of European Biochemical Societies. All rights reserved.
PII: S0014-5793(00)01502-7

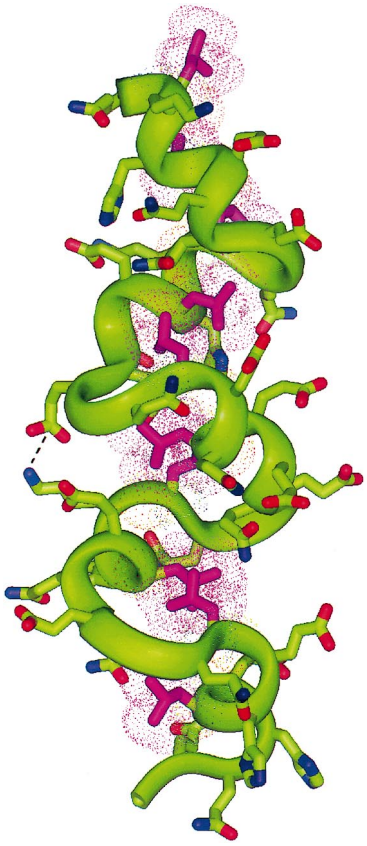
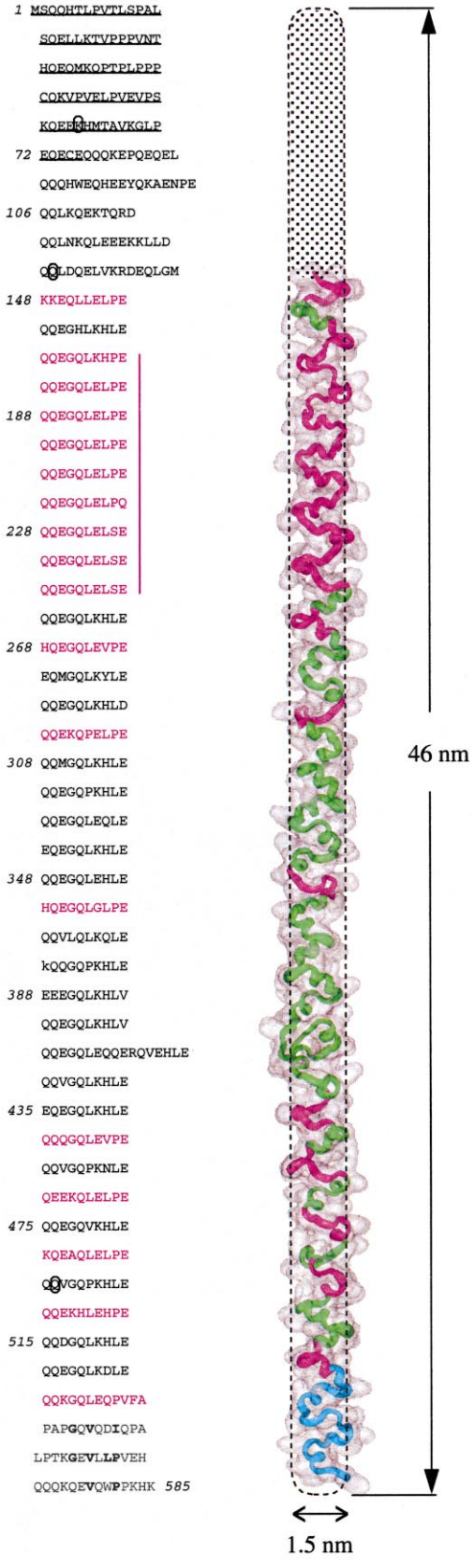


