# TMB Finding Pipeline: Novel Approach for Detecting β-Barrel Membrane Proteins in Genomic Sequences

M. Michael Gromiha,* Yukimitsu Yabuki, and Makiko Suwa

Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), AIST Tokyo Waterfront Bio-IT Research Building, 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan

Received June 21, 2007

We have developed a novel approach for dissecting transmembrane β-barrel proteins (TMBs) in genomic sequences. The features include (i) the identification of TMBs using the preference of residue pairs in globular, transmembrane helical (TMH) and TMBs, (ii) elimination of globular/TMH proteins that show sequence identity of more than 70% for the coverage of 80% residues with known structures, (iii) elimination of globular/TMH proteins that have sequence identity of more than 60% with known sequences in SWISS-PROT, and (iv) exclusion of TMH proteins using SOSUI, a prediction system for TMH proteins. Our approach picked up 7% TMBs in all the considered genomes. The comparison between the identified TMBs in *E. coli* genome and available experimental data demonstrated that the new approach could correctly identify all the 11 known TMBs, whose crystal structures are available. Further, it revealed the presence of 19 TMBs, homology with known structures, 60 TMBs similar to well annotated sequences, and 54 TMBs that have high sequence similarity with *Escherichia coli* β-barrel proteins deposited in Transport Classification Database (TCDB). Interestingly, the present approach identified TMBs from all 15 families in TCDB. In *human* genome, the occurrence of TMBs varies from 0 to 3% in different chromosomes. We suggest that our approach could lead to a step forward in the advancement of structural and functional genomics.

## INTRODUCTION

The β-barrel membrane proteins perform a variety of functions, such as pore formation, membrane anchoring, enzyme activity, bacterial virulence, mediating nonspecific, passive transport of ions and small molecules, selectively passing the molecules like maltose and sucrose, and are involved in voltage dependent anion channels.[1] In structural genomics, the identification of β-barrel membrane proteins is very important and necessary to understand their functions. The existence of several stretches of hydrophobic residues in TMH proteins paved the way to discriminate such a class of proteins with high accuracy of more than 90%.[2] Consequently, this type of proteins can be detected in genomic sequences with reasonable accuracy.[3] Although the transmembrane helical proteins in genomes are well characterized to be 20−30%, the occurrence of β-barrel membrane proteins is still not completely explored yet. Hence, developing new methods, which could identify the β-barrel membrane proteins, is a demanding task in structural and functional genomics.

Several methods have been proposed for discriminating transmembrane β-barrel proteins (TMBs) from globular and TMH proteins based on statistical analysis and machine learning algorithms.[4−15] We observed that the statistical methods have high accuracy in correctly identifying the TMBs, whereas the machine learning techniques excluded the globular and TMH proteins at high accuracy. On the other hand, few methods have been suggested to screen TMBs from genomic sequences.[16−18] Zhai and Saier[16] developed a β-barrel finder program based on secondary structure, hydropathy, and amphipathicity parameters and used it for identifying TMBs in *E. coli* genome. Berven et al.[17] proposed a program for identifying TMBs using two factors: (i) C-terminal pattern typical of many integral β-barrel proteins and (ii) integral β-barrel score based on the extent to which the sequence contains stretches of amino acids typical of transmembrane β-strands. Bigelow et al.[18] introduced a profile-based HMM for discriminating TMBs and suggested the probable TMBs in genomic sequences of 72 Gram-negative bacteria. These methods used a limited number of genomes, and/or they missed several known TMBs as the average occurrence of TMBs in genomes is predicted to be about 3%.

In this work, we have developed a new pipeline for detecting TMBs in genomes using *E. coli* and *human* genomes as model systems by combining statistical methods and elimination procedures. This approach was applied to 275 genomes, and we picked up about 7% of the sequences as TMBs. The identified TMBs in *E. coli* have been tested with known experimental data, such as three-dimensional structures of TMBs solved at high resolution, well annotated TMB sequences, and the β-barrel membrane proteins deposited in TCDB, Transport Classification Database.[19] Our approach correctly identified the TMBs similar to all known structures and representative examples from all 15 families in TCDB. Further, the merits and limitations of the method will be discussed.

* Corresponding author phone: +81-3-3599-8046; fax: +81-3-3599-8081; e-mail: michael-gromiha@aist.go.jp.

