Molecular Cell, Vol. 15, 879–888, September 24, 2004, Copyright ©2004 by Cell Press

A C^{α} Model for the Transmembrane α Helices of Gap Junction Intercellular Channels

Sarel J. Fleishman,[1] Vinzenz M. Unger,[2] Mark Yeager,[3,4] and Nir Ben-Tal[1,*]
[1]Department of Biochemistry
George S. Wise Faculty of Life Sciences
Tel-Aviv University
Ramat Aviv, 69978
Israel
[2]Department of Molecular Biophysics and Biochemistry
Yale University
P.O. Box 208024
New Haven, Connecticut 06520
[3]Department of Cell Biology
The Scripps Research Institute
10550 North Torrey Pines Road
La Jolla, California 92037
[4]Division of Cardiovascular Diseases
Scripps Clinic
10666 North Torrey Pines Road
La Jolla, California 92037

Summary

Gap junction channels connect the cytoplasms of apposed cells via an intercellular conduit formed by the end-to-end docking of two hexameric hemichannels called connexons. We used electron cryomicroscopy to derive a three-dimensional density map at 5.7 Å in-plane and 19.8 Å vertical resolution, allowing us to identify the positions and tilt angles for the 24 α helices within each hemichannel. The four hydrophobic segments in connexin sequences were assigned to the α helices in the map based on biochemical and phylogenetic data. Analyses of evolutionary conservation and compensatory mutations in connexin evolution identified the packing interfaces between the helices. The final model, which specifies the coordinates of C^{α} atoms in the transmembrane domain, provides a structural basis for understanding the different physiological effects of almost 30 mutations and polymorphisms in terms of structural deformations at the interfaces between helices, revealing an intimate connection between molecular structure and disease.

Introduction

A gap junction channel is formed by the end-to-end docking of two hexameric hemichannels or connexons (Kumar and Gilula, 1996). Each hexamer is formed by six connexin subunits (Cascio et al., 1995) that are composed of four hydrophobic transmembrane (TM) segments designated M1–M4 from the N- to the C terminus (Milks et al., 1988). The intercellular pore of gap junction channels is roughly 15 Å in diameter and allows transport of cytoplasmic secondary messengers, thereby mediating signaling and ion current flow between neighboring cells. Over the past several decades, the important role that gap junctions play in coordinating tissue and organ physiology, e.g., in the heart, ear, skin, and pancreas, has been increasingly recognized (Harris, 2001). A number of genetic conditions in humans and mouse models involving the skin, neurodegenerative and developmental diseases, and most cases of nonsyndromic hereditary deafness have been attributed to mutations in connexins (reviewed by Kelsell et al., 2001).

We previously used electron cryomicroscopy (cryo-EM) and image analysis to solve the structure of a recombinant gap junction channel formed by a C-terminal truncation mutant of Cx43. The three-dimensional (3D) density map at 7.5 Å in-plane resolution revealed the close packing of 24 α helices within each connexon (Unger et al., 1999). Since publication of the original map, improvements in the data analysis have allowed calculation of a map with 5.7 Å in-plane and 19.8 Å vertical resolution. Each of the helices is clearly resolved from its neighbors in the TM domain, and the helices' centers of gravity are also discernible, allowing accurate determination of the helix positions, tilt, and azimuthal angles. However, even in this improved map, connecting loops remained largely undefined either because of limitations in the vertical resolution (in the nonhelical structure of extracellular loops) or disorder (in the cytoplasmic domains). This precluded direct assignment of the helices in the map to the TM domains in the connexin sequence. Consequently, the molecular basis for ionic conduction, channel permeability, and gating properties among the various connexin isoforms could not be inferred directly from the cryo-EM map (Harris, 2001).

However, there is a large body of biochemical and biophysical evidence (reviewed by Harris, 2001) that provides insight into the TM boundaries for M1–M4 and subunit topology (Bennett et al., 1994) and the identities of the pore-lining helices (Kronengold et al., 2003; Skerrett et al., 2002; Zhou et al., 1997). We used these data to assign the TM segments M1–M4 to the helices observed in the cryo-EM map (Unger et al., 1999). We then combined the helix positions, tilt, and azimuthal angles from the improved cryo-EM map with computational methods for the analysis of evolutionary conservation and hydrophobicity of amino acid residues (Fleishman et al., 2004b) to generate a C^{α} trace model of the 24 helices in the connexon. Even though the cryo-EM map corresponds to Cx43, our analysis was based on the human Cx32 sequence since there is a wealth of biochemical, mutational, and genetic data for this isoform. Modeling Cx32 on the basis of the Cx43 structure is justified because the two proteins exhibit 50% sequence identity in the predicted TM residues of M1–M4 (Yeager and Gilula, 1992). Moreover, various connexins assemble to form heteromeric connexons (Harris, 2001). It is therefore very likely that connexins share a common architecture, at least in the TM domain. Consequently, the model we describe should serve as a template for other connexins.

Our approach followed that used by Baldwin et al. (1997) to predict the structure of the TM domain of vertebrate rhodopsin based on a cryo-EM map at 9 Å in-

*Correspondence: nirb@tauex.tau.ac.il

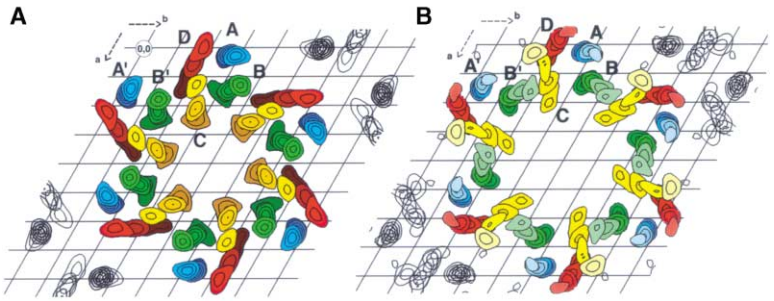

Figure 1. Overlay of Cross-Sections of the 3D Density Map of One Connexon Derived by Electron Cryocrystallography

Counting from the middle of the extracellular gap and toward the observer, sections +14, +18, and +24 (A) and +20, +24, +29, and +34 (B) were used. The approximate boundary between the membrane and the extracellular gap corresponds to section +8 (not shown). The vertical distance between consecutive sections is 2 Å. Densities belonging to the same helices are represented by the same base color, with the darkest and lightest shades corresponding to densities in sections +14 and +34, respectively. Helices were arbitrarily marked A–D and A′ and B′ (which are symmetry related to A and B) to provide a reference for discussion. The position marked (0,0) was used to generate grid coordinates for the locations of helices A–D given in Table 1. The spacing between grid lines is 10 Å, and the map was contoured starting at 1.5σ above the mean.

plane and 16.5 Å vertical resolution (Unger et al., 1997). The model of rhodopsin was shown to be quite accurate (Bourne and Meng, 2000) when compared to the subsequent high-resolution X-ray structure (3.2 Å rmsd) (Palczewski et al., 2000). We have used a similar approach (Fleishman et al., 2004b), which relies on the assumption that conserved amino acid residues preferentially pack at helix-helix interfaces, whereas the positions that face the lipid or the pore lumen are variable (Baldwin et al., 1997). In addition, it is unfavorable for charged residues to face the lipid, except for the terminal helical turns, where charged positions may interact favorably with the polar-headgroup region. Where conservation and hydrophobicity did not suffice to produce an unambiguous conformation, we applied a computational tool for identifying pairs of positions that exhibit correlated evolution, which is often associated with contact formation in the protein's tertiary structure (Fleishman et al., 2004a; Gobel et al., 1994). We thus computed a structure for the entire TM domain of the gap junction hemichannel.

