J Mol Model (2007) 13:233–246
DOI 10.1007/s00894-006-0144-9

ORIGINAL PAPER

A three-dimensional model of the human transglutaminase 1: insights into the understanding of lamellar ichthyosis

Karen M. Boeshans · Timothy C. Mueser · Bijan Ahvazi

Received: 23 March 2006 / Accepted: 11 June 2006 / Published online: 23 September 2006

Abstract The *stratum corneum*, the outer layer of the epidermis, serves as a protective barrier to isolate the skin from the external environment. Keratinocyte transglutaminase 1 (TGase 1) catalyzes amide crosslinking between glutamine and lysine residues on precursor proteins forming the impermeable layers of the epidermal cell envelopes (CE), the highly insoluble membranous structures of the *stratum corneum*. Patients with the autosomal recessive skin disorder lamellar ichthyosis (LI) appear to have deficient cross-linking of the cell envelope due to mutations identified in TGase 1, linking this enzyme to LI. In the absence of a crystal structure, molecular modeling was used to generate the structure of TGase 1. We have mapped the known mutations of TGase 1 from our survey obtained from a search of PubMed and successfully predicted the impact of these mutations on LI. Furthermore, we have identified Ca^{2+} binding sites and propose that Ca^{2+} induces a *cis* to *trans* isomerization in residues near the active site as part of the enzyme transamidation activation. Docking experiments suggest that substrate binding subsequently induces the reverse *cis* to *trans* isomerization, which may be a significant part of the catalytic process. These results give an interpretation at the molecular level of previously reported mutations and lead to further insights into the structural model of TGase 1, providing a new basis for understanding LI.

Keywords Keratinocyte transglutaminase 1 · Lamellar ichthyosis · Mutations · Metal ions · Isomerization · Molecular modeling

Abbreviations

TGase 1	keratinocyte transglutaminase 1
LI	lamellar ichthyosis
GTP	guanosine 5′-triphosphate
RMSD	root mean square deviation
DTT	1,4-dithio-DL-threitol
EDTA	ethylenediaminetetraacetic acid
TCA	trichloroacetic acid
CE	cell envelope
SPRs	small proline-rich proteins
SCRs	structurally conserved regions
SVRs	structurally variable regions
MolCad	computer aided molecular design
CTI	*cis* to *trans* isomerization

Introduction

Autosomal recessive congenital ichthyoses (ARCI) are a clinically and genetically heterogeneous group of disorders

This paper is dedicated to the memory of Dr. Peter M. Steinert (April 7, 2003).

K. M. Boeshans · B. Ahvazi
X-ray Crystallography Facility/Office of Science and Technology, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892-8024, USA

T. C. Mueser
Department of Chemistry, University of Toledo, Toledo, OH 43606, USA

B. Ahvazi (✉)
X-ray Crystallography Facility, NIAMS, MSC 8024, 50 South Dr., Building 50, Room 1345, Bethesda, MD 20892-8024, USA
e-mail: ahvazib@mail.nih.gov

with severe ichthyosis of the skin at birth without manifestations in other organ systems. The two major phenotypes, congenital ichthyosiform erythroderma (CIE) and lamellar ichthyosis (LI) are distinguished depending on the grade of erythema [1]. The LI is often present at birth in the form of a collodion membrane covering the neonate consisting of shiny taut skin that eventually dries and peels away, at which point large, thick brown scales develop [2, 3]. Patients with LI can be either homozygous for a single mutation or heterozygous for two different mutations. LI has been mapped to chromosome 14q11 [4, 5] and subsequently to the transglutaminase 1 (TGase 1) gene, TGM1, at position 14q11.2 in the majority of families with LI [6, 7]. Mutations in the TGase 1 gene produce functionally defective enzyme due to truncation and/or point substitutions located at a number of sites throughout the protein. A variety of mutations in the TGase 1 gene in the most severe phenotypes are associated with very low or undetectable enzyme activity [7–9].

Although the overall primary structure of TGase enzymes appear to be quite different, they all share a common amino-acid-sequence motif at the active site and an absolute requirement of calcium for their activity. The differences in the primary structure of TGases are most likely responsible for their diverse biological functions. In addition, among the mammalian TGases characterized thus far, the number of calcium ions, their location for activation, and nucleotide binding, if any, still remain elusive. The crystal structures of TGase 2 and TGase 3 show that they bind and hydrolyze GTP to GDP [10, 11]. Surprisingly, the amino-acid side chains that lie in the GDP binding pocket of TGase 2 are not precisely conserved in TGase 3. The crystallographic studies of TGase 3.GTPγS/GDP complexes have therefore challenged the long-standing notion that among all TGases, only TGase 2 binds GTP, and hydrolyzes it to GDP [11]. This provides further support for nucleotide binding by other TGase members. Interestingly, sequence alignment indicates that each enzyme possesses three *cis* peptide bonds. The role of these *cis*-peptide bonds is yet to be determined.

TGase 1, a 90 kDa membrane-bound enzyme, is the largest of the nine known TGases in the human genome. TGase 1 is synthesized as an 817-residue polypeptide and is modified by myristoyl or palmitoyl adducts near the *N*-terminus of the protein. It is expressed in the stratified *squamous epithelium*, predominantly in the upper spinous and granular layers beneath the *stratum corneum*. Reports of deficient cell envelopes in cases of LI indicate that TGase 1 plays an important role in the formation of the cell envelope (CE). TGase 1, like other TGases, is a Ca^{2+}-dependent enzyme, which catalyzes the transfer of the γ-carboxy amide group from protein-bound glutamine residues to the ε-amino group of protein-bound lysine residues or other primary amines [12]. This reaction results in the formation of N^{ε}-(γ-glutamyl)lysine isopeptide cross-links between (and within) polypeptide substrates, giving rise to stable protein polymers. The most abundant substrates in this cross-linking process are loricrin, small proline-rich proteins (SPRs), and involucrin, proteins that are rich in lysine and glutamine and function primarily as CE components. The full length TGase 1 apoprotein is inactive. TGase 1 is proteolyzed at two sites during maturation and enzymatic activity becomes upregulated as a result of epidermal differentiation. The three components remain associated in the active membrane-bound form. In vitro, the enzyme is fully activated by Ca^{2+} ions after proteolysis with the protease Dispase I, but the protease(s) responsible for the activation in vivo are still unknown. TGase 1 is anchored to the cytoplasmic side of the plasma membrane through thioester-linked palmitic and myristic acid [13, 14]. Fatty acylation sites are localized in a cluster of five cysteine residues in the extended *N*-terminus of TGase 1, the membrane anchorage region [15, 16]. Although TGase 1 deletion is lethal in knock-out mice, patients with LI, including those with undetectable TGase 1, live into adulthood with a severe scaling dermatosis and significant permeability barrier abnormalities [17].

While genotype and phenotype correlation in LI have been studied for several years, the exact nature of the relationship has yet to be elucidated fully [18]. In particular, the lack of crystal structure has hindered the dissection of TGase 1 functional domains, thus impeding any prediction of the impact of each mutation on LI. It is not yet clear how the mutations affect the structure and function of TGase 1. In the absence of a suitable X-ray crystal structure, comparative protein modeling based on sequence similarity among various proteins, offers a reasonable alternative for elucidation of TGase 1 structure. In the work presented here, we have used molecular modeling, bioinformatics and biocomputing approaches to develop a suitable structure of the TGase 1 enzyme. The inclusion modes of the TGase 1 enzyme with the peptides SQQ*VT from loricrin for the Gln* substrate and KTKQK* from small proline rich protein 1 as the Lys* substrate were determined. This study further provides an understanding of how peptide chains bearing reactive Gln* and Lys* residues may complex with TGase 1. The modeled structure has been validated using several structure/geometry verification tools. Docking studies of substrates and cofactors also provide insight into the possible binding modes leading to protein–protein interactions and catalysis. The results of these studies may help explain some of the biochemical abnormalities leading to LI.

Materials and methods

TGase 1 expression and activity assay

The human TGase 1 protein was expressed in *baculovirus*. cDNA was obtained by the reverse transcriptase-PCR method using mRNA isolated from normal human epidermal keratinocytes. A full-length gene was constructed in the pET11a vector (Novagen) and the nucleotide sequence was confirmed by automated sequencing. To generate recombinant virus, the transfer vector pVL1392 (Invitrogen), which contains a complete polyhedrin gene promoter for foreign gene expression, was used. The TGase 1 cDNA was ligated into pVL1392, and the resulting plasmid was named pVL-TG1. This was cotransfected with linearized viral DNA (Bac-N-Blue, Invitrogen) into sf9 cells by use of liposomes (Lipofectamine, LifeTechnologies). After plaque purification, a single clone, named vAcVL-TG1, was used to infect sf9 cells grown to a density of 13×10^6 cells/ml. The cells were incubated at 27 °C for 3 days after infection and harvested to generate a high-titer virus stock. This high-titer suspension was stored at 4 °C. For large-scale expression of TGase 1 protein, sf9 cells were grown to a density of 23×10^6 cells/ml, infected by the high-titer virus stock at a multiplicity of infection of 2, and grown for 3 days at 27 °C.

