Major Facilitator Superfamily Porters, LacY, FucP and XyIE of *Escherichia coli* Appear to Have Evolved Positionally Dissimilar Catalytic Residues without Rearrangement of 3-TMS Repeat Units

Åke Västermark\(^a\)  Bryan Lunt\(^b\)  Milton Saier\(^a\)

\(^a\)Department of Molecular Biology and \(^b\)Center for Theoretical Biological Physics, University of California at San Diego, La Jolla, Calif., USA

Abstract

Based on alleged functional residue correspondences between FucP and LacY, a recent study has resulted in a proposed model of 3-TMS unit rearrangements [Madej et al.: Proc Natl Acad Sci USA 2013; 110:5870–5874]. We rebut this theory, using 7 different lines of evidence. Our observations suggest that these two transporters are homologous throughout their lengths, having evolved from a common ancestor without repeat unit rearrangements. We exploit the availability of the high-resolution XyIE crystal structures in multiple conformations including the inward-facing state to render possible direct comparisons with LacY. Based on a distance map, we confirm the conclusion of Quistgaard et al. [Nat Struct Mol Biol 2013; 20:766–768] that the N-terminal 6 TMS halves of these transporters are internally less mobile than the second halves during the conformational transition from the outward occluded state to the inward occluded state and inward occluded state to inward open state. These observations, together with those of Madej et al. [2013], lead to the suggestion that functionally equivalent catalytic residues involved in substrate binding and transport catalysis have evolved in dissimilar positions, but apparently often in similar positions in the putative 3-TMS repeat units, from a single structural scaffold without intragenic rearrangement.

Introduction

The high-resolution structure of the xylose proton symporter, XyIE, of *Escherichia coli* (TC No. 2.A.1.1.3) was recently published, making it the first structurally determined major facilitator superfamily (MFS) transporter to be available in multiple conformations [Quistgaard et al., 2013; Sun et al., 2012]. The states elucidated by these investigators were the occluded outward, occluded inward and open inward conformations of XyIE.

MFS uni-, sym- and antiporters are believed to operate via an alternating-access single binding site mechanism that involves a primary ‘rocker-switch’ type movement of the two halves of the protein relative to each other. MFS transporters exhibit a pseudosymmetrical 6 + 6 TMS repeat structure with 2 extra TMS between the 2 repeat units in some members of the superfamily [Henderson, 1991; Henderson and Maiden, 1990; Marger and Saier, 1993; Reddy et al., 2012].
In a recent paper [Madej et al., 2013], the three-dimensional (3-D) structures of two members of the MFS, the fucose porter, FucP (TC No. 2.A.1.7.1) [Dang et al., 2010], and the lactose porter, LacY (TC No. 2.A.1.5.1) [Guan et al., 2007], both of \textit{E. coli}, were compared, where functional residue correspondences suggested to the authors that the architectures of LacY and FucP consisted of building blocks of 3-TMS units arranged in the opposite order: N-terminus A(1–3)-B(4–6)-C(7–9)-D(10–12)-C-terminus for LacY versus N-terminus D′(10–12)-C′(7–9)-B′(4–6)-A′(1–3)-C-terminus for FucP, with numbers of TMS corresponding to those in LacY. This suggestion is intriguing and requires that more than 1 rearrangement must have occurred between putative 3-TMS repeat units since LacY and FucP diverged from their common evolutionary precursor protein. This proposal leaves the mechanism of evolution open to speculation – in particular, whether the proposed 3-TMS unit rearrangement actually took place, and if so, whether or not the catalytic residues evolved before or after the proposed 3-TMS unit rearrangement took place.

Figure 1 shows a schematic view of XylE (modeled on FucP) with the proposed 3-TMS repeat unit rearrangement suggested by Madej et al. [2013] relative to that of LacY, where the Roman numerals refer to the TMS in LacY. Thus, the four 3-TMS repeat units are of opposite order in XylE or FucP relative to LacY, according to the ‘mix-and-match’ hypothesis of Madej et al. [2013]. In HMM comparisons, XylE and FucP are more similar to each other than either protein is to LacY, although all these proteins are clearly homologous to each other [Pao et al., 1998; Yen et al., 2010].