Fig. 2. A structural model for the repetitive fragment (residues 312–361). Backbone is present as a ribbon. Side chains have stick representation. Hydrogen atoms are not shown. Leucines are in purple and surrounded by dotted areas of van der Waals contours. The remaining side chains are colored by atoms (oxygen atoms are in red, carbon atoms in green, nitrogen atoms in blue). A thin dotted line denotes a possible inter-coil ionic bond between Lys-5 and Glu-2. The structure was constructed using the Insight II program [14]. The initial template was obtained by rotation–translation operation with the translation component $\Delta z = 9$ Å, the rotational component $\Delta\phi \approx +160°$ and the crossing angle $\omega \approx +40°$. To find the conformation of the cross-over, the α-helical segments containing sequence L-K-H-L were fixed, while the conformation of the remaining residues was manually varied by rotation of backbone torsion angles. Then, the α-helical segments were covalently connected by appropriate cross-overs. The resulting structures were subjected to energy minimization. During the 300 steps of minimization based on the steepest descent algorithm the backbone atoms of α-helical segments were restrained to their starting positions with force constant $K = 100$. To allay the concern that these constraints generated significant tensions in the minimized structure, the next 500 steps of the refinement were performed without any restrictions, using a conjugate gradients algorithm. The CHARMM force field [15] and the distance-dependent dielectric constant were used for the energy calculations. The program PROCHECK [16] was used to check the quality of the modeled structure. Figs. 2–4 were generated by Insight II [14]. Coordinates for the model of the repetitive domain (residues 312–361) will be deposited with the Brookhaven Protein Data Bank.

2. Formulation of the structural model for the repetitive domain

For globular proteins, structural prediction remains risky in the absence of strong sequence similarity to protein(s) of known structure [10]. However, for proteins with repetitive sequences, a priori predictions can be quite reliable. Examples include fibrous proteins with periodicities of 2–7 residues [11], superhelical proteins with repeats of 20–30 residues [12] and even bead-like proteins with 30–40 residue periodicities [13]. In all cases, predictions have been facilitated by assuming repetitive spatial arrangements within the tandem repeat sequences, by distinguishing conserved – and presumably struc-



turally important – residue positions in the repeats, and by incorporating constraints from indirect lines of evidence. Involucrin fulfills all of these requirements – it is highly repetitive and its shape and secondary structure are known.

In accordance with CD measurements, about 50% of involucrin is α-helical [8]. Assuming that the proline-rich N- and C-termini cannot form α-helical structures, one can conclude that α-helicity of the remaining repetitive segment is more than 50%. Thus, it is safe to assign more than one turn of α-helix for each 10 residue unit. The array of 10 residue repeats suggests a repetitive structure built by equivalent positioning of the 10 residue units. Given these constraints, a number of regular structures were automatically generated by a rotation–translation operation of 10 residue units (Fig. 1). The initial unit was a 10 residue α-helix with Leu-5 and -8 located in the middle of the helix and facing the axis of the molecule. Repetitive multiplication of α-helical segments and their covalent connection led to solenoid structures coiled around non-polar leucines. The length of involucrin [8] provides a range of possible values for the axial 10 residue unit pitch (Δz). If the repetitive region contributed solely to the molecular length, the pitch would be 11.8 Å (460 Å/39 repeats). If the N- and C-terminal segments comprising 32% of residues also made a proportional contribution to the length of the molecule, the pitch would be 8.0 Å calculated as $(1-0.32)\times 460$ Å/39 repeats. The analysis of the generated solenoids showed that the greater the pitch, the greater the exposure of non-polar leucines to the solvent. This suggests that the actual pitch is closer to 8.0 Å than to 11.8 Å. Therefore, for modeling of the involucrin structure the axial rise per 10 residue unit (Δz) was taken to be 9 Å, while the rotational component ($\Delta\phi$) and the crossing angle between the molecular axis and the α-helix (ω) were varied. The resulting structures were tested for the capability of the α-helices to be covalently connected to each other by unwinding up to five residues out of each α-helical segment and for close packing of the leucine core.

The conservation of Gly-3 strongly suggests that this residue belongs to the cross-over between the α-helices. Otherwise, being in the α-helix, this glycine would be directed toward the solution, which makes its conservation difficult to explain. This indicates that the α-helix starts somewhere after Gly-3, giving a clue to the direction of the cross-over entry into the α-helix. This constraint delimited a set of possible solenoid structures. Further analysis revealed that structures with α-helices oriented orthogonal to the molecular axis ($\omega\approx 90°$) are unlikely due to a loose inter-helical packing, in which the leucines do not contact each other. On the other hand, alignment of the α-helices along the molecule ($\omega\approx 0°$) hampers their covalent connection.