TGase activity of cell extract was measured by incorporation of [^{14}C] putrescine into casein. Assays were done at 37 °C in 0.5 ml aliquots of 100 mM Tris-HCl (pH 7.5), containing 1% N, *N*-dimethylated casein (Sigma), 0.5 μCi putrescine (118 mCi/mmol), 1 mM DTT, 5 mM $CaCl_2$ and 1 mM EDTA. Reaction blanks were identical except they contained no added calcium. Reactions were stopped with the addition of ice-cold 7% (w/v) TCA. Incorporated radioactivity was collected by filtration through GF/A glass fiber filters (Whatman) wetted with ice-cold 5% (w/v) TCA and then successively washed three times with 5 ml of ice-cold 5% (w/v) TCA. Filters were dried and solubilized in 5 ml of CytoScint (ICN). Inhibition assays contained 0–500 μM GTP at 0.5 mM $CaCl_2$ concentration.

Model construction and docking

Computational studies were performed on a Silicon Graphics Octane 2 workstation, equipped with dual 360 MHz R12000 processors. The sequences of TGase 1 were obtained from GenBank®. Only the mature sequences starting from residues Pro^{24} in TGase 1, as per the numbering in the complete sequence, were considered in deriving the homology models. A WUBLAST 2.0 [19] search and the Unity module of Sybyl 6.9 (Tripos Associates Inc., St. Louis, MO) were performed on the sequences. The Psi pred protein structure prediction server was used for the secondary structure prediction [20, 21]. Crystal structures from various sources with sequence identity of more than 30% (i.e. TGase 2 and TGase 3 with a significance score of 21 or more), and the model of TGase 5 were chosen to build the TGase 1 model structure. Comparative protein modeling was performed with the Composer module of Sybyl 6.9. Energy minimizations and molecular dynamics were accomplished in the Discover module of Insight II 2000 (Accelrys Inc., San Diego, CA). The geometrical and local environmental consistencies of the models were evaluated with the ProStat and Profiles-3D [22] modules of Insight II 2000 and the MatchMaker module of Sybyl 6.9. The alignment and identification of the structurally conserved regions (SCRs) and structurally variable regions (SVRs) was carried out in an analogous manner as described previously [23]. Coordinates for the SCRs were assigned from the template molecules. The coordinates of SVRs were obtained using the conformational search program GeneFold and MatchMaker modules. The SVR loop regions were built using the protein loop search protocol [24] as available in Composer. The model obtained was then refined by minimization using a similar protocol as described earlier [23].

Minor adjustments of the TGase 1 sequence alignment from residues Val^{110}–Pro^{817} were made to ensure correct positioning of all conserved residues. The appropriate loop models were selected on the basis of acceptable stereochemical parameters (allowable ϕ and ψ angles on a Ramachandran plot). The *N*-terminus of TGase 1 could not be modeled from the other TGase models. Instead, short segments were modeled based on homology with other known structures where possible. There was no template available for Met^1–Ser ^{23}. The short segment of Pro^{24} to Ala^{55} had a 42% sequence similarity to a segment in the Cd-6 metallothionein (PDB: 1dmf) and was modeled accordingly. The residues Asp^{66}–Leu^{109} were modeled based on human Fascin molecule (PDB: 1dfc). The identity score of 50.9% including significance score of 4.4 was obtained by Composer module for this segment.

Substrates were docked into the model of TGase 1 using an approach described previously [23]. Briefly, the substrates were placed in the binding pocket surrounding the catalytic core, and the model was energetically minimized with 500 steps of steepest decent minimization, followed by 500 steps of conjugate gradient minimization to remove the geometrical strain. Residues within 6 Å of the catalytic triad were allowed to move freely while the remaining atoms of the enzyme were held rigid during the docking process. The solvation model was used to introduce the effect of solvent implicitly [25]. In the first stage, the Monte Carlo minimization method was used to generate 20 distinct docking poses (minimum RMS distance of 1 Å) by random

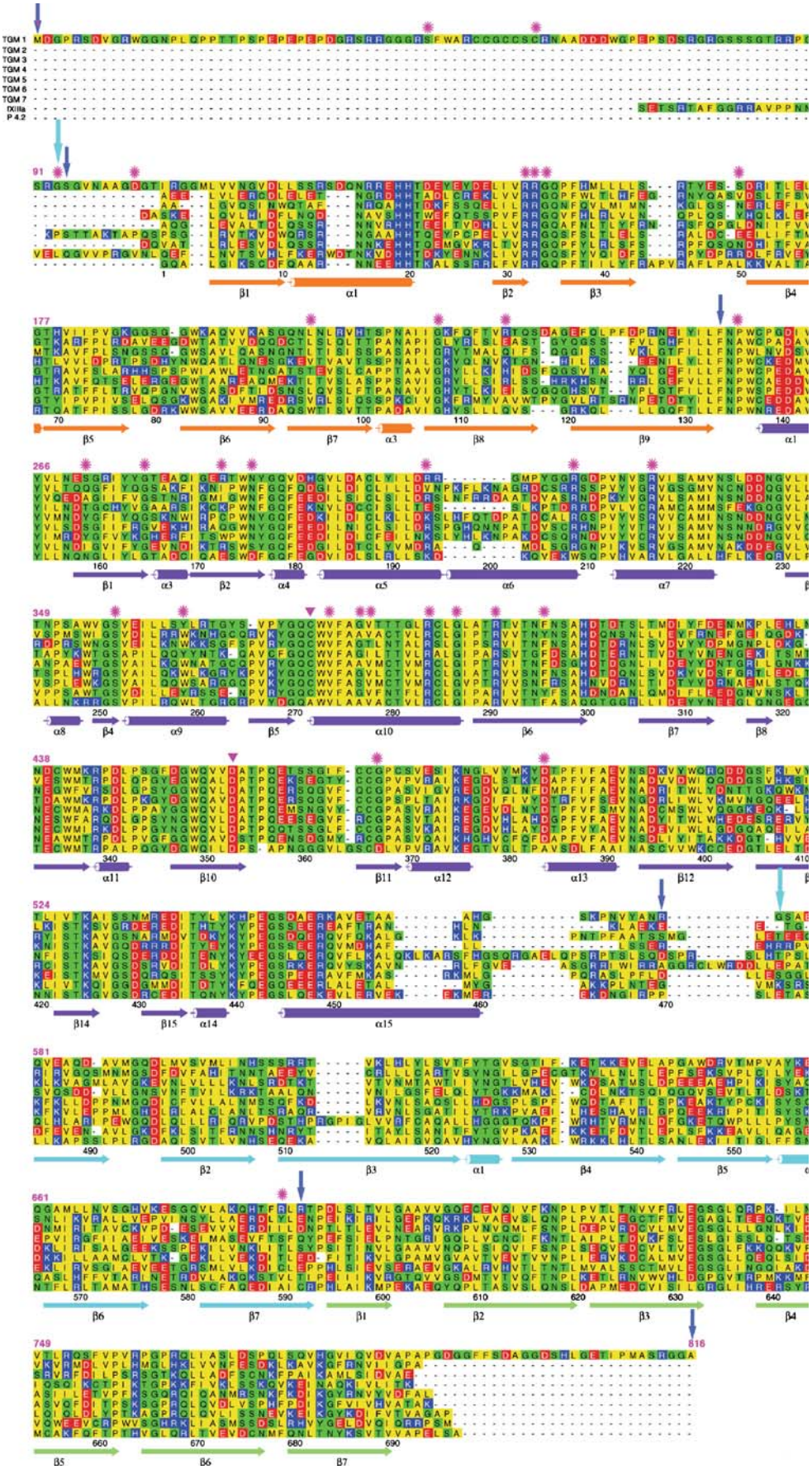

Fig. 1 Sequence alignment of TGase family enzymes in the human genome. The catalytic triad Cys^{376}, His^{435} and Asp^{458} is indicated by a triangle in *magenta*. The cleavage sites are shown with an arrow (*cyan*) at residues Gly^{93} and Gly^{573}. The amino acids highlighted in *red* are acidic (D, E), in *blue* are basic (H, K, R), in *yellow* are nonpolar (A, I, L, V, P, F, W, M), and in *green* are polar (G, S, T, Y, C, N, Q). The sequence numbering and secondary structure assignment is based on TGase 3 in the bottom (*black*) and corresponding to sequence numbering for TGase 1 on the top (*magenta*). The secondary structure domains are the β-sandwich (*orange*), the catalytic core (*purple*), the barrel 1 (*cyan*), and barrel 2 (*green*), and are separated by an arrow (*blue*). The 34 unique amino acid mutations (stars in *magenta*) in families with lamellar ichthyosis are indicated