Examination of the proposal of Madej et al. [2013] was complicated by the fact that the relevant high-resolution structures of FucP and LacY are available in dissimilar conformations [Dang et al., 2010; Guan et al., 2007]. In LacY, over 100 residues that have an impact on transport have been identified, but the residue correspondences in FucP, used to argue in favor of the proposed 3-TMS unit rearrangement, are limited to about 15 residues, an important one being an aspartate that influences counterflow [Madej et al., 2013].

In this report, we use multiple lines of evidence to argue against the proposed 3-TMS unit rearrangement theory proposed by Madej et al. [2013]. The evidence includes: (1) maximum bipartite matching of HMM:HMM comparisons; (2) superimposition of the inward open conformations of LacY and XylE (which, in contrast to FucP, are both available in the same conformation); (3) percent identity comparisons between LacY, FucP and XylE 3-TMS units using the Global Sequence Alignment Tool (GSAT) [Reddy and Saier, 2012]; (4) motif analyses using the Motif Expectation Maximization for Motif Elicitation (MEME) program [Bailey and Elkan, 1994]; (5) structural examination of FucP and LacY to search for characteristic hallmarks; (6) positive inside rule anal-
yses [von Heijne, 1992], and (7) use of a Δdistance map. This last approach led to the suggestion that the 6-TMS unit has functional importance during the conformational changes from the outward partially occluded state to the inward partially occluded state, and from the inward occluded state to the inward open state, but the proposed 3-TMS repeat units do not appear to move as independent units. The results do not argue against the notion of a primordial 3-TMS ancestral precursor unit which duplicated twice to give the present day 12 TMS MFS permeases, but they argue against the idea of the proposed 3-TMS unit rearrangements. Integrating our results, taken together with the observations of Madej et al. [2013] concerning the positions of functional residues in LacY and FucP, leads to the possibility that similar residues may have evolved at corresponding positions in dissimilar 3-TMS units of MFS permeases to give rise to substrate binding sites for a variety of dissimilar substrates.

**Results**

**Maximum Weight Perfect Bipartite Matching**

The maximum weight perfect bipartite matching (see Methods for details) for LacY and FucP, based on pairwise HMM:HMM alignment scores, is reported in table 1. We chose maximum weight perfect bipartite matching because each domain appears once and only once. The null hypothesis involves complete duplication without rearrangement, while the Madej hypothesis involves a bipartite matching involving complete duplication with multiple rearrangements. Comparing these two hypotheses, the scores in table 1 contradict the Madej hypothesis. Thus, the scores for 3-TMS unit matching (table 1) agree more with the nonrearranged (FucP-1 = LacY-1, FucP-2 = LacY-2, FucP-3 = LacY-3, FucP-4 = LacY-4) than with the rearranged Madej hypothesis of FucP-1 = LacY-4, FucP-2 = LacY-3, FucP-3 = LacY-2, FucP-4 = LacY-1. For example, the former arrangement gave the highest score (123), while the corresponding score for the Madej hypothesis was only 84. This shows that the null hypothesis is strongly favored.

**Superimposition of the Inward Open Configurations of XylE and LacY**

Figure 2a presents the structure of XylE in the open inward-facing conformation using the ‘pipes’ (for α-helices) convention. We found a remarkable coincidence of the structures of LacY and XylE when they were superimposed on each other (fig. 2b), particularly considering how divergent the sequences of these two proteins are. Their very significant degree of 3-D coincidence further supports our contention that these two proteins did not undergo rearrangement of their 3-TMS units during their evolutionary divergence from the common ancestral sequence.

**Table 1. Maximum weight bipartite matching scores supporting the null hypothesis**

<table>
<thead>
<tr>
<th>Unit</th>
<th>LacY-1</th>
<th>LacY-2</th>
<th>LacY-3</th>
<th>LacY-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>FucP-1</td>
<td>56.8&lt;sup&gt;a&lt;/sup&gt;</td>
<td>18.0&lt;sup&gt;d&lt;/sup&gt;</td>
<td>35.3&lt;sup&gt;c&lt;/sup&gt;</td>
<td>31.6&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>FucP-2</td>
<td>15.4&lt;sup&gt;d&lt;/sup&gt;</td>
<td>21.5&lt;sup&gt;a&lt;/sup&gt;</td>
<td>11.5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>19.9&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>FucP-3</td>
<td>24.2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>16.0&lt;sup&gt;b&lt;/sup&gt;</td>
<td>23.7&lt;sup&gt;a&lt;/sup&gt;</td>
<td>15.2&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>FucP-4</td>
<td>25.3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>16.0&lt;sup&gt;d&lt;/sup&gt;</td>
<td>12.6&lt;sup&gt;d&lt;/sup&gt;</td>
<td>21.1&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