Results

Helix Assignment

Analysis of superimposed cross-sections from the TM density of one connexon (Figure 1) revealed the following helix tilts (Table 1): 9.1° (A), 15.6° (B), 27.5° (C), and 29.2° (D). The contoured sections identified section 29 (second from the top in Figure 1B) as being the last section of helix A that exhibited significant density. Based on the necessity that the aqueous pore be shielded from the membrane lipids, we concluded that section 29 was located close to the cytoplasmic boundary of the membrane. Densities past section 29 for helices B, C, and D likely represented parts of the cytoplasmic domains (N-tail, C-tail, and the M2-M3 cytoplasmic loop).

There are 24 different assignments of the hydrophobic domains M1–M4 to the four helices in the cryo-EM map (Nunn et al., 2001). At the outset, we stress that no single helix assignment can be reconciled with all of the experimental data on connexins (Harris, 2001). This is in part due to the channel's plasticity and the heterogeneity of methods and connexin isoforms on which relevant studies were based. Our approach has therefore been to use primarily the cryo-EM map together with data on hydrophobicity and evolutionary conservation. We relied on other experimental evidence to provide support only in cases where there was substantial agreement between different studies. With Figure 1 as a reference, the following describes clues from different sources that were used to derive an assignment of helices A–D to the TM segments M1–M4 in the connexin sequences.

The substituted-cysteine accessibility method (SCAM) (Karlin and Akabas, 1998) demonstrated that specific residues on M1 as well as in the N-terminal part of E1 are accessible for labeling by water-soluble sulfhydryl reagents (Kronengold et al., 2003; Skerrett et al., 2002; Zhou et al., 1997). A detailed analysis showed that M3 was the major pore-lining helix (Skerrett et al., 2002). Notably, both M1 and M3 contain several evolutionarily

Table 1. Estimated Axes of the TM α Helices

	Tilt and Azimuthal Angles		Positions			
Helix	Θ (°)	φ (°)	a14 (Å)	b14 (Å)	a24 (Å)	b24 (Å)
A (blue)	9.1	0.0	2.4	24.8	2.4	28.8
B (green)	15.6	28.0	15.6	32.4	12.4	34.8
C (yellow)	27.5	90.0	23.2	27.6	11.6	21.6
D (red)	29.2	60.0	10.4	18.0	−0.8	18.0

Colors refer to Figure 1. Positions a14, b14, a24, and b24 were derived from the grid shown in Figure 1 using (0,0) as common origin. With +z pointing towards the observer, sections 14 and 24 are located +28 Å and +48 Å from the center of the extracellular gap. The values for the azimuthal angles (φ) were derived by centering orthogonal x,y-coordinate systems at the a14, b14 positions for each of the helices and measuring the angles between the x-axis, oriented parallel to b, and the projected paths of the helices connecting the points (a14, b14) and (a24, b24). Positive φ angles were measured counterclockwise from x in the direction of y. The values for the tilt angles (Θ) were measured as the angle between the projected path of the helices and the z axis. The estimated axes assume that the α helices are straight.

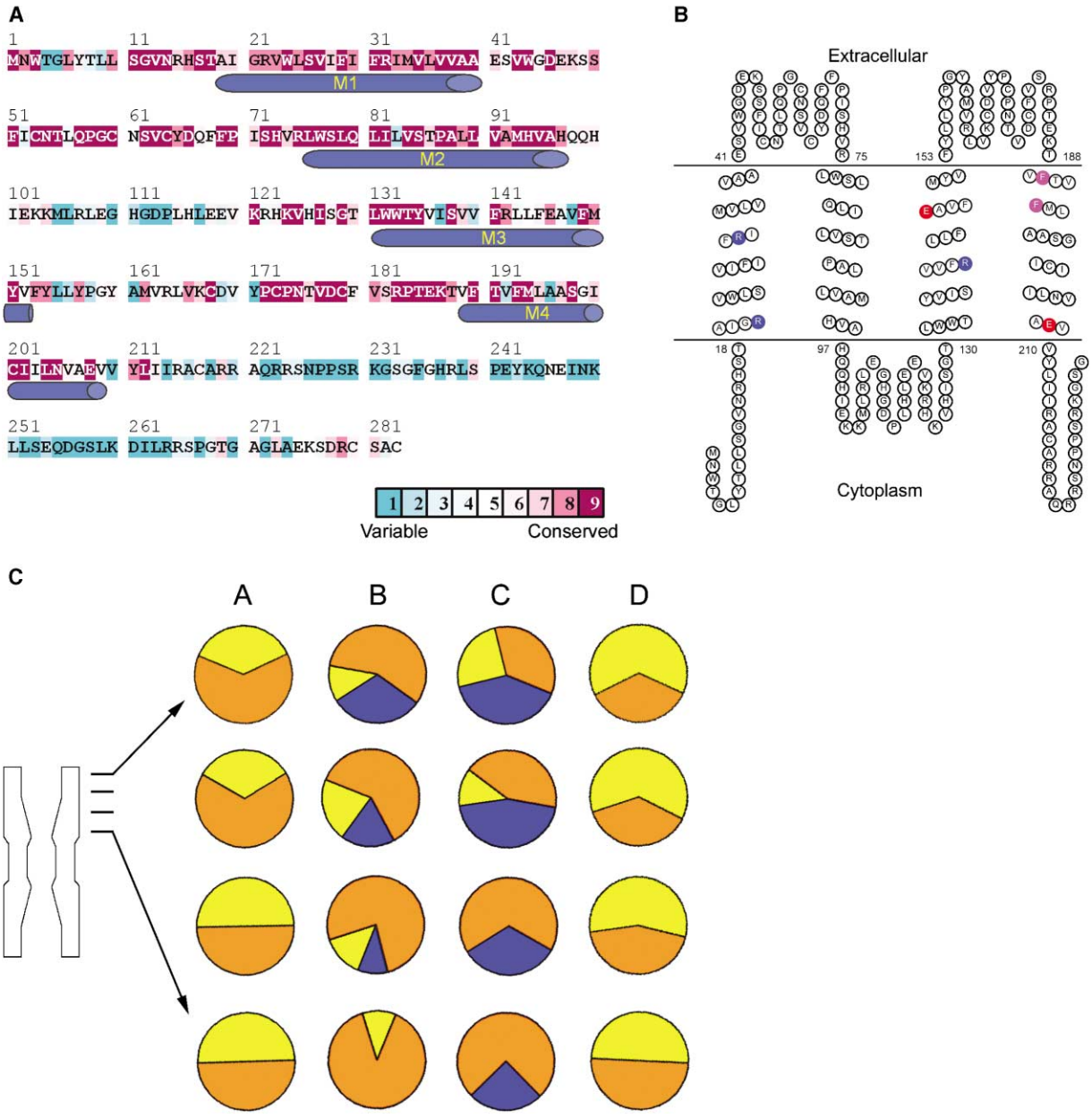

Figure 2. Connexin Architecture and Amino Acid Conservation

(A) The sequence of human Cx32 color-coded according to evolutionary conservation using the *ConSeq* server (Berezin et al., 2004), with turquoise-through-maroon corresponding to variable-through-conserved positions (see color bar). The hydrophobic segments M1–M4 are marked on the sequence.