combinations of translational and rotational motions each followed by 2,500 iterations of conjugate gradient minimization. During this stage, a purely repulsive bonded quartic potential was used for modeling the van der Waals interactions while the coulombic interactions were set to zero. In the second stage, the structures were subjected to molecular dynamics simulation for 5,000 iterations of equilibration at 300 K. The non-bonded interactions were calculated using a more refined Cell-multipole method [26] with a distance-dependent dielectric constant. Newton's equations of motion were integrated using the Verlet algorithm [27] with a time step of 1 fs using the NVT ensemble. Temperature was controlled via direct scaling of the atom velocity motion. Finally, the structures were judged complete when gradient steps of 0.001 kcal mol^{-1} $Å^{-1}$ or less were reached in conjugate gradient minimization. The various docking poses thus obtained were evaluated based on force-field-based interaction energies between the substrate and the enzyme. The interaction energy was used as a scoring function to rank the docking poses and the one with the lowest interaction energy was selected for further analysis. Further validation of the structures is provided by analysis of the stereochemistry of the structure with the programs Procheck [28]. The topography and electrostatic potential distribution of the molecular surface was characterized using the MolCad (computer aided molecular design) program. The visualization of spatial surface was performed through Connolly's program (designated as Connolly surfaces). MolCad was also used to generate electrostatic potential maps on the molecular surface. Figures were generated with the Molscript [29], Raster3D [30], and Sybyl [31] programs.

Results

The keratinocyte TGase 1 architecture

The crystal structure of TGase 3 was used as a homology model to generate a structure for the TGase 1 enzyme. Sequence alignments of the TGase family share more than 50% sequence homology, indicating that structural similarity is very likely (Fig. 1). The TGase 1 apoprotein is 816 residues in length. The two proteolytic cleavage sites have been identified, after residues Gly^{93} and Gly^{573}, which are associated with TGase 1 maturation [12]. The predicted model (Fig. 2a) for TGase 1 consists of five domains. The *N*-terminal residues Met^{1} to Arg^{92} contain the 10 kDa membrane anchorage region. This segment is unique to the TGase 1 enzyme and modeling was completed using homology to segments of Cd-6 metallothionein and human fascin. The β-sandwich domain (Ser^{94}–Phe^{246}), and catalytic core domain (Asn^{247}–Arg^{572}) comprise the 67 kDa fragment. The β-barrel 1 (Gly^{573}–Arg^{688}) and β-barrel 2 domains (Thr^{689}–Ala^{816}) comprise the 33 kDa fragment. These four domains have high similarity to the TGase 3 enzyme. The core domain harbors the active site residues Cys^{376}, His^{435}, and Asp^{458}. The refined TGase 1 model, like TGase 3, appears to have a shielded catalytic triad, which is buried within a narrow cleft with walls formed by two β-sheet strands of the catalytic core domain and the carboxyl-terminus of β-barrel 1 domain. Based on sequence homology, the core domain of TGase 3 possesses three Ca^{2+} binding sites 1, 2, and 3, respectively, which appear also to be present in the TGase 1 structure, as shown in Fig. 2a [32–34].

The TGase 1 model was subjected to energy minimization refinement to eliminate any steric conflicts. The superposition of Cα backbone atoms of the refined structures of TGase 1 and TGase 3 have an RMSD of 2.1 Å with the obvious exception of the membrane anchorage region (Fig. 2b). The structure was assessed for its geometrical and environmental consistencies. The energy-minimization refinement provided the allowed values for bond lengths, bond angles, amide torsions, ϕ, and ψ torsions for helices, ϕ torsions for proline residues and side chain torsions. The Ramachandran plot revealed a normal distribution of the ϕ and ψ torsions of the residues occupying the allowed region, indicating that no major distortion was introduced to eliminate steric conflicts. MatchMaker analysis also gave a good score for TGase 1, suggestive of reasonable 3D-models. Examples of the MatchMaker output include a histogram of sequence-structural compatibility, a sub-optimal alignment plot and a prediction of structural compatibility between the sequences which are given in Fig. 2b. The *Profiles-3D* analysis of the structure provided an overall self-compatibility score, again indicating a correct structure model as indicated in Fig. 2c.

Among all the known crystal structures of TGase enzymes, only TGase 2 and TGase 3 have been shown to bind and hydrolyze GTP [10, 11, 35]. Interestingly, the GDP-binding pockets of TGase 3 and TGase 2 are related to each other in their position and in several key amino-acid determinants, but these pockets are only superficially related to those characterized in G-proteins. Homology modeling suggests that TGase 1, TGase 5, and TGase 7 may also be able to bind to GTP/GDP (Figs. 1 and 3a). The effect of GTP on human TGase 1 activity was examined. The activities of TGase 2 and TGase 3 are specifically inhibited by the presence of 20–100 μM GTP, but no inhibition was observed in the TGase 1 activity (Fig. 3b). These results suggest that, while the pocket may exist for GTP nucleotide to bind to TGase 1, the GTP was not involved in the inhibitory regulation of TGase 1 activity. It is not clear from these studies if, (1) GTP affinity may be