The top row refers to the 4 proposed 3-TMS units of LacY, while the left column corresponds to the 4 proposed 3-TMS units of FucP. Thus, unit LacY-1 corresponds to TMS 1–3 in LacY, while unit FucP-1 corresponds to TMS 1–3 in FucP. Note that units 1 and 3, and units 2 and 4, are known to be homologous [Pao et al., 1998].

- <sup>a</sup> Combinations involving no rearrangements (sum of 123; null model).
- <sup>b</sup> Rearrangement proposed by Madej et al. [2013] (sum of 84; ‘mix-and-match’ model).
- <sup>c</sup> Homologous 3-TMS repeat units, assuming the well-established 6-TMS duplication for all MFS permeases (sum of 95).
- <sup>d</sup> Nonhomologous putative 3-TMS repeat units (sum of 61).
Conserved Motif Analyses

Fucose:H⁺ symporter (FHS; FucP) family members (TC No. 2.A.1.7) in the Transporter Classification Database are more numerous and diverse than oligosaccharide:H⁺ symporter (LacY) family members (TC No. 2.A.1.5). A well-known motif in the cytoplasmic loops between TMS 2–3 and TMS 8–9 (K-X3-K and R-X3-K, respectively) was located in both FucP and LacY. However, the motif in the second half of FHS family proteins is usually interrupted with an approximately 5-residue-long insertion in most of the sequences. The conservation of the R in this second repeat is only 42%, and that of the K only 36%. There is a significantly better-conserved R (63%) about 8 residues before the R of the motif, but that location appears to be inside TMS 8, toward the cytoplasmic side. In the alignment of FHS family members, the former 2 residues are the 2 well-conserved motifs of lysines and arginines that can be found in the cytoplasmic loops of close homologues of FucP, always between TMS 2 and 3 as well as TMS 8 and 9.

Motifs were examined for the OHS and FHS families using the MEME program [Bailey and Elkan, 1994]. This program was run using a 5- to 10-residue width setting with the maximal number of motifs set at 5, allowing any number of repetitions. For the OHS family, we used 4 sequences (TC No. 2.A.1.5.1–4). The R-X3-K motif was contained in a MEME motif, having the sequence [SI][DN][RK][IL]GLK[KN]TL (alternative residues at single positions are in brackets). This motif is found in all 4 OHS family sequences, at residue positions 70–80 and 290–300 in the multiple alignment, corresponding to the known loops between TMS 2 and 3, and TMS 8 and 9. In an effort to study how the E values and detection accuracy would be affected by a narrower motif definition, we set the acceptable motif length range to 3–6 and repeated the exercise with the same settings. The motif was represented by [DN][RK][IL]GLK.

The exercise was repeated with the same settings for the FHS family (19 sequences; TC No. 2.A.1.7.1–19), ini-
tially searching for a 5- to 10-residue-long motif. The motif between TMS 2 and 3 was identified and was [KR][IL][GS]YK. The motif expected to be present in TMS 8–9 did not show up in MEME analyses using these settings. It is well known that this motif between TMS 8 and 9 is, in general, less well conserved than that between TMS 2 and 3 [Henderson, 1991; Henderson and Maiden, 1990; Pao et al., 1998].


Figure 3 shows the schema of the second halves of FucP and LacY together with parts of the interdomain linkers between TMS 7 and 8. It can be seen that there is a characteristic ‘kink’ shared by the 2 proteins. These would not be in related positions assuming the ‘mix-and-match’ model of Madej et al. [2013].

Distribution of Positively Charged Residues (R and K) in FucP and LacY

Figure 4 shows the positions of residues with a formal positive charge (arginines and lysines) within FucP and LacY. It can be seen that the ‘positive inside rule’ [von Heijne, 1992] is strictly followed in both proteins. If the proteins evolved via the ‘mix-and-match’ pathway, positively charged residues would have had to have been eliminated and replaced in different positions. The null hypothesis does not require such modifications.