(B) Membrane topology of Cx32. Acidic and basic amino acids in the TM domain are marked red and blue, respectively. The transmembrane segments M1–M4 and the two extracellular loops E1 and E2 are indicated. Two aromatic residues are colored magenta. Numbers indicate the positions of the extramembrane domain boundaries. Part of the C terminus was truncated.

(C) As indicated in the schematic model (left), four cross-sections evenly distributed within the membrane region of one connexon were evaluated. The approximate total areas facing the aqueous pore (blue), the membrane lipid (yellow), and neighboring α helices (orange) were estimated in each section for each of the four helices A–D. The orientations of the pie charts are arbitrary. As suggested in Figure 1, this representation clearly reveals that only helices B and C have access to the aqueous pore. Furthermore, each of the helices has a characteristic accessibility pattern that was used in combination with the conservation profile of (A) to assign each helix to a specific TM sequence (see text).

conserved charged residues (Figure 2A). The important role of M3 in lining the pore is also suggested by the amphipathic pattern of its conserved polar and charged amino acids (Milks et al., 1988). In contrast, M2 is devoid of any charges, and M4 contains a single Glu residue at position 208 toward the cytoplasmic domain, which is likely to be outside the hydrophobic core of the bilayer (Figure 2B).

Hence, without committing to the specific identities of the pore-lining helices, a generalized assignment for

M1 and M3 could be made by assessing which of the helices in the structure had access to the aqueous pore. From Figure 1, it was clear that only helices B and C lined the aqueous pore, which suggested that these segments corresponded to M1 and M3. However, this first assignment step did not allow us to distinguish between the two possible alternatives. Nevertheless, if M1 and M3 corresponded to helices B and C, then it followed that helices A and D corresponded to TM segments M2/M4 in the connexin sequence.

After division of the four TM segments into two groups (i.e., B/C = M1/M3 and A/D = M2/M4), the number of options for a specific assignment could be limited by a comparison of connexin amino acid sequences using an approach similar to Baldwin's analysis of the G protein coupled receptor family (Baldwin, 1993). Specifically, residues in the lipid-facing positions of TM helices were the least conserved among the receptors. A similar analysis based on 60 connexin sequences (Berezin et al., 2004) showed that the relative conservation of the TM segments was M2 > M4 and M1 > M3 (Figure 2A). We reasoned that evolutionary variability within the TM segments indicated that amino acid residues in these positions were not very important for helix packing and were therefore more likely to face the membrane lipid or the large pore lumen.

A specific helix assignment could then be made by assessing the extent to which the α helices in the structure had access to the lipid and the aqueous pore. Cross-sections similar to those shown in Figure 1 were chosen throughout the membrane-spanning part of one connexon (Figure 2C). In each cross-section, we estimated what part of each of the helices faced the aqueous pore, packed against neighboring helices, or was exposed to the lipid. Helix C was found to be more exposed to the aqueous pore than B. Hence, of the M1/M3 pore-lining pair, the highly conserved M1 most likely corresponded to B, and M3 to the major pore-lining helix C. A similar analysis showed that helix D was more exposed to the lipid environment than was helix A (Figure 2C). Therefore, of the M2/M4 lipid-exposed pair, the conserved M2 most likely corresponded to the more buried helix A, and M4 to the lipid-exposed helix D. Interestingly, the evolutionary conservation of M3 showed a decrease in the central part of the bilayer (Figure 2A), which coincided with an increase in the exposure of helix C to the pore lumen (Figure 2C). Similarly, the conservation of M4 decreased toward the cytoplasmic side, correlating with an increase in its exposure to the membrane.

Helix Orientations

Canonical α helices were constructed based on the parameters defined in Table 1 (Figure 3). A starting C^{α} model for the 24 α helices in the hexameric connexon was built using the assignment M1 = B, M2 = A, M3 = C, and M4 = D. We used an exhaustive search and scoring function to sample the rotational orientation of each of the helices around their principal axes, while maintaining 6-fold symmetry around the channel axis (Fleishman et al., 2004b). This search yielded the optimal conformation shown in Figure 4A. It is evident that helices M3 and M4 show a very clear evolutionary variability versus conservation signal, with the variable residues mapping to one helical face. Indeed, the optimal conformation placed all of the evolutionarily variable positions of M3 and M4 in lumen- or lipid-exposed positions, respectively, whereas conserved faces were packed inside the protein core.

In contrast to M3 and M4, the residues in M1 and M2 are homogeneously conserved (Figure 2A), so that the orientations around their principal axes cannot be determined reliably on the basis of conservation alone. Correlated amino acid evolution has been used previously to identify interresidue contact (e.g., Gobel et al., 1994). The underlying assumption was that pairs of residues that form contact undergo dependent evolution, i.e., a substitution in one position would induce the other to change in order to maintain the protein fold.

To detect correlations, we applied a method that was especially designed for treating intermediate-sized protein families (50–100 sequences) (Fleishman et al., 2004a) such as connexins. We identified five pairs of correlations in the TM and juxtamembrane domains that are connected by solid lines in Figure 4B. Positions in the juxtamembrane domain (3 positions from the end of the hydrophobic stretch at most) were assumed to conform to α-helical ideality. Based on these correlations we manually oriented helices M1 and M2 to obtain a conformation in which each of the two positions of a correlated pair would be in proximity (Figure 4B). The correlations that pertained to helix M3 were in accordance with the helix's orientation around its principal axis as determined above by the evolutionary conservation analysis. Moreover, the five pairs of correlations were accommodated by the model, thus providing additional support for the model at various levels, including the TM-domain boundaries, helix assignment, and the orientations of the helices around their principal axes.

Structural Features

It is difficult to provide a detailed structural interpretation of the model at this resolution since the computed structure does not contain information regarding side chain conformations. Moreover, we estimate that the orientations of the helices around their principal axes may vary by up to 40°. Nevertheless, even at this level of uncertainty, it is possible to provide a rough description of the factors that stabilize the structure.

The lipid-exposed residues of M2 and M4 are mostly hydrophobic. In fact, these helices are devoid of charged amino acids, except for Glu208 on M4 (Figures 2B and 5A). This residue is just two amino acid positions from the C-terminal end of the hydrophobic segment and is located in the protein core, toward the cytoplasmic side of the protein. Hence, it is not exposed to the membrane environment and, due to the tilt of helix M4, might be surrounded by water from the cytoplasm. Position Arg22 on M1 faces the protein core on the cytoplasmic side of the protein (Figure 5B). Likely, this position "snorkels" (von Heijne, 1996) to the cytoplasmic side of the lipid bilayer according to the positive-inside rule (von Heijne, 1989). Another possibility is that Glu208 and Arg22, which are oriented toward one another, form a salt bridge.