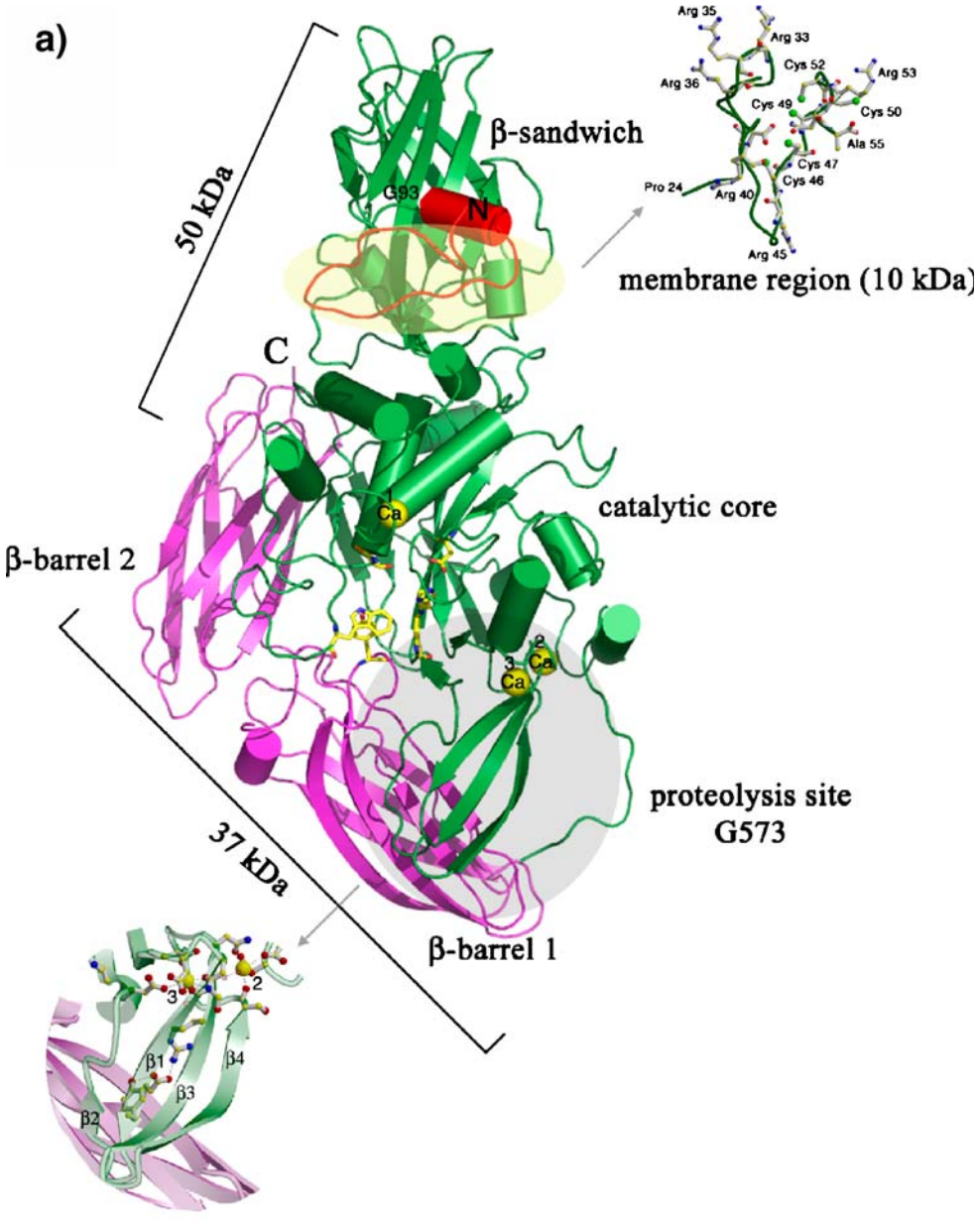

Fig. 2 Ribbon image of the model of human TGase 1 structure. **a** The two domains β-sandwich (*green*) and the catalytic core (*green*) comprises the 67 kDa fragment. The β-barrel 1, and β-barrel 2 (*magneta*) account for the 33 kDa fragment. The Ca^{2+} ions are shown in *yellow* and are numbered according to their sites. The fragment of the membrane anchorage region (residues Asp^{66}–Leu^{109}) was modeled (*red*) and the proteolysis sites are shown. The secondary structure comparison for the Pro^{24}–Ala^{55} residue region is shown separately. The side chain atoms of the catalytic triad residues Cys^{376}, His^{435}, and Asp^{458}, coordination of the three Ca^{2+} ions, and side chain atoms of Trp^{341}, Trp^{432}, and Tyr^{620} are shown in ball-and-stick. The inset of two catalytic Ca^{2+} ions at site 2, and 3, which hold the cluster, and form a flap motif is shown. **b** Examples of MatchMaker output, including a histogram of sequence-structural compatibility (*upper left*), a sub-optimal alignment plot (*upper right*), and a prediction of structural compatibility between the sequence of human TGase 1 (*forest green*) and TGase 3 (*purple*). **c** The figure represents a superposition of the ribbon image of the human TGase 5 (*yellow*), TGase 3 (*green*), and TGase 2 (*purple*) enzyme structures used to model TGase 1

diminished, (2) GTP binds with similar affinity without any effect or (3) the GTP pocket may be modified to have affinity for other inhibitors in TGase 1.

Calcium regulation in TGase 1

Calcium is an important physiological regulator of epithelial differentiation in vivo and in vitro. [36] The main evidence that calcium plays a role in vivo is the presence of a Ca^{2+} gradient in the epidermis that increases from the basal to the granular cell layer. High calcium concentrations are necessary for stratification and desmosome assembly [37]. In vitro, epidermal keratinocytes grown in low calcium concentrations continue to proliferate and lack desmosomes, but when shifted to media containing higher calcium concentrations greater than 0.1 mM they express differentiation markers such as involucrin, K1, K10, loricrin and profilaggrin. It is clear that the enzymatic activities of TGase 1 are controlled in a similar fashion to TGase 3 by local concentrations of Ca^{2+} and Mg^{2+} ions [38]. Under physiological conditions, intracellular free Ca^{2+} ion levels around 100 nM, and free Mg^{2+} concentrations in the low mM range, are sufficient to keep TGase 1 in a latent state. One would expect that, even after the proteolysis of TGase 1, the enzyme would display minimal activity intracellularly due to the absence of calcium. The enzyme would become active upon cytosolic Ca^{2+} ion entry, since higher Ca^{2+} ion concentrations reverse the inhibition [39].

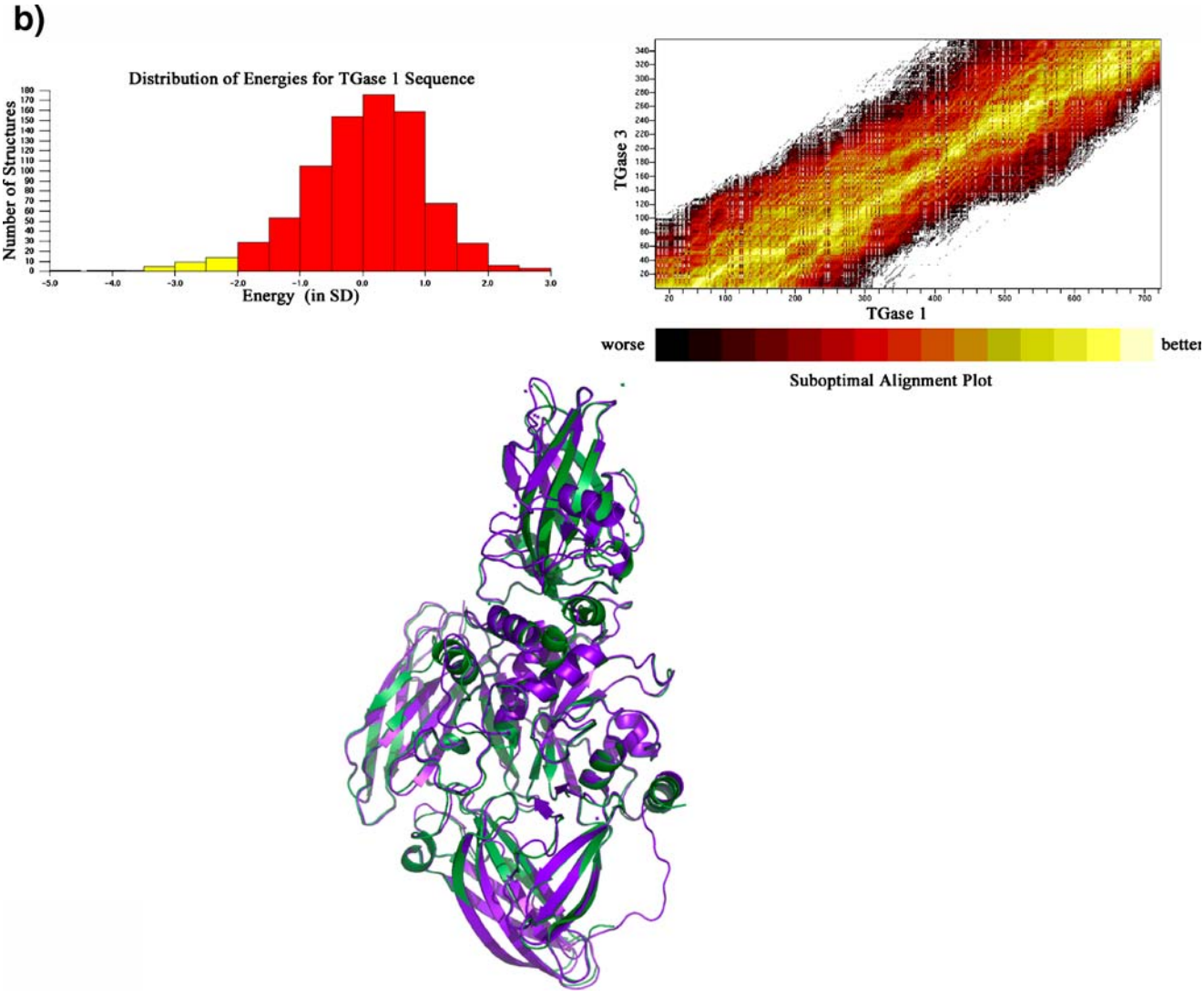

Fig. 2 (continued)

The amino-acid residues that form the three Ca^{2+} binding pockets are conserved in TGase 1, and as discovered in the TGase 3 crystal structure, are likely heptacoordinated with distorted pentagonal bipyramid geometries. One Ca^{2+} site (site 1, Fig. 2a) is located 13.7 Å away from the active-site Cys^{376} residue. In the activation of TGase 3 enzyme, a small surface loop effectively buries this Ca^{2+} ion in the interior of the enzyme. The Ca^{2+} ion site 2 is located about 23 Å below Cys^{376} and site 3 is about 22 Å from Cys^{376}. Notably, calcium binding moves a loop on the 'front' surface of the enzyme by about 9 Å, so that residue Asp^{429} can easily coordinate with the Ca^{2+} ion in site 3. The loop movement associated with this site opens a channel through the enzyme, exposing Trp^{341} and Trp^{432}, which are known to participate in the reaction mechanism [38]. The model of TGase 1 retains the three Ca^{2+} sites of TGase 3 suggesting a similar mode of activation. The binding of a Mg^{2+} ion in site 3 would release the loop, closing the channel to the active site, thus inactivating the enzyme [32]. It is conceivable that the channel opening could therefore be manipulated and controlled solely by intracellular cation levels.