Δdistance Mapping

Δdistance mapping was conducted as described in the Methods section. These results, depicted in figure 5a, compare the 2 occluded states (inward facing and outward facing) of XylE. This figure shows the map which compares pairwise distances between α-carbon atoms, in corresponding positions. Figure 5b reveals the same for the inward occluded and inward open configurations. Figure 5c shows a magnification of the C-terminal 6 TMS taken from figure 5b. These results suggest that during the conformational transitions between these 3 conformers, the primary internal distance remodelings occur within TMS 7–12, but not within TMS 1–6, although it should be noted that the transition between the outward occluded and outward open configurations could not be examined. Thus, in compliance with the suggestion of Quistgaard et al. [2013] and based on the available evidence, we concur that the primary intradomain motions during catalysis of transport are largely restricted to the second half of the protein, at least for the conformations currently available.

Discussion and Conclusions

Several observations are difficult to reconcile with the ‘mix-and-match’ hypothesis of Madej et al. [2013], but are easily explained by the null hypothesis that assumes that 3-TMS units in these proteins did not involve rearrangement during their evolutionary divergence:
(1) HMM:HMM comparisons suggested that FucP and XylE are more similar to each other than to LacY, although this was not obvious from the percent identity values or a phylogenetic tree generated with the SuperFamilyTree programs [Chen et al., 2011; Yen et al., 2009, 2010]. Maximum weight bipartite matching [Galil, 1986], training HMMs on 3-TMS units excised from alignments of close homologues of FucP and LacY, suggested homology throughout the lengths of the two proteins without rearrangement (table 1).

(2) Because the 3-D structures of XylE are available in 3 conformations including the inward open conformation analogous to the only available conformation of LacY, we used XylE as an inward open ‘stand-in’ for FucP. The inward open conformations of XylE (4JA4) and LacY (2V8N) superimposed such that the TMS matched throughout the lengths of both proteins from TMS 1 to 12 (fig. 2).

(3) FucP and XylE identity scores using the native proteins suggested that these two transporters are homologous throughout their lengths (table 2b).

(4) Conserved residues in homologues of LacY and FucP were present in motifs, particularly the MFS-characteristic motifs within loops of TM 2–3 and 8–9, represented by [RK][IL]GLK or [KR][IL]GSYK in the two sets of homologues. Their positions are difficult to reconcile with the ‘mix-and-match’ pathway of Madej et al. [2013]. The latter would require that both motifs would have to have been lost and reappear at another position via convergent evolution, a highly unlikely possibility considering the largely uniform sequence of this motif.

(5) In both FucP and LacY, between TMS 6 and 7, a characteristic well-conserved ‘kink’ precedes TMS 7 (fig. 3). The ‘mix-and-match’ hypothesis does not explain this conserved structure.

(6) Rearranging LacY relative to FucP, using the proposed rearrangement model, would necessitate switching of positively charged loops between the sides of some 3-TMS units to maintain the positive inside rule (fig. 4). Again, while not impossible, this occurrence is
The Δdistance map (fig. 5) indicated that the N-terminal bundle of 6 TMS in XylE is internally immobile during the conformational transition from the outward occluded state to the inward occluded state, and also from the inward occluded state to the inward open state. In the file 4GBY-4JA3.xlsx, available on request, a spreadsheet is provided for numerical exploration. This suggests that the halves, rather than the 3-TMS units, may be the basic functional units. Since this suggestion of necessity is based on studies with just one protein, XylE, and since the 3-D structure of XylE is not available in the outward open configuration, it is premature to conclude that this characteristic is generally valid for all conformational states and applies to MFS porters in general. Further studies will be required to establish the applicability of this interesting observation to the mechanism(s) of transport by the many members of the MFS [Chang et al., 2004]

According to Occam’s razor, the principle of parsimony states that among competing hypotheses, the one with the fewest assumptions should be selected. Considering the razor, all evidence points to the null hypothesis. In fact, most of the available evidence argues against the 3-TMS repeat unit rearrangement proposed by Madej et al. [2013], which was based solely on putative correspondences between functionally important residues.