10.1021/ci700222s CCC: $37.00 
Published on Web 10/24/2007

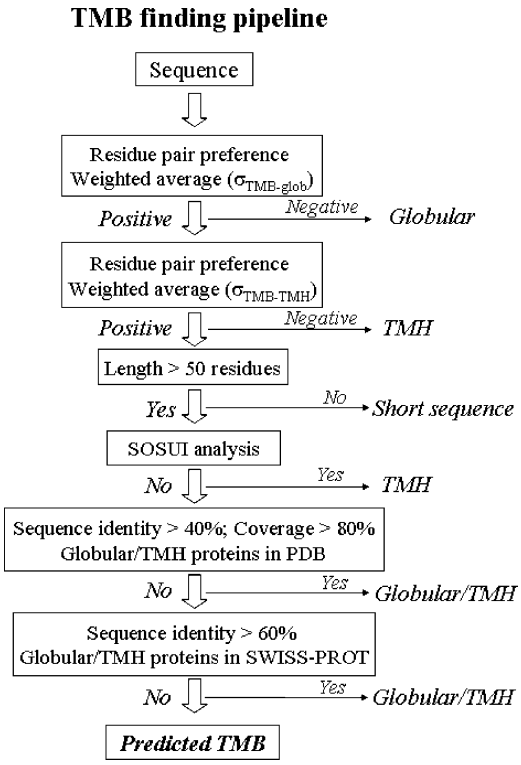


**Figure 1.** Pipeline for detecting TMB proteins in genomic sequences.

## MATERIALS AND METHODS

**Data Sets.** We have used several sets of data for the present work. They include the amino acid sequences of well annotated 377 TMBs, 20 TMBs of known experimental structure, 80 285 globular/TMH structurally known protein chains, 212 762 globular/TMH protein sequences, and the complete protein sequences of 275 genomes as explained below. The TMB structures have been selected from the Protein Data Bank[20] (PDB) with the criteria that they are solved at high resolution (<3 Å) and the sequence identity is less than 70% with the minimum length of 80 residues.[21] The amino acid sequences of 377 TMBs have been obtained from the PSORT database.[22] The sequences for globular/TMH proteins have been taken from PDB and SWISS-PROT. The amino acid sequences for all 275 completed genomes have been taken from the November 2005 release of the NCBI database (http://www.ncbi.nih.gov/). They include 23 genomes from archaea, 237 from bacteria, and 15 from eukaryote. The total number of proteins in these three kingdoms of life is 52 241, 686 562, and 165 186, respectively, with a total of 903 989 sequences.

**Detecting TMBs in Genomic Sequences.** We have developed a novel method, "TMB finding pipeline", for detecting TMBs in genomic sequences. The analysis of different statistical and machine learning methods for identifying TMBs showed that most of them missed several known TMBs, and the statistical method based on residue pair preference has successfully identified all the TMBs of known structure. In this method, for a new protein X, we have computed the compositional difference between globular and TMBs ($\sigma_{TMB-glob}$). The weighted average of $\sigma_{TMB-glob}$ with the dipeptide composition of protein X discriminates the TMB and globular protein.[12] We have followed the below-mentioned steps to identify the TMBs in genomic sequences as depicted in Figure 1: (i) identify the TMBs using the residue pair preferences between TMBs and globular proteins as explained in our earlier article,[12] (ii) repeat the same calculation with the residue pair preferences of TMBs and TMH proteins, (iii) eliminate the shorter sequences with less than 50 amino acid residues, (iv) exclude the TMH proteins using the program SOSUI, (v) eliminate the TMH and globular proteins that have more than 70% sequence identity and 80% coverage with known structures using the program, BLAST,[23] and (vi) eliminate the TMH and globular proteins that show the sequence identity of more than 60% with known sequences deposited in SWISS-PROT. We have used the concept of normalized score, which was found to be successful in detecting highly homologous sequences.[24]