The role of *cis* peptides in TGase 1

The TGase 1 model retains three *cis* peptides as they were found in the TGase 2 and TGase 3 structures, in which two are proline *cis* peptide bonds [10, 32]. One Pro^{372}–Tyr^{373} is located in the β−strand prior to the loop of the α-helix containing the active-site Cys^{376} residue and the second is at the loop containing Gly^{472}–Pro^{473}. The Lys^{487}–Tyr^{488}*cis*

Fig. 2 (continued)

peptide bond is located in a loop adjoining two α-helices of the core domain. Non-proline *cis* peptide bonds are very rare, occurring in only 0.03% of the peptide bonds in reported protein structures [40]. Extensive hydrogen bonding and hydrophobic side-chain interactions with neighboring residues contribute to the stabilization of energetically unfavored *cis* peptide bonds in proteins [41]. As expected, those found in the TGases are tightly bonded with neighboring structural motifs but the precise role of these conserved *cis* peptides in TGases remains unknown. The simplest explanation is that strain introduced during folding is localized in these *cis*-peptide residues. However, one can also speculate that the clustering around the active site implies an involvement in the catalytic cycle. Perhaps one or more of them undergo a reversible *cis* to *trans* isomerization (CTI) during the enzyme reaction cycle, either spontaneously upon binding of a substrate or during the catalysis by protonation or nucleophilic attack on the carbonyl carbon of the peptide bonds. It has been suggested that metastable *cis* bonds could store potential energy to drive biochemical reactions in enzymes [42]. It is possible that one or more of the *cis* peptide bonds in TGases isomerize upon substrate binding. Current evidence for this is unavailable since no TGase structure has been solved with a substrate or an inhibitor in the active site. If this transition is postulated, then the energy necessary to return the enzyme to the metastable *cis* form must be reintroduced to prepare for the next enzyme reaction cycle. Structures in the absence of substrate with and without bound Ca^{2+} ions contain the *cis* bonds and Ca^{2+} ions may stabilize this conformation. If substrate binding invokes a CTI then the reverse *trans* to *cis* conversion may be promoted by the Ca^{2+} ions, resulting in conformational changes that may expose some amino-acid side chains that are normally buried. These side chains may be involved in the open and closed states of this enzyme and thus provide an explanation for the activation of the enzyme by Ca^{2+} ions.

Substrate docking

Besides providing an understanding of the binding modes of substrates, docking studies have also been useful in validating homology models [38, 39, 43]. Some of the highly conserved residues in the TGases are believed to be necessary for catalytic hydrolysis [38]. For example, a thiolate/imidazolium ion pair can be formed between the catalytic Cys^{376} and His^{435} while Asp^{458} may facilitate appropriate orientation of the ion pair. Nucleophilic attack of the thiolate ion on an electron-deficient carbonyl of the substrate is proposed to result in the formation of a tetrahedral oxyanion intermediate bearing a negative charge [38, 39]. The negative charge is considered to be stabilized by the "oxyanion hole" formed by the indole hydrogen of a Trp residue. In addition, the hydroxyl of Tyr^{620} is within hydrogen-bonding distance of Cys^{376} and Trp^{341}, which is located at the loop of the β-barrel domain 1 that occludes the entrance to the active site. These bonds must be broken and structural motifs near the enzyme's surface must move to allow substrates to approach the active site. Also, there is evidence that two tryptophan residues Trp^{341} and Trp^{432} are involved in the enzyme reaction mechanism by forming an oxyanion intermediate first with the glutamine substrate and then with the lysine substrate in order to form the isopeptide cross-link bond [38].

We have modeled the peptides SQQ*VT (from loricrin) for the Gln* substrate and KTKQK* (from small proline rich protein 1) as the Lys* substrate (Fig. 4a), as these are used efficiently by TGase 1 [44]. These model substrates were docked in the active site of the modeled TGase 1 (Fig. 4b). Despite the overall similarity in the structures of

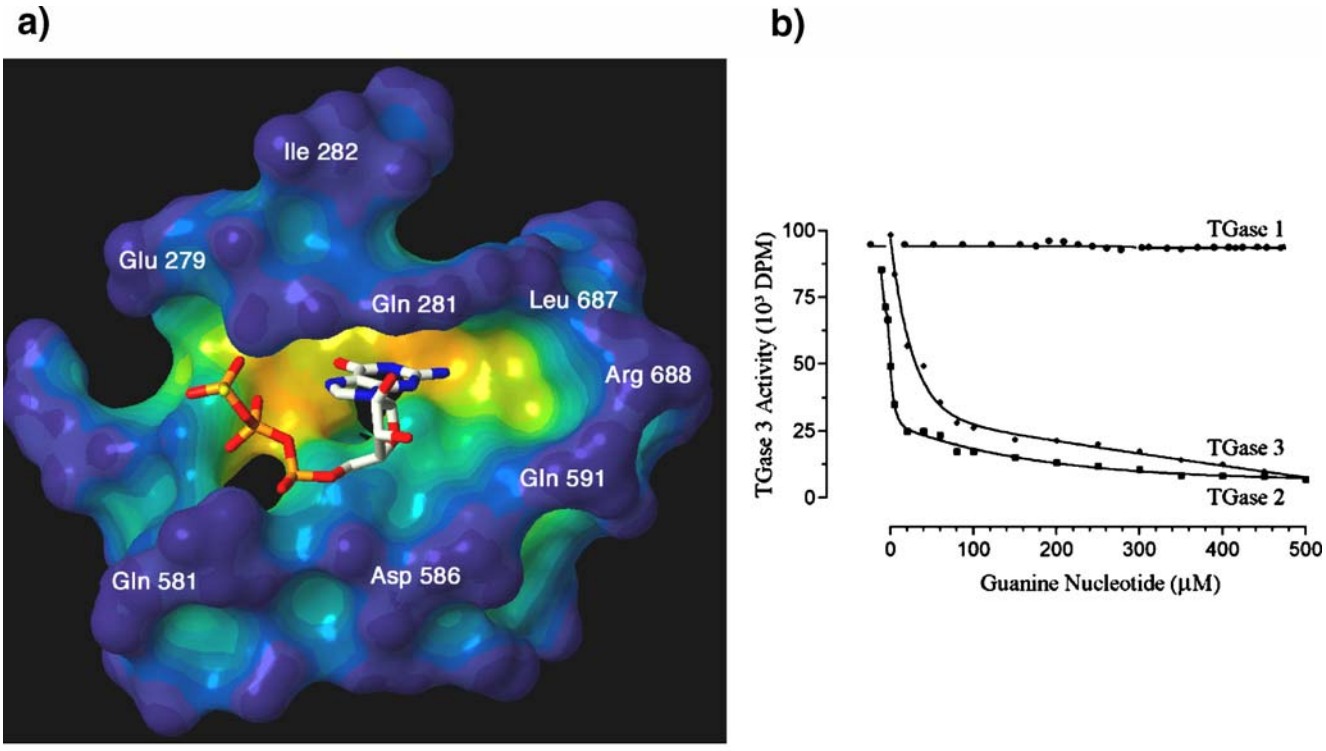

Fig. 3 **a** View of cleft surface (*yellow*) and the surrounding residues within 6.5 Å of the putative GTP nucleotide pocket located at the interface of the core domain (*top*) and the β-barrel 1 domain (*bottom*). **b** The inhibition of TGase 1 by increasing concentrations of GTP at 0.5 mM calcium concentration (■, TGase 2), (◆, TGase 3), and (●, TGase 1). The results represented in this figure are representative of one out of three experiments, each performed in triplicate

TGase 1 and TGase 3, there are a few significant differences in the residues forming the substrate binding sites. Comparison of the Cα backbone folding surrounding the active site reveals an important difference between these two structures. The region between substrate-binding pockets for TGase 1 appears to have a narrower opening than that in TGase 3. This difference is more apparent in the surface view of the two structures. To investigate this, we examined electrostatic surface representations of the access to the enzyme's active site (Fig. 4c). This difference in the binding-pockets of the two structures might explain some of the biochemical variation in the sensitivity of the two enzymes toward substrates. Since the binding-site residues are likely to govern selectivity and sensitivity of these enzymes, it is important to consider these differences in the design of substrates and inhibitors. In vitro studies have indicated that the TGase 3 enzyme cannot replace the essential cross-linking performed by the absence of TGase 1. For example, TGase 1 cross-links loricrin to the CE scaffold to form a large polymeric structure through interchain cross-linking with the preferential use of certain Gln and Lys residues, while TGase 3 uses the same residues with an intrachain cross-linking mechanism. [44] This indicates that there is a direct relationship between the types of cross-linking (inter- vs intrachain) and the specific residues used and the dual and complementary roles of these two TGases in this process [13, 44].