Most of the charged residues in M1 and M3 are posi-

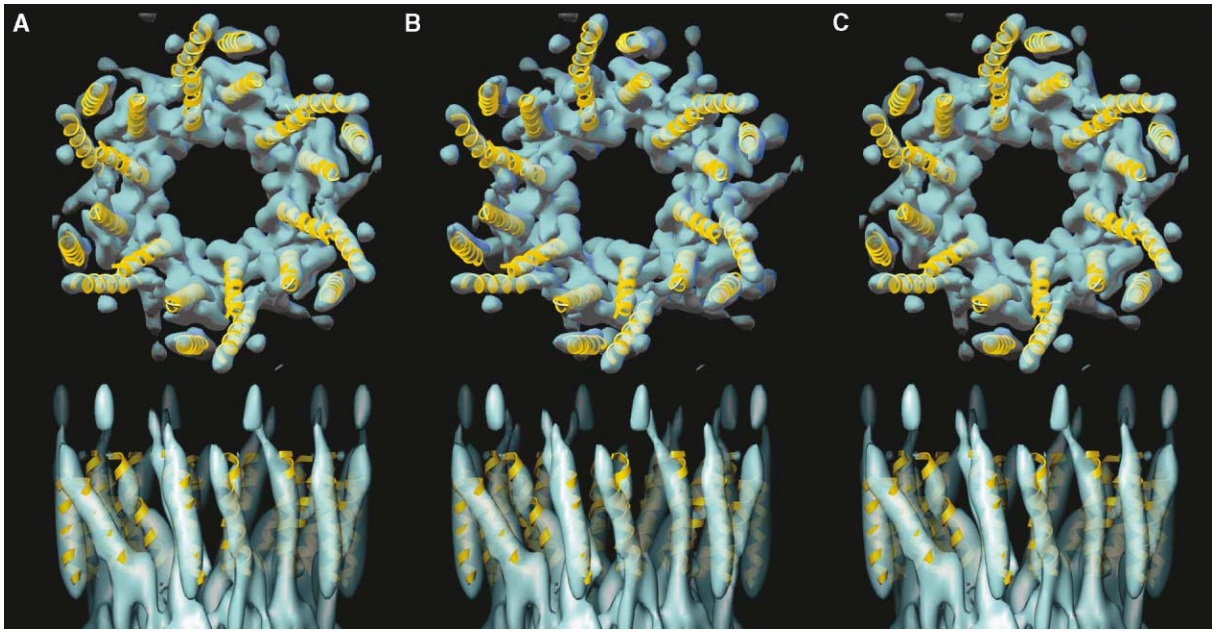

Figure 3. Fit of Canonical α Helices to the Cryo-EM Density Map of the Gap Junction Channel

Top and side views of one connexon showing the fit of canonical α helices (gold) to the cryo-EM density map of Cx43 (blue), according to the helix-axis parameters provided in Table 1. The left and right pairs are wall-eyed and cross-eyed stereo views, respectively.

tioned where they could extend their side chains into the pore lumen (Figure 5B). Arg142 and Glu146 on M3 are only partly pore lining, and interact in part with helix M1, in register with Arg32 of M1. Possibly, the two charged positions of M3, which are one helical turn above each other, form a salt bridge. Being roughly in register with one another, the three charged positions form a thin (4–5 Å) polar belt around the pore lumen about two-thirds of the way from the cytoplasmic to the extracellular ends of the TM domain (Figure 5B). Charged residues in the extracellular loops have been shown to be determinants of charge selectivity in gap junctions (Trexler et al., 2000). It is possible that this polar belt plays a secondary role in charge selectivity.

Roughly in register with one another, a number of conserved polar residues are found throughout the pro-

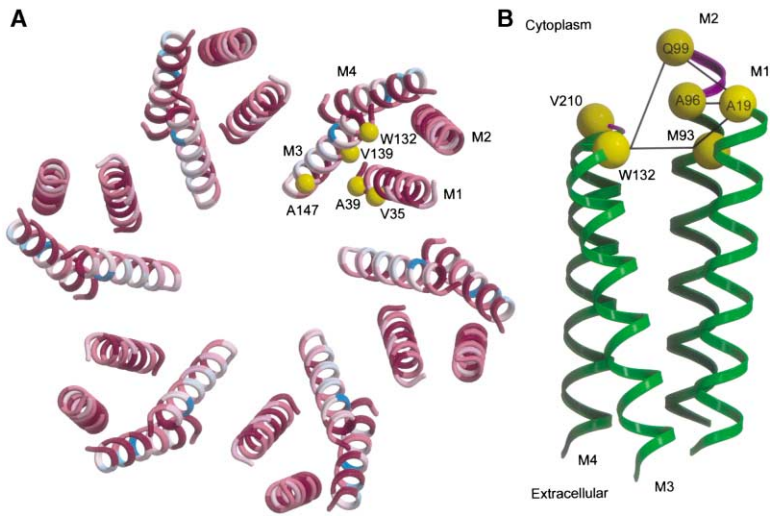

Figure 4. A Model for the Structure of the Gap Junction Connexon

(A) Conservation is color-coded as in Figure 2A. Helices were rotated around their principal axes and evaluated according to a scoring function that (1) favors the burial of conserved and charged amino acids in the protein interior and (2) the exposure of variable positions to the pore lumen or the lipid. Hydrophobic segments M3 and M4 show a clear conservation signal, with a well-defined variable face. Yellow spheres indicate putative specificity determinants, all of which map to pore-lining positions, where they may modulate permeability and conductance. Significantly, specificity determinants span five helical turns on the M3 segment in support of its role as the major pore-lining helix.

(B) M1 and M2 are almost homogeneously conserved (Figures 2A and 4A) and were oriented using a method for the detection of correlated positions (Fleishman et al., 2004a). Positions in the juxtamembrane domain (three positions from the end of the hydrophobic stretch at most) were assumed to extend the α helix (colored magenta). Correlated positions are connected by solid lines. The three correlated pairs of positions on M1 and M2 were assumed to interact, so the helices were rotated manually for these positions to be roughly in proximity. The orientation of M3 around its principal axis was determined solely on the basis of evolutionary conservation (Figure 4A), but the two pairs of correlations between positions on M3 and M2 are congruent with the orientation of M3, serving as partial verification of this helix's orientation around its principal axis. A sixth correlation between Gln99 (M2) and Val210 (M4) could not be reconciled with the model.

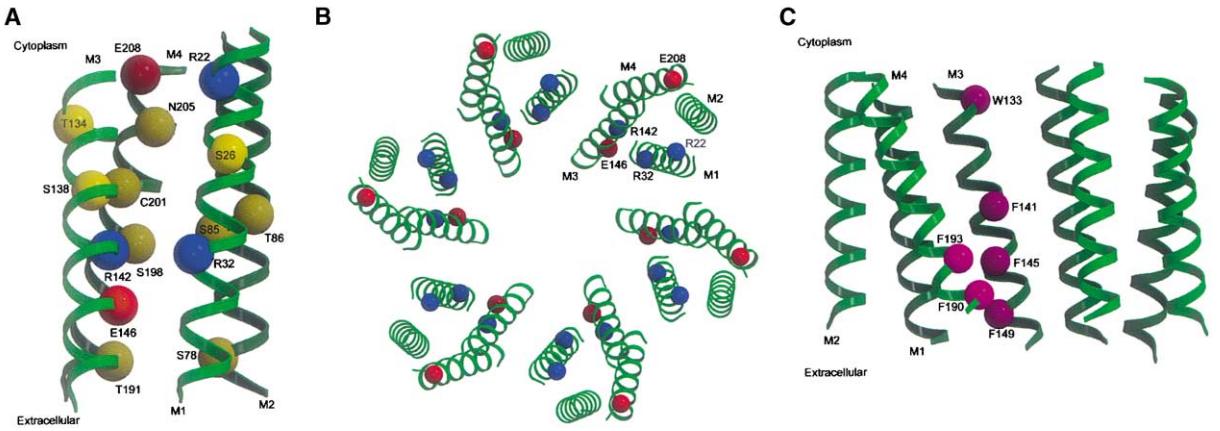

Figure 5. Structural Features of the TM Domain of the Gap Junction Connexon

(A) Polar and charged amino acid residues in the protein interior. The polar residues (yellow spheres) are roughly in register and could be involved in the formation of a network of hydrogen bonds that would stabilize interhelical contacts.