**Table 1.** List of TMBs from *E. coli*, Whose Structures Are Solved at High Resolution and the TMBs Identified by the Present Approach

| protein | PDB code | identified TMBs (NCBI gi nos.) | sequence identity with known TMB (%) | ref |
|---|---|---|---|---|
| OmpX | 1QJ8 | gi\|16128782 | 99.3 | 28 |
| Pagp | 1MM4 | gi\|16128605 | 100.0 | 29 |
| OmpA | 1QJP | gi\|16128924 | 98.3 | 30 |
| OmpT | 1I78 | gi\|16128548 | 99.0 | 31 |
| OmpLA | 1QD5 | gi\|16131671 | 100.0 | 32 |
| OmpF | 2OMF | gi\|16128896 | 100.0 | 33 |
| | | gi\|16128227 | 62.1 | |
| | | gi\|16130152 | 58.5 | |
| | | gi\|16128536 | 58.2 | |
| | | gi\|16129338 | 56.6 | |
| Phosphoporin | 1PHO | gi\|16128227 | 100.0 | 34 |
| | | gi\|16128536 | 64.4 | |
| | | gi\|16128896 | 62.1 | |
| | | gi\|16129338 | 61.1 | |
| | | gi\|16130152 | 60.1 | |
| FhuA | 2FCP | gi\|16128143 | 100.0 | 35 |
| | | gi\|16128773 | 22.9 | |
| FepA | 1FEP | gi\|16128567 | 98.5 | 36 |
| | | gi\|16130093 | 32.0 | |
| | | gi\|16131804 | 23.5 | |
| FecA | 1KMO | gi\|16132112 | 100.0 | 37 |
| | | gi\|16129410 | 24.2 | |
| | | gi\|16128143 | 23.7 | |
| | | gi\|16131804 | 23.1 | |
| | | gi\|16128773 | 21.5 | |
| | | gi\|16130093 | 20.6 | |
| BtuB | 1NQE | gi\|16131804 | 100.0 | 38 |
| | | gi\|16130093 | 26.0 | |
| | | gi\|16128567 | 23.3 | |
| | | gi\|16132112 | 23.1 | |

**Table 2.** Detection of TMBs in *E. coli* Genome

| procedure | eliminated | detected |
|---|---|---|
| list of complete sequences | 0 | 4237 |
| residue pair preference between globular and TMBs | 3201 | 1036 |
| residue pair preference between TMH and TMBs | 207 | 829 |
| excluding short peptides | 22 | 807 |
| excluding TMH proteins using SOSUI | 51 | 756 |
| excluding TMH and globular proteins using 70% sequence identity with known structures | 251 | 505 |
| excluding TMH and globular proteins using 60% sequence identity with known sequences | 418 | 87 |

**Table 3.** Comparison of Detected TMBs with the Present Approach and $\beta$-Barrel Membrane Proteins Obtained from *E. coli* in TCDB[a]

| TCDB code | family | N | sequence identity (%) | BBF |
|---|---|---|---|---|
| 1.B.1 | General Bacterial Porin (GBP) | 5 | 56−100 | ○ |
| 1.B.3 | Sugar Porin (SP) | 2 | 25−100 | ○ |
| 1.B.6 | OmpA-OmpF Porin | 2 | 37−100 | ○ |
| 1.B.9 | The FadL Outer Membrane Protein (FadL) | 1 | 99.8 | ○ |
| 1.B.10 | Nucleoside-specific Channel-forming Outer Membrane Porin (Tsx) | 1 | 100.0 | ○ |
| 1.B.11 | Outer Membrane Fimbrial Usher Porin (FUP) | 12 | 23−46 | ○ |
| 1.B.12 | Autotransporter (AT) | 7 | 21−99.9 | ○ |
| 1.B.14 | Outer Membrane Receptor (OMR) | 8 | 20−100 | ○ |
| 1.B.17 | Outer Membrane Factor (OMF) | 3 | 24−100 | x |
| 1.B.18 | Outer Membrane Auxiliary (OMA) | 2 | 64−100 | x |
| 1.B.21 | OmpG Porin (OmpG) | 3 | 24−100 | x |
| 1.B.22 | Outer Bacterial Membrane Secretin (Secretin) | 1 | 24.3 | x |
| 1.B.25 | Outer Membrane Porin (OPr) | 1 | 100.0 | ○ |
| 1.B.33 | Outer Membrane Protein Insertion Porin (OmpIP) | 4 | 34−100 | x |
| 1.B.35 | Oligogalacturonate-specific Porin (KdgM) | 2 | 24−100 | x |

[a] N: number of TMBs identified in each family; BBF: identification of TMBs using $\beta$-barrel finder program;[16] ○: identified representative TMBs; x: failed to identify the TMBs.