Ultimately, our analysis led to a unique structural solution, which fits stereochemical constraints and explains the residue conservations within the repeats. It turned out that the α-helices need to be tilted at $\omega\sim +40°$ to obtain feasible inter-helical cross-overs (Fig. 2). In this arrangement, each repeat Q-E-G-Q-L-K-H-L-E-Q has a β-β-ε-β-α-α-α-α-α-α conformation, with the symbols α, β and ε denoting residue backbone conformations close to the α-helical, β-structural conformations, and a conformation which is mirror-symmetrical to β, respectively. Leu-5 forms inter-helical contact with Leu-8, forming a longitudinal non-polar core of the solenoid. The size and geometry of leucines are perfect for such packing and this may explain their conservation. Another conserved residue, Gly-3, has a glycine-specific conformation from the lower right quadrant of the Ramachandran plot. At the same time, the cross-over between the helices tightly embraces the apolar core of the structure and does not leave much room for other side chains but glycine. Together, these observations explain the glycine conservation in the sequence. Furthermore, one position of the involucrin repeat is frequently occupied by positively charged lysine and two others have negatively charged glutamates. In the suggested structure, the lysine can form stabilizing inter-helical ionic bonds with the glutamates (Fig. 2). It is known that involucrin serves as the major glutamyl donor in the transglutaminase-catalyzed cross-linking reaction [2,17,18]. In agreement with this, the solenoid structure possesses extended surface areas with Gln-1, -4, and -10 directed to the solvent that are ideally suited to form maximum cross-links with CE proteins [2] and lipids [19].

In the suggested structure, the adjacent α-helices are packed approximately crosswise, forming a so-called helix-turn-helix, or α-α-corner fold [20,21], which is widespread in proteins. The overall structure of involucrin represents a long left-handed solenoid. The rod-like shape of involucrin implies that the inter-helical connections should be relatively rigid. The α-α-corners may provide such rigidity. It is known that some small proteins and domains are merely composed of an α-α-corner and short irregular 'tails' [22]. This suggests that α-α-corner represents a stable kind of fold which can adopt its unique structure per se. An important point is that the handedness of α-α-corners of the chosen involucrin structure is the same as the unique handedness of this fold in the known protein structures [21].

3. Proline-containing repeats

About 30% of the repeats from the higher primate involucrins have another consensus sequence Q-E-G-Q-L-E-L-P-E-Q with glutamate instead of lysine in position 6, leucine in-

←

Fig. 3. Left: The complete amino acid sequence of human involucrin. A part of the head domain which is conserved in the known involucrins is underlined. The sequence of the head domain is subdivided by about 14 residue fragments, which are aligned in such a manner that the first two positions of each fragment have polar residues and the next positions are enriched in prolines and non-polar residues. These covert repeats, being folded in the solenoid, may explain an elongated shape of the head domain. The central domain (148–546) has 39 copies of aligned 10 residue repeats. The proline-containing repeats are in purple. The negatively charged segment which is specific for the human sequence is denoted by a vertical purple line. The tail domain (blue) is inferred to be made up of covert copies of the 10 residue repeats, although its length and sequence appear to be less well conserved. Putative structurally important residues of the tail domain are in bold. The major cross-linking sites (Lys-62, Gln-133 and Gln-496), which are discussed in the text, are encircled. Right: Ribbon representation of the involucrin structure drawn to scale using the coordinates of the modeled structure. The molecule is contoured by solvent-accessible surface. The parts of the structure with ambiguously predicted polypeptide conformations are in purple (proline-containing repeats) and in blue (tail domain). The dotted contour line over the structure shows the involucrin dimensions determined by electron microscopy. A detailed structure is not proposed for the head domain and for the purpose of illustration the part assigned to it is filled by a dotted pattern.

stead of histidine in position 7 and proline instead of leucine in position 8. Usually, single copies of such proline-containing repeats are interspersed with segments containing several typical repeats. Furthermore, human involucrin has nine additional copies of this type of repeat arranged in tandem, which probably are of comparatively recent evolutionary origin [3]. The rod-like shape of human involucrin [8] and the proposed solenoid topology in the typical repetitive regions suggest that its structure should consist of an unbroken run of solenoid coils, even in the places where the typical repeats are interrupted by the proline-containing one. On the other hand, the conformation of the proline-containing coils should be different from the typical one due to the fact that occurrence of proline in position 8 should disfavor the α-helix (residues 5–10) proposed for the typical repeat. Our stereochemical analysis suggests that the proline-containing repeat has segments with extended conformations alternating with short α-helices or β-turns. In general, these conformations are more irregular and ambiguously predicted compared with the conformation suggested for the typical repeat. For the purpose of illustration (Fig. 3), we chose one of these possible conformations, α-α-α-β-α-β-β-β-α-α. Such units form a non-polar core in the solenoid arrangement and have close inter-coil packing with each other and with the typical units. The nine proline-containing tandem repeats at the N-terminal part of the repetitive segment of human involucrin have a strong negative charge (−3 per 10 residue repeat). This suggests that inter-coil interactions of these repeats may not be sufficient to give them a stable structure on their own. Stabilization may require positively charged cations, as was observed in the β-helical solenoid of seralysins [23] which use stacks of Ca^{2+} ions for structure stabilization. On the other hand, the polar heads of CE lipids may complete the formation of these Ca^{2+} binding sites and this may explain the observed Ca^{2+}-dependent binding of involucrin to the native-like synthetic lipid vesicles [24]. Remarkably, the other four proline-containing repeats alternate with typical ones at the C-terminal part (residues 445–505). In the proposed model, this region can adopt a structure with the proline-containing units located on one side. This region may be another site with membrane binding potential.