(B) Acidic and basic residues in the protein interior and facing the pore lumen are indicated by red and blue spheres, respectively. Arg22 is near the boundary of the hydrophobic domain and could be accessible to the cytoplasmic side of the membrane (von Heijne, 1989). Glu208 also resides at this boundary and is likely to be exposed to the cytoplasm. The pore-lining charged residues form a slender (4–5 Å) belt of charge around the pore lumen. None of the charged residues is exposed to the membrane.

(C) Aromatic residues on M3 and M4 are shown as purple spheres. The two Phe positions on M4 coincide with the position of a protrusion of density on helix D in the cryo-EM map (Unger et al., 1999). Stacked aromatic residues have been shown to generate such protrusions of density (Henderson et al., 1990). The clustering of aromatic residues from M3 and M4 could stabilize interhelical contacts. Furthermore, the ridge of aromatic residues on M3 could serve to shield the water-filled pore from the lipids in this region of the protein structure, in which helices are not tightly packed.

tein core (yellow spheres in Figure 5A). An attractive hypothesis is that these residues form a hydrogen bonding network to stabilize interhelical contacts. This could explain why many of these positions are intolerant to substitution; even fairly conservative mutations at these positions have been implicated in disease. We note, however, that no terms in the scoring function used to orient the helices around their principal axes favored a particular hydrogen bonding pattern among amino acid residues (Fleishman et al., 2004b).

Significantly, the criteria used for orienting M3 and M4, i.e., evolutionary conservation and hydrophobicity (Fleishman et al., 2004b), did not take into account interactions among aromatic residues. Nevertheless, a prominent structural feature of the model is the clustering of five conserved Phe residues near the extracellular side of the bilayer between helices M3 and M4 (Figure 5C), which may stabilize interhelical contacts. There is also a ridge of aromatic residues on M3 that extends almost without interruption between the extracellular and the intracellular ends of the channel, from Trp133 on the cytoplasmic side to Phe149 on the extracellular side of the bilayer (Figure 5C). Notably, the density map shows that helices C (M3) and B′ (M1) are separated by a relatively large distance (Figure 1). This ridge of aromatic residues could shield the water-filled pore from the lipid.

It is also notable that the previous (Unger et al., 1999) and current cryo-EM maps show a relatively large "shoulder" of density on helix D toward the extracellular side of the gap junction channel. Such protrusions of density can arise from stacked aromatic residues in intermediate-resolution maps (Henderson et al., 1990). A map of bacteriorhodopsin that we computed at the resolution of the gap junction map (5.7 Å in-plane and 19.8 Å vertical) showed a thickening of density corresponding to Phe153 and Phe156 in helix E (data not shown). Although aromatic residues are present in all four TM segments of connexins, only M4 contains two conserved Phe residues near the extracellular side of the bilayer (positions 190 and 193) that occupy the same helical face (magenta circles in Figure 2B). In contrast, helix M2 contains only one aromatic residue (Trp77) in its extracellular part. While it is not an ultimate proof, the interpretation of the shoulder of density on helix D provides independent support for the assignment of helix D to M4 and the orientation of this helix around its principal axis.

Specificity Determinants

Gap junction channels manifest very little ionic selectivity and yet do show differences in ionic preferences between different connexin isoforms. Based on this behavior, one would expect that pore-lining residues would vary among different types of connexins (paralogs) but be conserved for identical connexins in different species (orthologs) (Harris, 2001). Such positions are termed specificity determinants, as their identities determine the specific functional behavior of the given channel.

We analyzed the connexin sequences to identify putative specificity determinants. Connexins of similar functions in different species (orthologs) are the products of speciation events, whereas those with different functions (paralogs) arise from gene duplication (Graur and Li, 1999). It is therefore expected that orthologous sequences would cluster in the termini of the phylogenetic tree, whereas the events leading to paralogy would be reflected in deeper nodes. Thus, using a phylogenetic tree (Yang, 1997) and reconstructed ancestral sequences (Schmidt et al., 2002), we automatically traced the evolutionary history of each amino acid position in

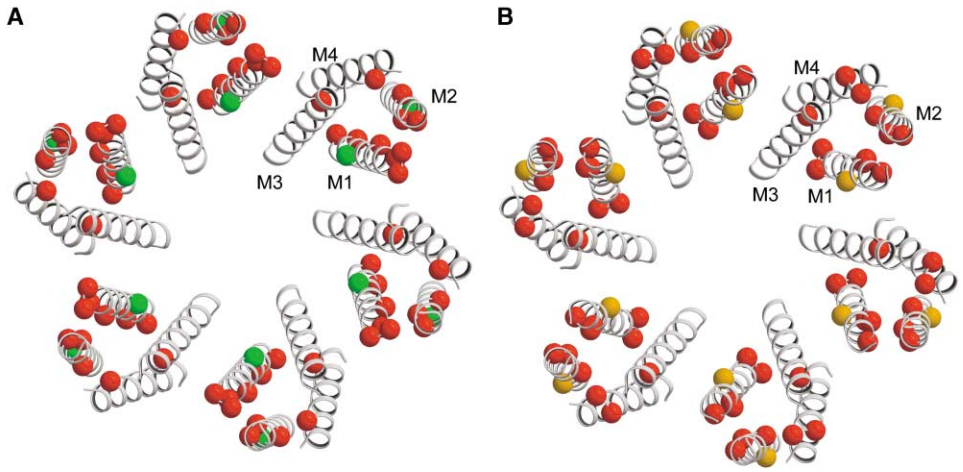

Figure 6. The Distribution of Disease-Causing and Benign Polymorphisms in the Gap Junction Model

(A) The model provides an explanation for the differential effects of mutations that cause nonsyndromic hereditary deafness, erythrokeratodermia variabilis (EKV), and polymorphisms in the TM domain. Physicochemically conservative disease-causing mutations (e.g., Val for Ile) were colored red, and radical polymorphisms (e.g., Ser for Tyr) were colored green. As expected, conservative disease-causing mutations all map to structurally dense regions of the protein, whereas the radical polymorphisms map to more spacious regions.

(B) Similarly, 11 of 13 conservative Charcot-Marie-Tooth (Fischbeck et al., 1999) causing mutations (red spheres) map to structurally packed regions, whereas only two such mutations (orange spheres) map to pore-lining or lipid-exposed helix faces.

search of those that exhibited relatively minor evolutionary differences in the branches separating terminal nodes and large differences in the inner branches (see Experimental Procedures).

We identified five putative specificity determinants on M1 and M3, all of which are pore lining as expected (yellow spheres in Figure 4A). Notably, the putative specificity determinants on M3, the major pore-lining helix, span five helical turns from the cytoplasmic end of the channel, up to roughly two-thirds of the way toward the extracellular side of the bilayer. Since pore-lining positions are expected to specify the different conductance and permeability traits of connexins (Harris, 2001), these results serve as independent verification of our model and predict which residues have important effects on channel properties.