**Comparison with Experiments.** Recent experimental studies showed that the structures of 11 TMB proteins from *E. coli* have been solved at high resolution, and the details are presented in Table 1. First, we have examined the predictive ability of our approach for identifying all these TMBs in *E. coli* genome. Second, we have analyzed all/identified proteins of *E. coli* that have sequence identity (varies from 20 to 100%) with 377 well annotated sequences in PSORT database. Third, we have compared the detected TMBs with the experimentally known $\beta$-barrel membrane proteins from *E. coli* deposited in TCDB.

## RESULTS AND DISCUSSION

**Detecting TMBs in *E. coli* Using the Present Approach.** We have identified the TMBs in genomic sequences using the new approach as described in the Methods section. An example is illustrated with *E. coli* genome as a model system in Table 2. The complete genome of *E. coli* contains 4237 proteins, and 1036 proteins are detected as TMBs using the residue pair preference between globular and TMBs. We have excluded TMH and globular proteins using other elimination procedures, and finally we obtained 87 proteins as TMBs. This constitutes 2.05% of the total sequences, while the statistical methods picked up 18−25% of the proteins as TMBs. Our method shows the remarkable achievement in genomic research that it correctly eliminated most of the false positives, which is the major problem with many statistical methods, and identified almost all the real TMBs, which is the main shortcoming of machine learning techniques.

**Comparison of Detected TMBs with Known Experimental Structures.** Table 1 shows the details of the TMBs along with their Protein Data Bank codes. The present approach identified 19 TMBs, which have the sequence identity in the range of 20−100% with known structures of TMBs. Interestingly, similar sequences of all the experimentally known structures of TMBs have been identified by the present method with the sensitivity of 100%. On the other hand, the method based on the Hidden Markov Model failed to identify the protein Pagp from *E. coli*.[7]

The results presented in Table 1 indicates that the proteins OmpF, phosphoporin, FhuA, FepA, FecA, and BtuB have more than one representative sequence in *E. coli* genome. It is noteworthy that all the proteins in the data set with 22 transmembrane $\beta$-strand segments (FhuA, FepA, FecA, and BtuB) have at least two similar sequences in *E. coli* genome. On the other hand, few identified TMBs have significant sequence similarity with $\beta$-barrel proteins from other organisms, such as sucrose porin and maltoporin from *Salmonella typhimurium* (1A0S and 2MPR) and osmoporin from *Klebsiella pneumoniae* (1OSM).

**Comparison of Detected TMBs with Well Annotated Sequences.** We have examined the TMBs identified in *E. coli* using our method with the sequences of 377 experimentally known well annotated TMBs from PSORT database.[22] The present approach identified 60 protein sequences in *E. coli* genome that have the sequence identity in the range of 20−100% with well annotated 377 TMBs. On the other hand, among the 4237 sequences in the whole genome of *E. coli* only 63 have the sequence identity in the range of 20−100%. This analysis showed that our method correctly picked up 95.2% of the known TMBs even when the homology is remote (20% sequence identity). Interestingly, all the proteins in *E. coli* that have the significantly high sequence identity (>70%) are correctly picked up by our approach with the accuracy of 100%.

**Evaluation of TMBs Identified by the Present Approach Using $\beta$-Barrel Membrane Proteins in TCDB.** We have evaluated the TMBs identified by the present "TMB finding pipeline" approach with the data available for $\beta$-barrel membrane proteins in TCDB. Among the 45 different classes of $\beta$-barrel membrane proteins 15 families have the sequences obtained from *E. coli*. We have verified our results with representative sequences in these 15 families, and the results are shown in Table 3. We have successfully identified the representative TMBs in all these families. Our results showed the presence of 54 TMBs that have sequence identity with the TMBs deposited in TCDB. Interestingly, the present approach could identify the sequences that have very less sequence identity (<40%). The detection of TMBs belonging to all families of TCDB indicates the excellent performance of the present approach.