Discussion

LI, a severe skin disease that results from loss of function of TGase 1 due to mutations, dramatically illustrates the importance of TGase enzymes in CE formation. It has been shown that TGase 1 is essential for restoring the barrier function of the skin [17]. The TGase 1 model indicates the existence of three Ca^{2+} binding sites. The inset in Fig. 2a reveals that the Ca^{2+} ion at site 2 binds together upper portions of two β-strands of the core domain, designated β3 and β4, with two α-helices prior to the β3 strand and preceding the β4 strand, by coordination with side chains. Likewise, a Ca^{2+} ion at site 3 binds together upper regions of strands β1 and β2, which are adjacent to the mobile loop. Thus the β-strands and α-helices with connecting loops are linked together. Therefore, a plausible role of the two catalytic Ca^{2+} ions is to hold this cluster together to form a flap motif on the front surface of activated TGase 1. This motif overlays part of the β-barrel 1 domain and is

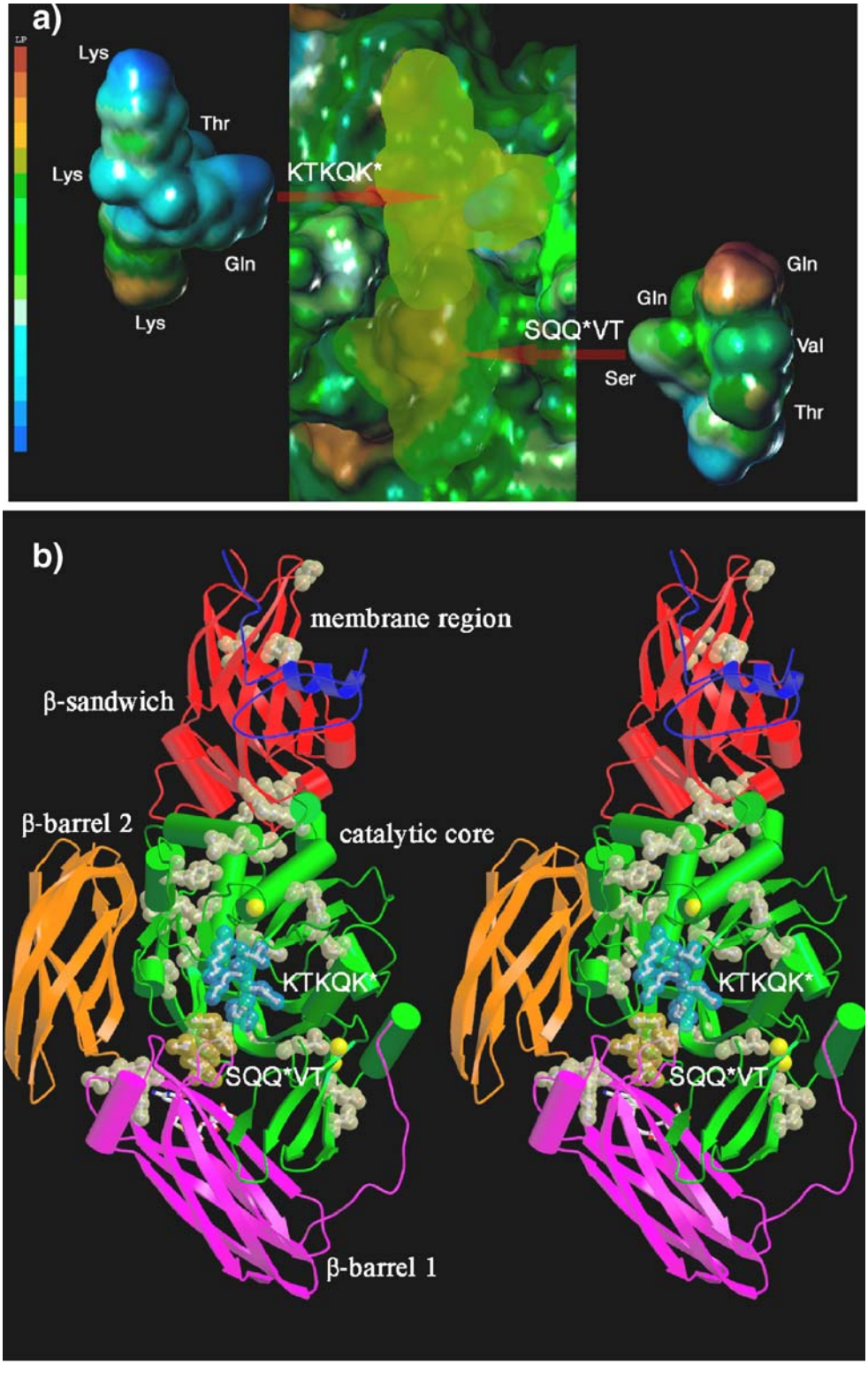

Fig. 4 **a** Mapping of the MOLCAD surface property (colored by hydrophobic region) of the TGase 1 enzyme active site is shown: lipophilic area (*brown*) and hydrophilic area (*blue*) potential over a range of 0.25–0.05, where blue to green to brown represents the spectrum from hydrophilic to hydrophobic regions. The surfaces of both peptides SQQ*VT (from loricrin) for the Gln* substrate on the right and KTKQK* (from small proline rich protein 1) as the Lys* substrate on the left of TGase 3 enzyme also are shown. **b** Stereo view of the SQQ*VT (*gold*) and KTKQK* (*light blue*) substrates determined by molecular docking. The five domains are the membrane region (*blue*), the β-sandwich (*red*), the catalytic core (*green*), the β-barrel 1 (*magenta*), and β-barrel 2 (*orange*). The Ca^{2+} ions in the catalytic core domain are shown in *yellow*. The relative position of TGase 1 mutated residues from LI patients is shown in *cream* color in CPK. **c** Stereo view of the electrostatic surface potential of TGase 1 enzyme. The acidic and basic residues are colored *red* and *blue*, respectively. The electrostatic potentials, including Ca^{2+} ions, have been mapped onto the surface plan from −15 kT (*deep red*) to +15 kT (*deep blue*). The electrostatic surface potential fragment from Pro^{24}–Ala^{55} region of the membrane anchorage region is shown separately. The electrostatic potentials, including that of Ca^{2+}, have been mapped onto the surface of TGase 3. The open channel is clearly evident when Ca^{2+} ion is present at site 3. The loop bearing residues ^{320}DKGSDS^{325} has moved from its position in the native enzyme, closing the channel when Ca^{2+} ion is absent

connected by hydrogen bonds. Interestingly, the flap motif region of the core domain shows low sequence homology with other members of the TGase family. The overlay of the TGase Cα backbone reveals that, while the strands of the flap motif are common, their orientations are different and this region is the most variable. Importantly, this region in TGases may play different roles. Indeed analysis of the temperature factor <*B*> values (Å^2) of crystal structures of TGase 2 and TGase 3 revealed variations due to thermal motions, suggesting that this flap motif is likely to be