The Locations of Mutations and Polymorphisms

To see whether the model can provide insight on the molecular basis for the effects of mutations that have been reported clinically, we analyzed mutations related to skin, deafness, and developmental diseases that are documented in the Connexin-Deafness Homepage (http://www.crg.es/deafness). The logic underlying our analysis is that mild substitutions such as Val for Ile will cause disease only if they occur in regions of the protein that are structurally well packed. Similarly, radical substitutions such as Ser for Tyr will only be tolerated if they occur in structurally spacious regions.

Figure 6A displays the structure of the gap junction hemichannel with all 11 physicochemically conservative substitutions of Cx26 causing nonsyndromic deafness and erythrokeratodermia variabilis (EKV) in red, and the two radical but benign substitutions (polymorphisms) in green. Strikingly, all mutations indeed map to structurally packed regions, whereas both polymorphisms map to either the pore region or the lipid-exposed face. Analysis of all 13 mild substitutions of Cx32 causing Charcot-Marie-Tooth (CMT) neuropathy (Fischbeck et al., 1999) (Figure 6B) reveals a similar pattern, with only two disease-causing mutations (orange) mapping to spacious regions of the protein structure. A list of all of the mutations shown in Figure 6 is available upon request and at http://ashtoret.tau.ac.il/~sarel/GJ.html.

Discussion

Determining the positions of amino acid residues in the gap junction channel has defied experimental approaches for many years. In part, this is due to the complicated organization of gap junctions when compared to other membrane channels whose structures have been solved. That is, gap junction channels are composed of two connexons in separate membranes and can form different channel varieties depending on the types of connexins that are associated within a connexon.

We used a cryo-EM map of the gap junction channel (Figure 1) to guide the positioning of model α helices in the membrane (Figure 3). The four hydrophobic segments M1–M4 in the connexin sequence were assigned to the helices according to biochemical and evolutionary evidence. The orientation of each of the helices around its principal axis was then computed by analyses of evolutionarily conserved (Figure 4A) and correlated amino acid substitutions (Figure 4B). The resultant conformation placed positions that we identified as specificity determinants in pore-lining locations, as expected (Figure 4A).

We note that the validity of the model is entirely contingent on the assignment of the hydrophobic segments, M1–M4, to the helices A–D in the cryo-EM map (that is, A = M2, B = M1, C = M3, and D = M4). While no single assignment is completely in harmony with all available

biochemical evidence (Harris, 2001), the assignment we have used is compatible with a large body of data.

Several different lines of evidence have converged in the computation and verification of the model. The agreement between these methods is encouraging, but the model should be treated only as an approximation. In fact, there are some inherent inaccuracies in the modeling. For instance, the effective resolution of the cryo-EM map perpendicular to the membrane plane is only 19.8 Å, thereby precluding accurate vertical positioning of the helices. However, the helices are all relatively short, and their tilt angles are fairly small (Table 1). Hence, it is reasonable to position the geometric centers of the helices in the middle of the membrane-spanning part of the cryo-EM map. We note that the correlated pairs of positions are roughly in register (Figure 4B), as are the polar amino acids in the protein core (Figure 5A), serving as support for the positions of the helices' geometric centers.

Another complication is that the limited resolution of the cryo-EM map does not allow us to detect deviations from α-helicity. Nevertheless, the fit of canonical α helices to the cryo-EM map is energetically reasonable (Nunn et al., 2001), and the map does not show any kinks in the TM domain. For comparison, large kinks have been observed in the cryo-EM map of vertebrate rhodopsin at 9 Å in-plane resolution, which still yielded a correct assignment for the positions and orientations of the helices (Baldwin et al., 1997). We cannot rule out the existence of small kinks and bulges at this resolution (Ri et al., 1999), but these would likely have only a local effect on the resultant model (Fleishman et al., 2004b).

The limited vertical resolution of the cryo-EM map also does not reveal the connecting loops between the TM helices, thus precluding the unambiguous assignment of the molecular boundary of each connexin subunit. There are two reasonable subunit boundaries, encompassing either the helices marked as ABCD or A′B′CD in Figure 1. Certainly, more experiments are needed to distinguish these alternatives, and the model provides a detailed structural template for testing these possibilities biochemically. Nevertheless, it is important to note that this ambiguity regarding the connexin subunit boundary is independent of and does not adversely affect the assignment of TM sequences to the helices in the cryo-EM map (i.e., A = M2, B = M1, C = M3, and D = M4).

We are encouraged that the model provides an explanation why substitutions at certain positions can lead to disease (mutations), whereas in other positions, substitutions result in no apparent phenotype (polymorphism). Helices M1 and M2 are considerably more sensitive to mutations than M3 and M4, consistent with the tighter packing of M1 and M2 according to the model (Figure 6). The somewhat higher incidence of mild disease-causing mutations toward the cytoplasmic ends of M1 and M2 coincides with a closer approach of these two helices in this region. We note that sequence conservation alone is not as informative as the model in identifying the portions of the sequences in which substitutions would have deleterious consequences (Figure 6). That is, residues on M1 and M2 are all highly conserved (Figure 2A), but only substitutions in relatively narrow segments on these helices, which are packed in the protein interior, result in disease. Given the striking compatibility of data on mutations and polymorphisms with the model, it appears that the effects of a significant fraction of disease-causing mutations in the TM domain may be explained quite simply in terms of deformations of local structure at the interfaces between helices.

Without a model that explicitly defined amino acid positions, it has been difficult previously to plan rational biochemical experiments. Many studies tested connexin chimeras by swapping large segments from various isoforms (e.g., Hu and Dahl, 1999; Oh et al., 2000; Trexler et al., 2000). While such approaches have provided important insight into broad characteristics, such as charge selectivity and channel permeability, they do not provide an understanding of fine structural and functional details. In recent years, scanning mutagenesis and SCAM provided more detailed information (e.g., Kronengold et al., 2003; Skerrett et al., 2002; Zhou et al., 1997). Nevertheless, without a detailed model, it has not been possible to assess their reliability within one consistent structural framework. Another difficulty in interpreting results from SCAM analyses is that negative results at particular positions (i.e., no labeling) cannot be reliably associated with inaccessibility of these residues. As the labeling reaction depends very strongly on the local environment, neighboring side chains might obstruct accessibility to an otherwise pore-lining position.

The model we describe provides the first integration of a large body of biochemical, mutational, structural, and computational data on the structure of gap junction channels. The model should prove valuable for deriving testable hypotheses related to structure and function. For instance, the model provides certain clues regarding the factors that stabilize interhelical contacts and the determinants of connexin oligomerization. Studies on the roles of the pore-lining positions in affecting channel permeability and selectivity may also be focused with the help of the model, in particular to the residues that we identified as putative specificity determinants (Figure 4A). Moreover, the model can guide studies on the folding of individual connexins and their association to form connexons. A fascinating prediction of the model is that the phenotypic effects of a disease-causing mutation on one helix can be rescued by a substitution on a neighboring helix.