4. The overall structure of the central repetitive domain of involucrin

Fig. 3 shows the molecular model of the overall structure of the central repetitive segment of human involucrin. The model is ∼55% α-helical and agrees well with the CD data. The solenoid has a 9 Å axial rise per 10 residue unit and predicts that the repetitive domain contributes most (76%) of the estimated 460 Å length of human involucrin. Furthermore, the width of the solenoid varies between 14 and 19 Å and fits the electron microscopy estimation. Comparison of the dimension of the modeled repetitive domain with the known length of the whole molecule predicts that the head and tail domain structures should also be elongated. Indeed, these domains, especially the tail, like the central domain, have repeats, however of longer and more variable length and less conserved sequence (Fig. 3). The elongated shape and repetitive patterns suggest that the head and tail domains may also have a solenoid arrangement.

In contrast to the previous models [8,9], the present arrangement explains how such a thin and long molecule made of a single chain can form a non-polar core. However, the evaluation of the model with the VADAR program [25] suggests that the non-polar core is not buried enough to stabilize the structure when the solenoid has a 9 Å pitch. In this case, fractional accessible surface areas (ASA) of Leu-5 and -8 are 19% and 11%, respectively, while, in accordance with VADAR, these leucines should have fractional ASAs less than 5% in a stable structure. The compactness predicted for the stable structure can be reached after extensive energy minimization (1500 steps) without applying the 9 Å constraint on the axial pitch. The length of the resultant structure is about 420 Å (∼7.3 Å pitch). This value still matches the range of lengths 460 ± 70 Å estimated by electron microscopy [8].

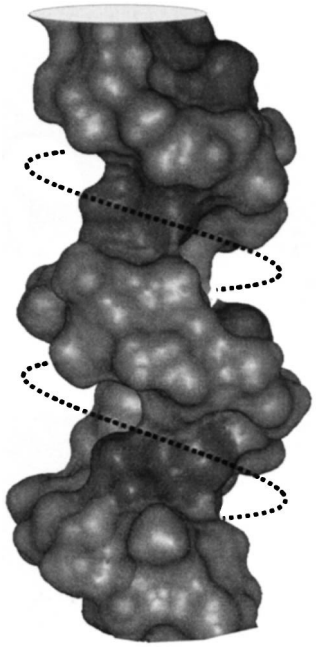


Fig. 4. Diagram showing the solvent-accessible surface of a fragment of the repetitive domain (residues 312–361) of human involucrin. A dotted line winding around the involucrin groove denotes a possible mode of peptide binding. For simplicity, only C_β atoms of the outside side chains are shown.

5. The surface groove of involucrin as a potential docking site for linear peptides

If one looks at the surface of the modeled structure of involucrin, an interesting feature stands out: a groove twines around the molecule (Fig. 4). A peptide in an extended conformation fits well in the groove of the modeled involucrin. The bottom of this groove is partially hydrophobic because it is mostly built by that part of the leucines which is free from intramolecular contacts. The hydrophobicity of the groove may favor binding of an extended polypeptide chain enriched in non-polar residues. It is known that involucrin molecules are cross-linked between each other during CE assembly [2]. Moreover, primarily Lys-62 of the head domain is initially used for cross-linking with Gln-133 and Gln-496 [26]. Therefore, the head domain may have a tendency to non-covalently associate with another involucrin molecule prior to its enzymatic cross-linking. Our analysis revealed that the head domain sequence has relatively hydrophobic and proline-rich

regions alternating with hydrophilic areas (Fig. 3). One may suggest that the hydrophobic regions of the head domain may have an extended conformation and bind the groove of the central domain of another involucrin.