Zhai and Saier[16] developed a program, $\beta$-barrel finder, for identifying $\beta$-barrel membrane proteins in prokaryotic ge-

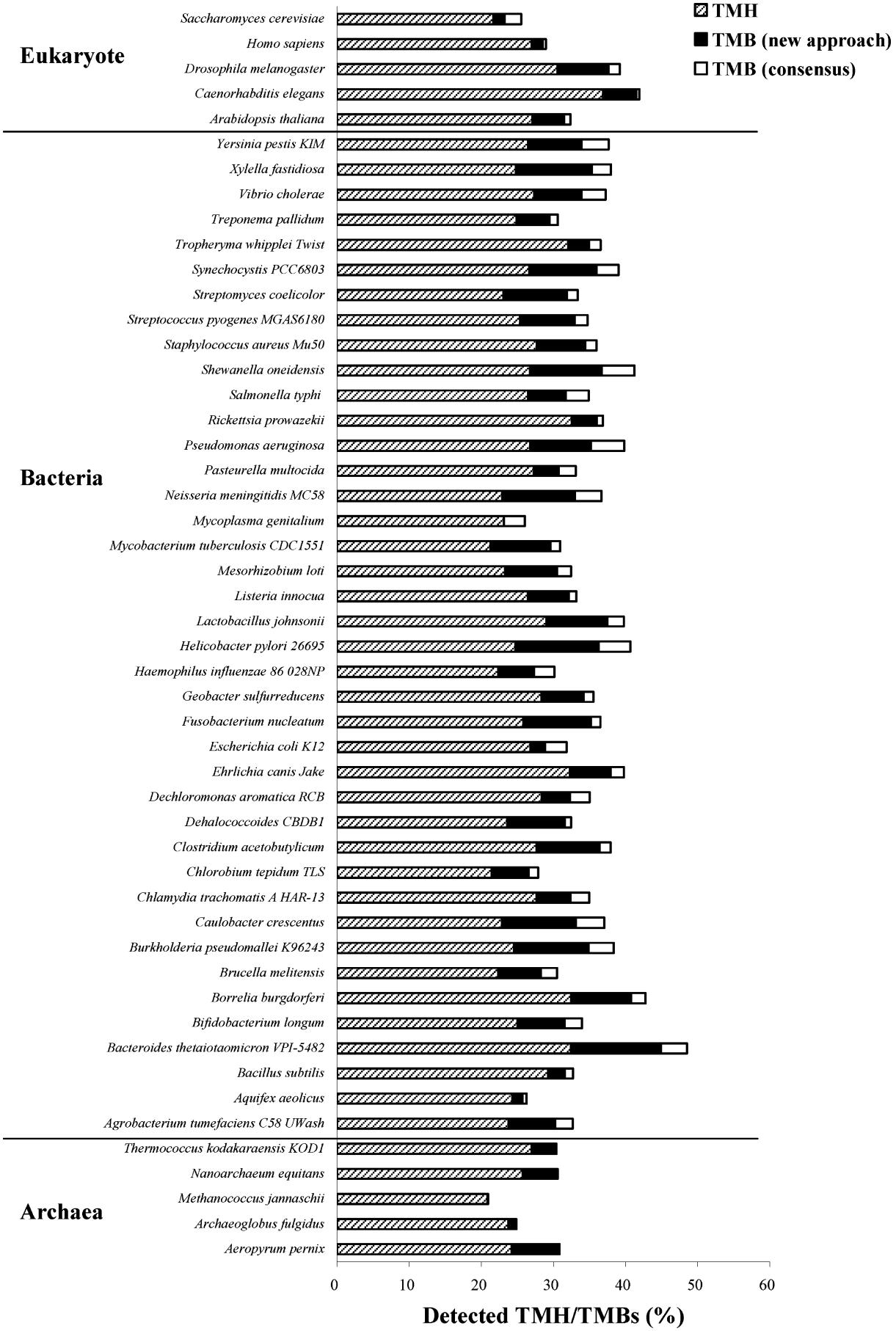


**Figure 2.** Percentage of TMH and TMBs in selected genomes. Shaded bars: TMH proteins obtained with the SOSUI program. Filled bars: TMBs obtained with the new approach (TMB finding pipeline). Open bars: TMBs obtained using the consensus of statistical methods and support vector machines.