Experimental Procedures

Electron Cryomicroscopy and Image Analysis

Preparation of two-dimensional crystals, cryo-EM, and lattice straightening were performed as described before (Unger et al., 1999). A list that contained the data from 69 crystalline areas was edited to exclude measurements where the sampling of reciprocal space was too sparse to allow a meaningful fit of lattice lines. The final fit was limited to a maximum z* value of 0.065 $Å^{-1}$ generating 1734 unique structure factors compared to 1022 that were included in the previous reconstruction (Unger et al., 1999). Using image data with signal-to-noise ratios $\geq$1.8, the overall merging phase residual for each crystal was $<25^{\circ}$ compared with the entire data set. The 3D map was computed using an inverse B factor of -350. Analysis of the point-spread function indicated a maximum in-plane resolution of 5.7 Å and a vertical resolution of 19.8 Å.

Sequence Data

60 connexin sequences were obtained from SWISS-PROT (Bairoch and Apweiler, 2000) and aligned using CLUSTAL W (Thompson et

al., 1994) with default parameters. For each position in the alignment, evolutionary conservation was computed using the *ConSeq* server (Figure 2A) (Berezin et al., 2004), and hydrophobicity using the Kessel and Ben-Tal scale (Kessel and Ben-Tal, 2002).

The topology of Cx32 was determined experimentally (Milks et al., 1988). Definition of the N- and C termini of the four TM segments (Bennett et al., 1994) was adjusted slightly to include hydrophobic stretches that were as long as possible. That is, we eliminated positions from the hydrophobic segments' termini that were occupied by polar or charged amino acids in any of the sequences in the multiple-sequence alignment of 60 homologs. The resulting topology and boundaries of the hydrophobic stretches are shown in Figures 2A and 2B.

Scoring Function

The conformational search was performed using the scoring function described by Fleishman et al. (2004b). In brief, this scoring function favors the burial of evolutionarily conserved amino acid positions in the protein core and the exposure of variable positions to the lipid or the pore. Conformations that expose charged amino acids to the lipid milieu are penalized. Since the gap junction pore is relatively large, pore-lining and lipid-exposed residues were treated equally as unburied positions, with no need for introducing modifications to the functions. However, since charged residues can be exposed to the lumen of the pore with no consequence on desolvation energy, we abolished the penalty for exposure of charged positions on the pore-lining helices M1 and M3 (Figure 5B). Each conformation was scored according to the following equation:

$$Score = \sum_i (2(B^i - \tfrac{1}{2})(H^i - C^i)), \quad (1)$$

where B^i quantifies the extent of burial of amino acid i in the protein core (Fleishman and Ben-Tal, 2002). It assumes values of 0 to 1; 1 signifying complete burial against another helix, and 0 complete exposure to the lipid or the pore lumen. The function is computed by iterating over all of the helices in the structure other than the one on which i is located, and taking into account i's distance from, and orientation with respect to, each of these helices. B^i is then taken as the maximum of the values calculated for each of the helices (Fleishman and Ben-Tal, 2002; Fleishman et al., 2004b). Thus, high values of B^i imply that i is in close contact with another helix, whereas low values indicate that it is not interacting with any of the helices.

The C^i values are the normalized evolutionary-rate scores assigned by *Rate4Site* (Figure 2A) (Berezin et al., 2004; Pupko et al., 2002). High-through-low values of C^i are assigned to variable-through-conserved positions, respectively. Proline residues are ignored in calculating the conservation scores, as they are often conserved due to kinks they induce in the helix secondary structure rather than due to the formation of interhelical contacts (Baldwin et al., 1997).

H^i is the free energy of transfer from water to lipid of amino acid i according to the Kessel and Ben-Tal scale (Kessel and Ben-Tal, 2002). H^i values are taken into account only if they are greater than 7 kcal/mole, and only for residues i that are exposed to the membrane, i.e., for which the burial scores B^i are less than 0.5. Thus, the hydrophobicity scale serves as a significant penalty on the exposure of the most polar residues to the membrane environment. The terminal turns (4 amino acid residues) from each side of the TM segments were ignored in computing this penalty, since the polar environment at the lipid-water interface could accommodate these residues (von Heijne, 1989).

Conformational Search

Canonical C^{α}-trace models of four α helices were constructed according to the helix-axes parameters derived from the cryo-EM map (Table 1), and their geometric centers were placed at the hypothetical membrane midplane. The amino acid identities of positions in the hydrophobic segments M1–M4 were assigned to the relevant positions on these helices. The channel's axis of symmetry was inferred from the map (Figure 1), and 6-fold symmetry around this axis was strictly maintained throughout all conformational searches. Hence, only the rotation angles around the principal axes of each of four helices comprising a single connexin were explored, and applied to all 24 helices.

Each helix was rotated around its principal axis independently, in 5° steps, and its optimal orientation was derived. Then, the optimal orientations of all helices were superimposed to yield the optimal conformation of the entire complex.

Correlated Mutations

The multiple-sequence alignment of 60 connexin homologs was used to compute the phylogenetic tree of maximum-likelihood (Schmidt et al., 2002). Subsequently, the most likely ancestral (now-extinct) sequences were inferred (Yang, 1997). We then identified correlated positions in the TM domain of connexins (Fleishman et al., 2004a) (Figure 4B). The informational-entropy threshold (Shannon, 1948), which is a measure of the heterogeneity of amino acid identities in a particular position in the alignment, was set to 1.1 in order to remove highly conserved positions. To obtain confidence intervals for each of the computed correlations, 400 bootstrap iterations (Bradley and Tibshirani, 1993) with replacement were conducted. The lower (r_{low}) and upper (r_{high}) boundaries of the 95% confidence interval were determined as the correlation coefficient at the 2.5 and the 97.5 percentiles, respectively, and the trimmed mean (r) of correlation coefficients was calculated. Pairs of positions showing lower confidence boundaries of $r_{low} < 0.1$ were eliminated as were pairs with trimmed means of $r < 0.5$.

Specificity Determinants

The phylogenetic tree and ancestral-sequence reconstruction (see Correlated Mutations, above) were used to detect putative specificity determinants in the connexin family. Conserved positions in the sequence alignment exhibiting information entropy (Shannon, 1948) of less than 1.1 were eliminated. For each position in the alignment, and in each phylogenetic branch, we measured the physicochemical distance between the amino acid identities occupying those positions using the Miyata substitution matrix (Miyata et al., 1979). Multiple and back substitutions in a single branch were not considered. Each node in the phylogenetic tree was assigned a "depth" value, which was an integer calculated as the minimal distance between that node and any terminus, counting intervening nodes. Thus, the termini were assigned depth values of 0, neighboring nodes values of 1, etc.

For each amino acid position, we then computed the Pearson correlation coefficient between physicochemical distances traversed in each phylogenetic branch and the average depths of each of the nodes that were connected by that particular branch. Hence, high correlation coefficients were associated with positions that exhibited relatively low variability among terminal nodes (orthologs) and relatively high variability in deeper nodes (separating paralogs). We conducted 400 bootstrap iterations with replacement (Bradley and Tibshirani, 1993), and calculated the trimmed mean of the 95% confidence interval of these correlation values (r). The lower (r_{low}) and upper (r_{high}) bounds of the 95% confidence interval were determined as the correlation coefficient at the 2.5 and the 97.5 percentiles, respectively. Positions showing lower confidence bounds of $r_{low} < 0$ were eliminated as were positions with trimmed means $r < 0.1$.

Acknowledgments

The authors thank Michael Bennett, Ted Bargiello, and Vyto Verselis for helpful discussions. This study was supported by a Research Career Development Award from the Israel Cancer Research Fund to N.B.-T. S.J.F. was supported by a doctoral fellowship from the Clore Israel Foundation. During part of this work, V.M.U. was supported by a postdoctoral fellowship from the American Heart Association. M.Y. was supported by NIH grant RO1HL48908 and a Clinical Scientist Award in Translational Research from the Burroughs-Wellcome Fund. Additional information is available on request and at http://ashtoret.tau.ac.il/~sarel/GJ.html.