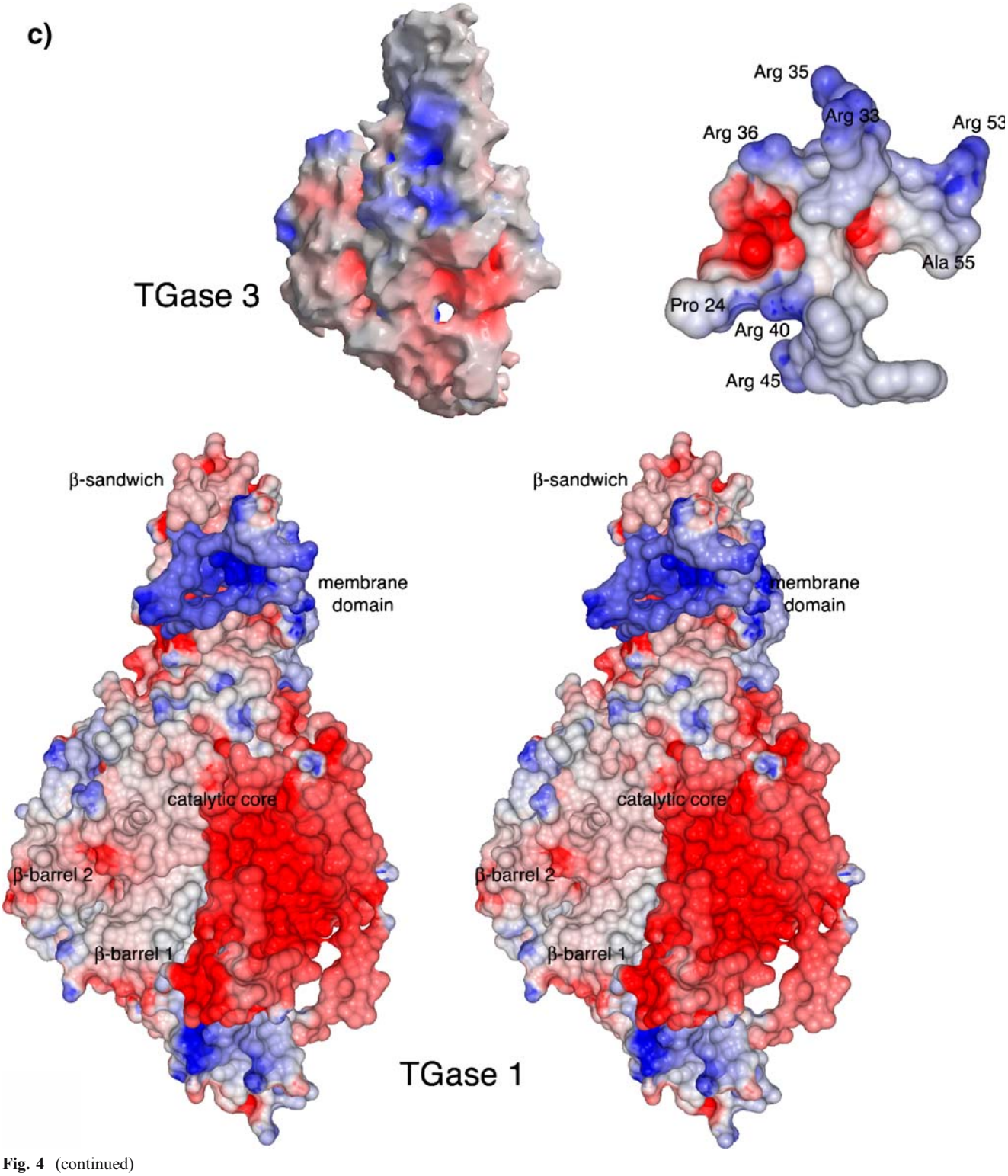

Fig. 4 (continued)

flexible [10, 11, 32]. We suggest that the Ca^{2+} ions in TGase 1 not only serve to anchor large clusters of motifs on the enzyme's surface in conjunction with conserved *cis* peptide bonds, but also serve as anchor points to stabilize and tighten structural motifs, especially around the active site. The routes of energy transfer between substrate, Ca^{2+} ion binding, and protein interactions are yet to be determined. We also postulate the existence of a CTI mechanism in which the three *cis* peptides discussed earlier play a pivotal role in the TGase reaction. CTI is suggested

to play a major role in the biochemical cycle of several enzymes where substrate binding and product release are coordinated with isomerization [45, 46]. Because of the structural restrictions in folded proteins, often one conformation, either *cis* or *trans*, is favored.

In the native state, thermal agitations may slowly drive polypeptide backbone interconversion. The *trans* to *cis* spontaneous conversion rate is rather slow but the reverse conversion rate is faster by 3 orders of magnitude, amounting to 10^{-3}–10^{1} s^{-1} at room temperature and pH 7 [47, 48]. The rate constant of CTI of the peptidyl-prolyl imide bond is of the order of 10^{-2} s^{-1}. The CTI interconversion rate in TGases may be accelerated dramatically, perhaps in the range of 10^{3}–10^{6} fold, by coupling the substrate binding to the isomerizing peptide bond as described for peptidyl-prolyl *cis–trans* isomerases [49–51]. In our proposed mechanism, the peptide bond in the active TGase first toggles from a *cis* to a *trans* conformation upon substrate binding. The tertiary structure may then be "locked in" if the energy barrier to the global energy minimum is relatively high. A spontaneous reverse interconversion slowly follows. For the peptide bond to isomerize, most of the free energy spread over the enzyme should focus into this specific *cis* peptide bond. The target substrate approach to TGase 1 may accelerate the reverse CTI, which then inflicts stress on the enzyme. CTI is not just a switching mechanism but also a force generator from the energy supplied by Ca^{2+} ions [38]. This binding may be sufficient for overcoming the rotational barrier in TGase 1. Lowering the activation energy may accelerate CTI rate by several orders of magnitude. The actual barrier height for CTI within a protein is highly dependent on the specific conditions within the CTI pocket, as evident in peptidyl-prolyl *cis–trans* isomerases that catalyze CTI without investing Ca^{2+} ion binding energy [49–51].

It is well established that among the mammalian TGases characterized thus far, TGase 2 and TGase 3 have been shown to bind and hydrolyze GTP to GDP [10, 11]. For instance the binding of GTP to TGase 3 results in the loss of Ca^{2+} binding at site 3 and occupation of this site by Mg^{2+} ions [10, 11, 35]. It has been shown that the channel opening and closing to the active site could be modulated based on which metal is bound at site 3 [11]. Furthermore it has been proposed that the front of the channel could be used by the two substrates for enzyme reaction, which is based on detailed analyses of current structural information of the zymogen and the activated forms of TGase 3 [39]. Therefore, we examined the model of the TGase 1 structure in the presence of GTP in terms of its implications for substrate access to the active site residues. To do this, we relied on electrostatic surface representations of the structures that allow examination of access to the enzyme's active site. In the TGase 3-GDP complex, two simultaneous conformations are associated with both the closed and open substrate channels respectively [11, 38]. This would indicate that the TGase 1, in a similar manner to TGase 3, is able to readily shift between the active and inactive states when bound to GDP. The marked differences in the amino acid residues that make up the nucleotide-binding pocket of the TGase isozymes suggests that other TGases might use still other residues to bind nucleotides (Fig. 1).

We are intrigued by the observed channel modulation directed toward the active site, as this channel appearance coincides closely with the activation of the enzyme. It appears likely that the channel could serve as 'ports of entry' for the two substrates and might be synchronized with CTI upon substrate binding. Thus, the geometrical configuration of the *cis*-peptide bond is not static but can isomerize from a metastable *cis* to a more stable *trans* configuration in TGase 1 in the presence of a substrate analog. We propose that a specific Q* substrate could promote a similar isomerization. If so, this could serve as a pivot for the flexible flap motif. There is also another conserved *cis* peptide bond just prior to the catalytic active site residue, which could conceivably regulate enzyme activity. While the role of these bonds in TGases remains unresolved, it was suggested that a CTI of *cis*-peptide bonds might be involved in TGase-substrate activation [38]. Although the *cis*-peptide bonds in proteins are energetically unfavorable (metastable), such bonds store potential energy to drive biochemical reactions in other enzymes [49–51]. However, isomerization is typically very slow and may occur only once to activate an enzyme [47, 48].

The model of 34 unique mutation sites from our survey of PubMed suggests that these mutations could affect misfolding and stability of the TGase 1 enzyme, as shown in Table 1. Most of these residues are involved in hydrogen bonding interactions either in the same domain or across a domain interface. The majority of mutations are attributed to TGase 1 cross-linking deficiency affecting the integrity of CEs in the epidermis. These mutations could include single nucleotide changes leading to amino acid alteration, nonsense mutations resulting in a premature termination codon, and finally mutations in introns which effect splicing and leading to frame shifts resulting in deletion of amino acid(s). Most of the reported mutations may also influence the protein's intrinsic solvation properties, local surface polarity or surface hydrophobicity, and perhaps the physiological pH of its environment. Conformational changes near the catalytic core and calcium binding regions are likely to impair enzymatic activity of TGase 1. There are many ways in which mutations can affect enzyme activity including active site modification, errors in protein folding, and protein–protein interactions. When interface residues are mutated, complementarity (electrostatic interactions and shape) is perturbed. Deleting a large side chain creates a

Table 1 Key residues of patients with TGase 1 missense mutations (34 amino acid mutations) were obtained from a search of PubMed. The residues with an *asterisk* represent the nonconserved residues

	Missense Mutation	Domain	Point Mutation	Exon
1	S41Y*	Membrane	C→A	2
2	C52S*		T→A	2
3	G93D*		G→A	2
4	D101V*	β-sandwich	A→T	2
5	R141C		C→T	3
6	R141H		G→A	3
7	R141P		G→C	3
8	R142C		C→T	3
9	R142H		G→A	3
10	G143E		G→A	3
11	G143R		G→A	3
12	S159C*		C→G	3
13	L204Q		T→A	4
14	G217S		G→A	4
15	R224H*		G→A	4
16	P248L	Catalytic core	C→T	4
17	S271P*		T→C	5
18	G277R		G→A	5
19	R285Q*		G→A	5
20	N288T		A→C	5
21	R306W*		C→T	6
22	R314L		G→T	6
23	R314C		C→T	6
24	R322Q		G→A	6
25	S357R		C→G	7
26	Y364D		T→G	7
27	V378L		G→C	7
28	G381R		G→A	7
29	V382M		G→A	7
30	R388H		G→A	7,8
31	R388P		G→C	8
32	G391D		G→A	8
33	R395L		G→T	8
34	F400V		T→G	8
35	D429V		A→T	8
36	G472S		G→A	10
37	D489G		A→G	10
38	V517M*		G→A	11
39	E519G*		A→G	11
40	R686C*	β-barrel 1	C→T	13

hole that solvent may or may not fill depending on its size; removing or adding a polar group leaves a hydrogen-bond donor or acceptor unsatisfied. The mutations therefore may abrogate protein–protein interaction involved in activation of keratinocyte TGase 1 as a key enzyme for the processing of CE proteins such as involucrin, loricrin and filaggrin during epidermal differentiation.