nomes. The characteristic features used in the method are similar to that used in Gromiha et al.[25] for predicting the membrane spanning $\beta$-strand segments in bacterial porins. They have detected 118 TMBs in *E. coli* genome, and this method failed to detect the TMBs in several families, such as Outer Membrane Factor (OMF), outer Membrane Auxiliary (OMA), OmpG Porin (OmpG), Outer Bacterial Membrane Secretin (Secretin), etc. Only representatives from nine families could be identified among 15 families available in TCDB. The method based on k-nearest neighbor missed few TMBs with the e-value > 3 although these proteins have been used in the training set to develop the method (10). The BOMB server failed to identify eight proteins.[17] The profile based HMM missed six TMBs, and no representative protein was detected in the families, outer Membrane Auxiliary (OMA), OmpG Porin (OmpG), and Outer Bacterial Membrane Secretin (Secretin). This analysis demonstrates that the present approach performs extremely well compared with other methods.

**Performance of the Present Approach in Recently Identified Proteins.** Recently, Marani et al.[26] reported eight new outer membrane proteins in *E. coli*, and their UniProt codes are as follows: CSGF_ECOLI, YHJY_ECOLI, YTFM_ECOLI, YFAL_ECOLI, YAIO_ECOLI, YLII_ECOLI, YFAZ_ECOLI, and YAGZ_ECOLI. We have tested our approach and found that it could identify all these proteins as TMBs.

**Identification of TMBs in Completed Genomes.** We have applied our approach for identifying TMBs in all completed genomes, and the results for selected genomes are presented in Figure 2. We observed that the present method identified 5.1%, 7.5%, and 6.5% of the proteins as TMBs in eukaryote, bacteria, and archaea, respectively. The respective TMH proteins in these kingdoms of life are 27.9%, 26.1%, and 24.6%. Interestingly, our method did not detect any TMB in *Mycoplasma genitalium*, which agreed well with experiments (*M. genitalium* has no outer membrane and $\beta$-barrel membrane protein).

**Detection of TMBs in *Human* Genome.** We have identified the TMBs in all the 24 chromosomes of *human* genome using the new approach. We observed that 0−3% of the proteins are identified as TMBs, while SOSUI predicted 15−39% of the proteins as TMH type. The chromosome 13 has the highest occurrence of 3% TMBs followed by the chromosomes 21 (2.9%) and 4 and 8 (2.2% each). The chromosome Y has no occurrence of TMBs. Interestingly, the lowest occurrence of TMH proteins is identified in chromosome Y as in the case for TMBs.[27]

**Possible Improvements.** In this work, we have proposed a new approach for detecting TMBs in genomic sequences. This method could identify most of the experimentally known TMBs, as illustrated above using *E. coli* genome as a model system. The identified TMBs agreed very well with known structures of TMBs, well annotated sequences, and the $\beta$-barrel membrane proteins deposited in TCDB.[19] However, there may have been few false positives in the identified TMBs. These can be eliminated by developing methods to exclude other types of proteins, such as DNA and RNA binding proteins, etc. Further, increasing the number of TMB sequences when available, incorporating alignment profiles, etc. may improve the accuracy of identifying TMBs in genomic sequences.

## CONCLUSIONS

We have developed a novel approach for detecting TMBs in genomic sequences based on reside pair preference in globular, TMH and TMBs. The method has been refined by excluding globular proteins using the BLAST sequence alignment and TMH proteins with at least two membrane spanning segments in SOSUI. Our method identified 2% of the proteins in *E. coli* as TMBs, and it picked up all the known TMBs with 100% accuracy. The result has been tested with well annotated TMBs and $\beta$-barrel membrane proteins in TCDB, and we observed an excellent agreement with experimental data. Further, it detected 0−3% of the proteins in *human* chromosomes as TMBs. Considering the whole data set, the number of TMBs is approximately one-fourth of that of TMH proteins. We suggest that this approach could be a powerful tool to detect TMBs.

**Abbreviations.** TMB, transmembrane $\beta$-barrel protein; TMH, transmembrane helical protein

## ACKNOWLEDGMENT

The authors wish to thank Mr. Sivasundaram Suharnan and Mr. Srinesh Kundu for their help and Dr. Yutaka Akiyama for encouragement.