Received: March 26, 2004
Revised: July 6, 2004
Accepted: July 22, 2004
Published: September 23, 2004

References

Bairoch, A., and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. *28*, 45–48.

Baldwin, J.M. (1993). The probable arrangement of the helices in G protein-coupled receptors. EMBO J. *12*, 1693–1703.

Baldwin, J.M., Schertler, G.F., and Unger, V.M. (1997). An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. J. Mol. Biol. *272*, 144–164.

Bennett, M.V., Zheng, X., and Sogin, M.L. (1994). The connexins and their family tree. Soc. Gen. Physiol. Ser. *49*, 223–233.

Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, R., Cassadio, R., and Ben-Tal, N. (2004). ConSeq: the identification of functionally and structurally important residues in protein sequences. Bioinformatics, in press.

Bourne, H.R., and Meng, E.C. (2000). Structure. Rhodopsin sees the light. Science *289*, 733–734.

Bradley, E., and Tibshirani, R. (1993). An Introduction to the Bootstrap (New York: Chapman and Hall).

Cascio, M., Kumar, N.M., Safarik, R., and Gilula, N.B. (1995). Physical characterization of gap junction membrane connexons (hemi-channels) isolated from rat liver. J. Biol. Chem. *270*, 18643–18648.

Fischbeck, K.H., Abel, A., Lin, G.S., and Scherer, S.S. (1999). X-linked Charcot-Marie-Tooth disease and connexin32. Ann. N Y Acad. Sci. *883*, 36–41.

Fleishman, S.J., and Ben-Tal, N. (2002). A novel scoring function for predicting the conformations of tightly packed pairs of transmembrane alpha-helices. J. Mol. Biol. *321*, 363–378.

Fleishman, S.J., Yifrach, O., and Ben-Tal, N. (2004a). An evolutionarily conserved network of amino acids mediates gating in voltage-dependent potassium channels. J. Mol. Biol. *340*, 307–318.

Fleishman, S.J., Harrington, S., Friesner, R.A., Honig, B., and Ben-Tal, N. (2004b). An automatic method for predicting the structures of transmembrane proteins using cryo-EM and evolutionary data. Biophys. J., in press.

Gobel, U., Sander, C., Schneider, R., and Valencia, A. (1994). Correlated mutations and residue contacts in proteins. Proteins *18*, 309–317.

Graur, D., and Li, W.H. (1999). Fundamentals of Molecular Evolution, Second Edition (Sunderland, MA: Sinauer Associates).

Harris, A.L. (2001). Emerging issues of connexin channels: biophysics fills the gap. Q. Rev. Biophys. *34*, 325–472.

Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., and Downing, K.H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. J. Mol. Biol. *213*, 899–929.

Hu, X., and Dahl, G. (1999). Exchange of conductance and gating properties between gap junction hemichannels. FEBS Lett. *451*, 113–117.

Karlin, A., and Akabas, M.H. (1998). Substituted-cysteine accessibility method. Methods Enzymol. *293*, 123–145.

Kelsell, D.P., Dunlop, J., and Hodgins, M.B. (2001). Human diseases: clues to cracking the connexin code? Trends Cell Biol. *11*, 2–6.

Kessel, A., and Ben-Tal, N. (2002). Free energy determinants of peptide association with lipid bilayers. In Current Topics in Membranes, S. Simon and T. McIntosh, eds. (San Diego: Academic Press), pp. 205–253.

Kronengold, J., Trexler, E.B., Bukauskas, F.F., Bargiello, T.A., and Verselis, V.K. (2003). Single-channel SCAM identifies pore-lining residues in the first extracellular loop and first transmembrane domains of Cx46 hemichannels. J. Gen. Physiol. *15*, 389–405.

Kumar, N.M., and Gilula, N.B. (1996). The gap junction communication channel. Cell *84*, 381–388.

Milks, L.C., Kumar, N.M., Houghten, R., Unwin, N., and Gilula, N.B. (1988). Topology of the 32-kd liver gap junction protein determined by site-directed antibody localizations. EMBO J. *7*, 2967–2975.

Miyata, T., Miyazawa, S., and Yasunaga, T. (1979). Two types of amino acid substitutions in protein evolution. J. Mol. Evol. *12*, 219–236.

Nunn, R.S., Macke, T.J., Olson, A.J., and Yeager, M. (2001). Transmembrane α-helices in the gap junction membrane channel: systematic search of packing models based on the pair potential function. Microsc. Res. Tech. *52*, 344–351.

Oh, S., Abrams, C.K., Verselis, V.K., and Bargiello, T.A. (2000). Stoichiometry of transjunctional voltage-gating polarity reversal by a negative charge substitution in the amino terminus of a connexin32 chimera. J. Gen. Physiol. *116*, 13–31.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., et al. (2000). Crystal structure of rhodopsin: a G protein-coupled receptor. Science *289*, 739–745.

Pupko, T., Bell, R.E., Mayrose, I., Glaser, F., and Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics *18*, S71–S77.

Ri, Y., Ballesteros, J.A., Abrams, C.K., Oh, S., Verselis, V.K., Weinstein, H., and Bargiello, T.A. (1999). The role of a conserved proline residue in mediating conformational changes associated with voltage gating of Cx32 gap junctions. Biophys. J. *76*, 2887–2898.

Schmidt, H.A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics *18*, 502–504.

Shannon, C.E. (1948). A mathematical theory of communication. Bell Sys. Tech. J. *27*, 379–423, 623–656.

Skerrett, I.M., Aronowitz, J., Shin, J.H., Cymes, G., Kasperek, E., Cao, F.L., and Nicholson, B.J. (2002). Identification of amino acid residues lining the pore of a gap junction channel. J. Cell Biol. *159*, 349–360.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

Trexler, E.B., Bukauskas, F.F., Kronengold, J., Bargiello, T.A., and Verselis, V.K. (2000). The first extracellular loop domain is a major determinant of charge selectivity in connexin46 channels. Biophys. J. *79*, 3036–3051.

Unger, V.M., Hargrave, P.A., Baldwin, J.M., and Schertler, G.F. (1997). Arrangement of rhodopsin transmembrane α-helices. Nature *389*, 203–206.

Unger, V.M., Kumar, N.M., Gilula, N.B., and Yeager, M. (1999). Three-dimensional structure of a recombinant gap junction membrane channel. Science *283*, 1176–1180.

von Heijne, G. (1989). Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. Nature *341*, 456–458.

von Heijne, G. (1996). Principles of membrane protein assembly and structure. Prog. Biophys. Mol. Biol. *66*, 113–139.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. *13*, 555–556.

Yeager, M., and Gilula, N.B. (1992). Membrane topology and quaternary structure of cardiac gap junction ion channels. J. Mol. Biol. *223*, 929–948.

Zhou, X.W., Pfahnl, A., Werner, R., Hudder, A., Llanes, A., Luebke, A., and Dahl, G. (1997). Identification of a pore lining segment in gap junction hemichannels. Biophys. J. *72*, 1946–1953.

Accession Numbers

The coordinates of the C^{α} model have been deposited in the Protein Data Bank (accession code 1TXH).