Taking the interesting observations of Madej et al. [2013] into account, we suggest that the functionally important residues involved in substrate and proton recognition as well as catalysis may have evolved late, after the formation of the basic master plan of the stable MFS fold. This is reasonable in view of the many dissimilar substrates transported by MFS porters.

In unpublished phylogenetic studies, we have obtained preliminary evidence that not all MFS permeases duplicated their 6 TMS units at the same time, suggesting that the 12-TMS framework was not transmitted intact through evolutionary history without at least some loss and gain of one or the other of the two duplicated halves in some MFS families. However, since 6-TMS homodimeric or heterodimeric MFS permeases have not been detected in nature, we suspect that the proposed late duplication events occurred with starting material that had 12 TMS, and that a duplication of one of these 6-TMS units occurred either simultaneously with, or subsequently to, loss of the redundant repeat unit. However, these suggestions are still highly speculative.

While replacement of one of the two 6-TMS halves of a permease might reasonably have occurred, translocation of 3-TMS units to give rise to inversion of 3-TMS units with a change in order, as suggested by Madej et al. [2013] in their comparison of LacY with FucP, seems highly improbable as it requires multiple rearrangements. This postulate is certainly contrary to the data presented in this communication, which evaluates the proteins globally rather than examining functional residues. Further studies, currently in progress, will be required to establish the detailed pathways that gave rise to the tremendous sequence diversity with retention of the basic common fold that characterizes the widely distributed members of the MFS.

Using the GSAT (http://tcdb.org/progs/?tool=gsat), we have calculated percent identity values for the 4 putative 3-TMS building blocks of LacY, FucP and XylE. The percent identity values for the rearranged and nonrearranged orders were 45.5 (‘mix-and-match’ model) and 64.1 (null model) for the LacY/FucP comparisons, and 40.1 (‘mix-and-match’ model) and 52.8 (null model) for the FucP/XylE comparisons, respectively. These values favor the nonrearranged configuration.

| Table 2. GSAT percent identity results supporting the null hypothesis |
|------------------|------------------|------------------|------------------|
|                  | Percent identity |                  |                  |
|                  | LacY-1           | LacY-2           | LacY-3           | LacY-4           |
| Unit             |                  |                  |                  |                  |
| FucP-1           | 12.8a            | 8.1              | 21.4             | 9.3b             |
| FucP-2           | 5.6              | 10.4a            | 13.9b            | 4.0              |
| FucP-3           | 8.2              | 1.1b             | 20.0a            | 13.8             |
| FucP-4           | 21.2b            | 10.2             | 14.0             | 20.9a            |

|                  |                  |
|------------------|------------------|------------------|------------------|
|                  | Percent identity |                  |                  |
|                  | FucP-1           | FucP-2           | FucP-3           | FucP-4           |
| Unit             |                  |                  |                  |                  |
| XylE-1           | 19.2a            | 3.1              | 11.5             | 16.7b            |
| XylE-2           | 5.1              | 21.1a            | 10.4b            | 9.8              |
| XylE-3           | 3.9              | 5.0b             | 9.0a             | 25.3             |
| XylE-4           | 8.0b             | 5.5              | 9.8              | 3.5a             |

a Null model.  
b ‘Mix-and-match’ model proposed by Madej et al. [2013].

deeemed less likely than the simple retention of charges as dictated by the null hypothesis.

(7) The Δdistance map (fig. 5) indicated that the N-terminal bundle of 6 TMS in XylE is internally immobile during the conformational transition from the outward occluded state to the inward occluded state, and also from the inward occluded state to the inward open state. In the file 4GBY-4JA3.xlsx, available on request, a spreadsheet is provided for numerical exploration. This suggests that the halves, rather than the 3-TMS units, may be the basic functional units. Since this suggestion of necessity is based on studies with just one protein, XylE, and since the 3-D structure of XylE is not available in the outward open configuration, it is premature to conclude that this characteristic is generally valid for all conformational states and applies to MFS porters in general. Further studies will be required to establish the applicability of this interesting observation to the mechanism(s) of transport by the many members of the MFS [Chang et al., 2004]
Methods

Δdistance Mapping

Each Δdistance map was determined for 2 different conformations of XylE by subtracting the 2 contact maps from each other and displaying the Δdistance values using conditional formatting in Excel. The file 4GBY was used to design the software. First, only lines starting with ‘^ATOM’ were selected, leaving 3,587 lines. Second, selecting the α-carbons left 475 lines, from Tyr5 to Glu479. Third, a Perl script was created that could read the file, line by line. The program parses out the name of the residue and its x, y- and z-coordinates. It then loops over the other residues, calculating the 3-D distance, building up a 475 × 475 matrix of pairwise distances. This is visualized using conditional formatting in Excel.