## REFERENCES AND NOTES

(1) Koebnik, R.; Locher, K. P.; Van Gelder, P. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol.* **2000**, *37*, 239−53.
(2) Hirokawa, T.; Boon-Chieng, S.; Mitaku, S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **1998**, *14*, 378−379.
(3) Granseth, E.; Daley, D. O.; Rapp, M.; Melen, K.; von Heijne, G. Experimentally constrained topology models for 51,208 bacterial inner membrane proteins. *J. Mol. Biol.* **2005**, *352*, 489−94.
(4) Gnanasekaran, T. V.; Peri, S.; Arockiasamy, A.; Krishnaswamy, S. Profiles from structure based sequence alignment of porins can identify beta stranded integral membrane proteins. *Bioinformatics* **2000**, *16*, 839−842.
(5) Wimley, W. C. Toward genomic identification of beta-barrel membrane proteins: composition and architecture of known structures. *Protein Sci.* **2002**, *11*, 301−312.
(6) Martelli, P. L.; Fariselli, P.; Krogh, A.; Casadio, R. A sequence-profile-based HMM for predicting and discriminating beta barrel membrane proteins. *Bioinformatics* **2002**, *18*, S46−S53.
(7) Bagos, P. G.; Liakopoulos, T. D.; Spyropoulos, I. C.; Hamodrakas, S. J. A Hidden Markov Model method, capable of predicting and discriminating beta-barrel outer membrane proteins. *BMC Bioinformatics* **2004**, *5*, 29.
(8) Liu, Q.; Zhu, Y.; Wang, B.; Li, Y. Identification of beta-barrel membrane proteins based on amino acid composition properties and predicted secondary structure. *Comput. Biol. Chem.* **2003**, *27*, 355−361.
(9) Natt, N. K.; Kaur, H.; Raghava, G. P. Prediction of transmembrane regions of beta-barrel proteins using ANN- and SVM-based methods. *Proteins* **2004**, *56*, 11−18.
(10) Garrow, A. G.; Agnew, A.; Westhead, D. R. TMB-Hunt: a web server to screen sequence sets for transmembrane beta-barrel proteins. *Nucleic Acids Res.* **2005**, *33*, W188−192.
(11) Gromiha, M. M.; Suwa, M. A simple statistical method for discriminating outer membrane proteins with better accuracy. *Bioinformatics* **2005**, *21*, 961−968.
(12) Gromiha, M. M.; Ahmad, S.; Suwa, M. Application of residue distribution along the sequence for discriminating outer membrane proteins. *Comput. Biol. Chem.* **2005**, *29*, 135−42.
(13) Gromiha, M. M. Motifs in outer membrane protein sequences: Applications for discrimination. *Biophys. Chem.* **2005**, *117*, 65−71.
(14) Park, K. J.; Gromiha, M. M.; Horton, P.; Suwa, M. Discrimination of outer membrane proteins using support vector machines. *Bioinformatics* **2005**, *21*, 4223−9.
(15) Gromiha, M. M.; Suwa, M. Discrimination of outer membrane proteins using machine learning algorithms. *Proteins* **2006**, *63*, 1031−1037.