Based on detailed analyses of current structural information as well as molecular modeling of Gln and Lys substrates, we have presented here a new model of TGase 1 and its mechanism of action. The model offers explanations for: (1) the absolute requirement of three Ca^{2+} ions; (2) the purpose of the channel opening to the catalytic active site; (3) the role of the three *cis* peptide bonds; (4) the energy needed to move certain sequence motifs on the enzyme surface that occlude access to the active site region; (5) how the glutamine and then lysine substrates approach the active site triad residues to perform the chemistry of reaction; and (6) why the TGase reaction as currently understood is so slow. What is far from clear, at this time is an explanation for the marked degree of substrate specificity of TGase 1. Nevertheless, we believe this model will stimulate efforts to generate experimental data, which may offer clues to TGase 1 specificity and the design of TGases isoform specific inhibitors.

The presented model of mutations of TGase 1 structure further reveals some interesting clues to understanding LI. The abnormalities in CE structure in epidermis and appendages as a result of defective TGase 1 may lead to symptoms of large dark, lamellar scales on the entire body, the classic phenotype of LI. It is important to identify and understand the effect of occurring mutations on the critical regions of TGase 1. These mutations could affect normal cross-linking of CEs in maturing epidermal keratinocytes, which is required as a critical scaffold for the formation of the extracellular lamellar membranes, and thus affects the permeability barrier of the skin. Further understanding of the causes of LI will improve genetic counseling and prenatal diagnosis.

Acknowledgments We thank Dr. Henry Hennings of NCI's Laboratory of Cellular Carcinogenesis and Tumor Formation for critical review of our manuscript. We are indebted to Dr. Ellis Kempner for stimulating discussion and critical review of this manuscript. This research was supported by the Intramural Research Program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health.

References

1. Anton-Lamprecht I (1992) The skin. In: Papadimitriou JM, Henderson DW, Spagnolo DV (eds) Diagnostic ultrastructure of non-neoplastic diseases. Edinburgh, Churchill and Livingstone, pp 459–550
2. Williams ML, Elias PM (1985) Arch Dermatol 121:477–488
3. Russell LJ, DiGiovanna JJ, Rogers GR, Steinert PM, Hashem N, Compton JG, Bale SJ (1995) Nat Genet 9:279–283
4. Russell LJ, DiGiovanna JJ, Hashem N, Compton JG, Bale SJ (1994) Am J Hum Genet 55:114–1152
5. Laiho E, Niemi KM, Ignatius J, Kere J, Palotie A, Saarialho-Kere U (1999) Eur J Hum Genet 7:625–632
6. Laiho E, Ignatius J, Mikkola H, Yee VC, Teller DC, Niemi KM, Saarialho-Kere U, Kere J, Palote A (1997) Am J Hum Genet 61:529–538
7. Williams M, Elias PM (2000) Curr Probl Dermatol 12:170–176
8. Huber M, Rettler I, Bernasconi K, Frenk E, Lavrijsen SP, Ponec M, Bon A, Lautenschlager S, Schorderet DF, Hohl D (1995) Science 267:525–528

9. Choate KA, Williams ML, Khavari PA (1998) J Invest Dermatol 110:8–12
10. Liu S, Cerione RA, Clardy J (2002) Proc Natl Acad Sci 99:2743–2747
11. Ahvazi B, Boeshans KM, Idler W, Baxa U, Steinert PM, Rastinejad F (2004) J Biol Chem 279:7180–7192
12. Kim IG, McBride W, Wang M, Kim SY, Idler WW, Steinert PM (1992) J Biol Chem 267:7710–7717
13. Steinert PM, Kim SY, Chung SI, Marekov LN (1996) J Biol Chem 271:26242–26250
14. Nemes Z, Marekov LN, Fesüs L, Steinert PM (1996) Proc Natl Acad Sci 96:8402–8407
15. Phillips MA, Stewart BE, Qin Q, Chakravarty R, Floyd EE, Jetten AM, Rice RH (1990) Proc Natl Acad Sci 87:9333–9337
16. Phillips MA, Qin Q, Mehrpouyan M, Rice RH (1993) Biochemistry 32:11057–11063
17. Pilgram GSK, Vissers DCJ, van der Meulen H, Pavel S, Lavrijsen SPM, Bouwstra JA, Koerten HK (2001) J Invest Dermatol 117:710–717
18. Hennies HC, Kuster W, Wiebe V, Krebsova A, Reis A (1998) Am J Hum Genet 62:1052–1061
19. Gish W (1996–2002) http://blast.wustl.edu/blast/
20. Godzik A, Kolinski A, Skolnick J (1992) J Mol Biol 227:227–238
21. McGuffin LJ, Bryson K, Jones DT (2000) Bioinformatics 16:404–405
22. Lüthy R, Bowie JU, Eisenberg D (1992) Nature 356:83–85
23. Sabnis Y, Desai PV, Rosenthal PJ, Avery MA (2003) Prot Sci 12:501–509
24. Jones TA, Thirup S (1986) EMBO J 5:819–822
25. Stouten PFW, Froemmel C, Nakamura H, Sander C (1993) Mol Simul 10:97–120
26. Greengard L, Rokhlin VI (1987) J Comp Phys 73:325–348
27. Verlet L (1967) Phys Rev 159:98–103
28. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1992) J Appl Cryst 26:283–291
29. Merritt EA, Bacon DJ (1997) Method Enzymol 277:505–525
30. Kraulis PJ (1991) J Appl Cryst 24:946–950
31. SYBYL 6.7 (1995) Tripos Assoc. Inc., St. Louis, MO
32. Ahvazi B, Kim HC, Kee SH, Nemes Z, Steinert PM (2002) EMBO J 21:2055–2067
33. Ahvazi B, Boeshans KM, Jang S-I, Kalinin P, Steinert PM (2002) Minerva Biotecnol 14:165–169
34. Ahvazi B, Boeshans KM, Idler W, Baxa U, Steinert PM (2003) J Biol Chem 278:23834–23841
35. Ahvazi B, Boeshans KM, Steinert PM (2004) J Biol Chem 279:26716–26725
36. Menon GK, Elias PM (1991) Arch Dermatol 127:57–63
37. Hennings H, Michael D, Cheng C, Steinert P, Holbrook K, Yuspa SH (1980) Cell 19:245–254
38. Ahvazi B, Steinert PM (2003) Exp Mol Med 35:228–242
39. Ahvazi B, Boeshans KM, Rastinejad F (2004) J Struct Biol 147:200–207
40. Weiss MS, Jabs A, Hilgenfeld R (1998) Nat Struct Biol 5:676
41. Jabs A, Weiss MS, Hilgenfeld R (1999) J Mol Biol 286:291–304
42. Stoddard BL, Pietrokovski S (1998) Nat Struct Biol 5:3–5
43. Folkers G (1998) In: Codding PW (ed) Structure-based drug design: experimental and computational approaches. Kluwer Academic Publisher, Norwell, MA, pp 271–283
44. Candi E, Tarcsa E, Idler WW, Kartasova T, Marekov LN, Steinert PM (1999) J Biol Chem 274:7226–7237
45. Ludwig M, Pattridge K, Metzger A, Dixon M, Eren M, Feng Y, Swenson R (1997) Biochemistry 36:1259–1280
46. Heroux A, White E, Ross L, Davis R, Borhani D (1999) Biochemistry 38:14495–14506
47. Stewart D, Sarkar A, Wampler J (1990) J Mol Biol 214:253–260
48. Lin LN, Brandts JF (1993) Biochemistry 22:564–573
49. Scherer G, Kramer M, Schutkowski M, Reimer U, Fischer G (1998) J Am Chem Soc 120:5568–5579
50. Schiene-Fischer C, Fischer G (2001) J Am Chem Soc 123:6227–6231
51. Reimer U, Fischer G (2002) Biophys Chemist 96:203–212