The same was repeated for 4JA3_A (Ser8 to Glu465) and 4JA4_A (Tyr9 to Glu479). The data for all files were restricted to Tyr9-Glu465 to get comparable matrices. 4GBY contains all α-carbons between Tyr9 and Glu465. However, the 4JA* files are approximately 30 residues shorter. In 4JA3, residues 265–275 (inclusive), 11 residues are missing; in 304–309 (inclusive), 6 residues are missing; in 398–406 (inclusive), 9 residues are missing; and in 434–439 (inclusive), 6 residues are missing. In total, 32 residues are missing in 4JA3. The total length is 425 residues in 4JA3, and 428 residues in 4JA4. In 4JA4, residues 229–232 (4 residues), 264–275 (8 residues), 304–314 (11 residues), and 435–440 (6 residues) are missing. Thus, 29 residues are missing in total. The corresponding missing regions were removed from both files in a given comparison.

Maximum Bipartite Matching

We represented FucP and LacY by multiple sequence alignments of high-scoring homologues found by searching the National Center for Biotechnology Information nonredundant database with PHMMER with the sequences from PDB models 3O7Q and 2V8N, respectively. A minimal E value of 1 × 10−5 was used as the threshold for inclusion of hits.

These alignments were then used to train profile hidden Markov models with HH-Suite’s ‘hhmake’. In order to better represent the specific proteins in question, maximal allowed sequence identity was changed from hhmake’s default to 99%, and sequences used in training were required to have a minimum of 70% coverage with the query protein. These HMMs were then compared via HMM:HMM pairwise alignment using ‘halign’ [Söding, 2005].

For each of the four 3-TMS units in each protein, the appropriate columns of the previously produced alignment were copied to create alignments for each subunit in isolation. To achieve this, we followed the numbering of Madej et al. [2013]. Thus, for FucP, we used: FucP-1, residues 22–115; FucP-2, residues 116–229; FucP-3, residues 258–345; and FucP-4, residues 347–431. For LacY, we used: LacY-1, residues 7–102; LacY-2, residues 104–187; LacY-3, residues 220–309; and LacY-4, residues 311–400.

For each helix 3-TMS unit, a separate profile HMM was trained on only that subregion. Again, HMM:HMM alignment was performed using ‘halign’ for every pair of 3-TMS units in FucP and LacY. Because halign’s scores are not symmetric, [score (A,B) is not necessarily equal to score (B,A), though they are usually close], this was done twice for each pair in both orders, and the average of the 2 scores was used. Maximum accuracy realignment was used for this process with a setting of ‘–mact 0.05’ to encourage global alignments of subunits, an assumption in both the null and Madej hypotheses.

A ‘bipartite matching’, from graph theory, is a grouping into pairs of objects of one category with objects of another category (the two ‘parts’ of bipartite), such that each pair satisfies a definition of compatibility (there is an ‘edge’ between them). Also, no two elements of a single category have edges between them [Galil, 1986].

Motif Elicitation

Using Jalview [Waterhouse et al., 2009], we studied conserved positions containing lysines and arginines in the cytoplasmic loops of LacY family members (OHS; TC No. 2.A.1.5) and FucP family members (FHS; TC No. 2.A.1.7). MEME (version 4.9.1) employs a statistical modeling technique to create position-dependent letter-probability matrices [Bailey and Elkan, 1994]. The procedures followed were conducted as dictated in the program manuals.

Average Properties

To calculate average charges of residues, the distribution of residues in each column of the full-length alignments was tabulated, discounting gaps. Residues R, H and K were given a charge of ‘+1’, and D and E were given a charge of ‘−1’; all others were ‘0’. Then the average charge in that column was calculated. This information is visualized on the respective structures by coloring the appropriate residue red (positive) or blue (negative) for average charge values greater than 0.2.

Acknowledgements

The work reported was supported by NIH grants GM077402-05A1 and GM094610.