(16) Zhai, Y.; Saier, M. H., Jr. The β-barrel finder (BBF) program, allowing identification of outer membrane β-barrel proteins encoded within prokaryotic genomes. *Protein Sci.* **2002**, *11*, 2196−2207.
(17) Berven, F. S.; Flikka, K.; Jensen, H. B.; Eidhammer, I. BOMP: a program to predict integral β-barrel outer membrane proteins encoded within genomes of Gram-negative bacteria. *Nucleic Acids Res.* **2004**, *32*, W394−W399.
(18) Bigelow, H. R.; Petrey, D. S.; Liu, J.; Przybylski, D.; Rost, B. Predicting transmembrane beta-barrels in proteomes. *Nucleic Acids Res.* **2004**, *32*, 2566−2577.
(19) Busch, W.; Saier, M. H., Jr. The transporter classification (TC) system. *Crit. Rev. Biochem. Mol. Biol.* **2002**, *37*, 287−337.
(20) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(21) Bagos, P. G.; Liakopoulos, T. D.; Hamodrakas, S. J. Evaluation of methods for predicting the topology of beta-barrel outer membrane proteins and a consensus prediction method. *BMC Bioinformatics* **2005**, *6*, 7.
(22) Gardy, J. L.; Spencer, C.; Wang, K.; Ester, M.; Tusnady, G. E.; Simon, I.; Hua, S.; deFays, K.; Lambert, C.; Nakai, K.; Brinkman, F. S. PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria. *Nucleic Acids Res.* **2003**, *31*, 3613−3617.
(23) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.
(24) Yabuki, Y.; Mukai, Y.; Swindells, M. B.; Suwa, M. GENIUS II: a high-throughput database system for linking ORFs in complete genomes to known protein three-dimensional structures. *Bioinformatics* **2004**, *20*, 596−58.
(25) Gromiha, M. M.; Majumdar, R.; Ponnuswamy, P. K. Identification of membrane spanning beta strands in bacterial porins. *Protein Eng.* **1997**, *10*, 497−500.
(26) Marani, P.; Wagner, S.; Baars, L.; Genevaux, P.; de Gier, J. W.; Nilsson, I.; Casadio, R.; von Heijne, G. New Escherichia coli outer membrane proteins identified through prediction and experimental verification. *Protein Sci.* **2006**, *15*, 884−889.
(27) Ono, Y.; Fujibuchi, W.; Suwa, M. Automatic gene collection system for genome-scale overview of G-protein coupled receptors in eukaryotes. *Gene* **2005**, *364*, 63−73.
(28) Vogt, J.; Schulz, G. E. The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence. *Struct. Fold Des.* **1999**, *7*, 1301−1309.
(29) Hwang, P. M.; Choy, W. Y.; Lo, E. I.; Chen, L.; Forman-Kay, J. D.; Raetz, C. R.; Prive, G. G.; Bishop, R. E.; Kay, L. E. Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13560−13565.
(30) Pautsch, A.; Schulz, G. E. High-resolution structure of the OmpA membrane domain. *J. Mol. Biol.* **2000**, *298*, 273−282.
(31) Vandeputte-Rutten, L.; Kramer, R. A.; Kroon, J.; Dekker, N.; Egmond, M. R.; Gros, P. Crystal structure of the outer membrane protease OmpT from Escherichia coli suggests a novel catalytic site. *EMBO J.* **2001**, *20*, 5033−5039.
(32) Snijder, H. J.; Ubarretxena-Belandia, I.; Blaauw, M.; Kalk, K. H.; Verheij, H. M.; Egmond, M. R.; Dekker, N.; Dijkstra, B. W. Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature* **1999**, *401*, 717−721.
(33) Cowan, S. W.; Garavito, R. M.; Jansonius, J. N.; Jenkins, J. A.; Karlsson, R.; Konig, N.; Pai, E. F.; Pauptit, R. A.; Rizkallah, P. J.; Rosenbusch, J. P, et al. The structure of OmpF porin in a tetragonal crystal form. *Structure* **1995**, *3*, 1041−1050.
(34) Cowan, S. W.; Schirmer, T.; Rummel, G.; Steiert, M.; Ghosh, R.; Pauptit, R. A.; Jansonius, J. N.; Rosenbusch, J. P. Crystal structures explain functional properties of two E. coli porins. *Nature* **1992**, *358*, 727−733.
(35) Ferguson, A. D.; Hofmann, E.; Coulton, J. W.; Diederichs, K.; Welte, W. Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. *Science* **1998**, *282*, 2215−2220.
(36) Buchanan, S. K.; Smith, B. S.; Venkatramani, L.; Xia, D.; Esser, L.; Palnitkar, M.; Chakraborty, R.; van der Helm, D.; Deisenhofer, J. Crystal structure of the outer membrane active transporter FepA from Escherichia coli. *Nat. Struct. Biol.* **1999**, *6*, 56−63.
(37) Ferguson, A. D.; Chakraborty, R.; Smith, B. S.; Esser, L.; van der Helm, D.; Deisenhofer, J. Structural basis of gating by the outer membrane transporter FecA. *Science* **2002**, *295*, 1715−1719.
(38) Chimento, D.P .; Mohanty, A. K.; Kadner, R. J.; Wiener, M. C. Substrate-induced transmembrane signaling in the cobalamin transporter BtuB. *Nat. Struct. Biol.* **2003**, *10*, 394−401.

CI700222S