6. Conclusion

The model for human involucrin described here is the first insight into the 3D atomic structure of the CE meshwork. This structure can be added to a collection of proteins with a solenoid topology, which emerged relatively recently (for review see [27,28]). The model suggests the sites and modes of interaction of involucrin molecules with each other and with CE lipids. These predictions can be tested experimentally. Moreover, the great importance of involucrin hopefully will lead to the experimental determination of its structure. This in turn will provide a chance to assess the validity of the model and the correctness of our understanding of protein structures.

Acknowledgements: I am grateful to U. Blum-Tirouvanziam, A.E. Kalinin, P.J. Steinbach, P.M. Steinert and A.S. Steven for reading the manuscript, for discussion and for criticism.

References

[1] Steven, A.C. and Steinert, P.M. (1994) J. Cell Sci. 107 (Pt 2), 693–700.
[2] Steinert, P.M. and Marekov, L.N. (1997) J. Biol. Chem. 272, 2021–2030.
[3] Eckert, R.L. and Green, H. (1986) Cell 46, 583–589.
[4] Tseng, H. and Green, H. (1988) Cell 54, 491–496.
[5] Eckert, R.L., Yaffe, M.B., Crish, J.F., Murthy, S., Rorke, E.A. and Welter, J.F. (1993) J. Invest. Dermatol. 100, 613–617.
[6] Rice, R.H. and Green, H. (1979) Cell 18, 681–694.
[7] Etoh, Y., Simon, M. and Green, H. (1986) Biochem. Biophys. Res. Commun. 136, 51–56.
[8] Yaffe, M.B., Beegen, H. and Eckert, R.L. (1992) J. Biol. Chem. 267, 12233–12238.
[9] Lazo, N.D. and Downing, D.T. (1999) J. Biol. Chem. 274, 37340–37344.
[10] Moult, J., Hubbard, T., Bryant, S.H., Fidelis, K. and Pedersen, J.T. (1997) Proteins 1 (Suppl.), 2–6.
[11] Fraser, R.D.B. and MacRae, T.P. (1973) Conformation in Fibrous Proteins and Related Synthetic Polypeptides, Academic Press, London.
[12] Kajava, A.V., Vassart, G. and Wodak, S.J. (1995) Structure 3, 867–877.
[13] Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4, 1609–1614.
[14] Dayring, H.E., Tramonato, A., Sprang, S.R. and Fletterick, R.J. (1986) J. Mol. Graph. 4, 82–87.
[15] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) J. Comp. Chem. 4, 187–217.
[16] Laskowski, R.A., McArthur, Moss, D.S. and Thornton, J.M. (1993) J. Appl. Crystallogr. 26, 282–291.
[17] Simon, M. and Green, H. (1985) Cell 40, 677–683.
[18] Rorke, E.A. and Eckert, R.L. (1991) J. Invest. Dermatol. 97, 543–548.
[19] Nemes, Z., Marekov, L.N., Fesus, L. and Steinert, P.M. (1999) Proc. Natl. Acad. Sci. USA 96, 8402–8407.
[20] Pabo, C.O. and Sauer, R.T. (1984) Annu. Rev. Biochem. 53, 293–321.
[21] Efimov, A.V. (1984) FEBS Lett. 166, 33–38.
[22] Efimov, A.V. (1994) FEBS Lett. 355, 213–219.
[23] Baumann, U., Wu, S., Flaherty, K.M. and McKay, D.B. (1993) EMBO J. 12, 3357–3364.
[24] Nemes, Z., Marekov, L.N. and Steinert, P.M. (1999) J. Biol. Chem. 274, 11013–11021.
[25] Wishart, D.S., Willard, L. Richards, F.M. and Sykes, B.D. (1995) VADAR: Version 1.3, University of Alberta, Edmonton, Alta.
[26] Steinert, P.M. and Marekov, L.N. (1999) Mol. Biol. Cell 10, 4247–4261.
[27] Kobe, B. (1996) Nature Struct. Biol. 3, 977–980.
[28] Groves, M.R. and Barford, D. (1999) Curr. Opin. Struct. Biol. 9, 383–389.