Biophysics, Vol. 48, Suppl. 1, 2003, pp. S141–S145.
Original Russian Text Copyright © by Kalinina, Gelfand, Mironov, Rakhmaninova.

Amino Acid Residues Forming Specific Contacts between Subunits in Tetramers of the Membrane Channel GlpF

O. V. Kalinina[1], M. S. Gelfand[1,2,3], A. A. Mironov[1,2], and A. B. Rakhmaninova[1]

[1]*Department of Bioengineering and Bioinformatics, Moscow State University,*
1-73 Vorobievy Gory, Moscow, 119992 Russia; E-mail: abr@belozersky.msu.ru
[2]*State Scientific Center GosNIIGenetika, 1 1st Dorozhny proezd, Moscow, 117545 Russia*
[3]*Institute for Problems of Information Transmission, Russian Academy of Sciences,*
19 Bolshoi Karetny per., Moscow, 127994 Russia

Received November 11, 2003; in final form, January 24, 2004

Abstract—Bacterial proteins from the MIP family of membrane channels are divided into two subfamilies, glyceroaquaporins and aquaporins. The earlier developed method SDPred [1,2] was applied to predict amino acid residues responsible for specific functional features of these two subfamilies. Five of such residues mapped to the external side of the monomer. The analysis of the tetrameric glyceroaquaporin GlpF from *Escherichia coli* reported here demonstrates that these residues—20Leu, 24Phe, 43Glu, 108Tyr, and 193Ser—form tight intersubunit contacts. Moreover, these residues form contacts between each other in two spatial clusters. This allows us to suggest that 20Leu, 24Phe, 43Glu, 108Tyr and 193Ser form specific patterns on the external side of GlpF monomers that are responsible for correct recognition of monomers and preclude formation of chimeric oligomers with aquaporins from the other family.

Key words: protein design, lattice model, knot, folding, Go model, free energy landscape, mutation

INTRODUCTION

The MIP (Major Intrinsic Protein) family is a large group of selective membrane channels responsible for passive transport of water and some other small molecules such as glycerol, urea, ammonia, ÑÎ$_2$, etc. [3]. These proteins play a major role in osmotic regulation. They are found in diverse organisms from bacteria and archaea to yeasts, plants, and animals. Basing on sequence similarity, they were divided into six subfamilies [4], of which only two were found in bacteria: aquaporins (water channels) and glyceroaquaporins (channels conducting glycerol and water, the latter much less effectively than aquaporins [5]).

The structure of several proteins of the MIP family has been solved by X-ray analysis. These are glyceroaquaporin GlpF from *Escherichia coli* [6], aquaporins AQP1 from cattle [7] and human [8], and AqpZ from *E. coli* [5]. The 3D structure of monomers is a sheaf of six long transmembrane alpha-helices and two shorter alpha-helices reaching only half of the membrane. The conducting channel is in the middle of the sheaf (Fig. 1a). In the membrane these proteins usually form tetramers [5, 6], and it is discussed whether the channel formed by the monomers is an additional ion-conducting channel [3, 6].

Families of homologous proteins whose functions are generally similar but the substrate specificities different pose the problem of identifying the amino acid residues responsible for the differences in specificity. This problem is especially interesting when the 3D structure of the proteins is not known, and it has been addressed in a number of recent studies [9–11]. In particular, we developed an algorithm for identification of specificity-determining positions (SDPs) [1, 2]. It is based on statistical analysis of multiple protein alignment divided into groups of proteins with the same specificity, and does not require any additional structural data. On the other hand, such data are useful for testing the algorithm.

For the bacterial proteins from the MIP family, we identified 21 SDPs [1]. Mapping of these proteins

Table 1. Contacts of residues in A-type (a) and B-type (b) clusters with residues from other GlpF subunits (contacts between SDPs are in bold; contact with 194Met is in italics)

(a)

subunit I		subunit II		subunit IV		
residue	atom	residue	atom	residue	atom	distance (Å)
Glu43	OE1	Ser38	O			4.8
Glu43	**OE2**	**Glu43**	**OE2**			**4**.1
Glu43	CG			Trp42	CD1	3.7
Glu43	**OE2**			**Glu43**	**OE2**	**4.1**

(b)

subunit I		subunit II		
residue	atom	residue	atom	distance (Å)
Leu20	CD2	Ile158	CD1	4.3
Leu20	CD1	Leu162	CD2	45
Phe24	CZ	Ile158	CG2	3.9
Phe24	CZ	Leu186	CD1	3.9
Phe24	CE2	Val189	CG2	3.8
Phe24	CE2	le190	CG1	3.7
Phe24	**CA**	**Ser193**	**CB**	**3.9**
Phe24	**O**	**Ser193**	**OG**	**4.2**
Phe24	**O**	**Ser193**	**CB**	**3.3**
Gly27	O	Ser193	O	3.2
Cys28	CA	Ser193	CA	3.8
Tyr108	**OH**	**Ser193**	**O**	**2.6**
Tyr108	*CE1*	*Met194*	*CE*	*3.7*
Tyr108	CE1	Leu197	CD1	3.9

into the known GlpF structure [6] demonstrated that five SDPs—20Leu, 24Phe, 43Glu, 108Tyr, and 193Ser—lie on the outer side of the monomer and could participate in the tetramer formation. The aim of this study was to check this conjecture.

METHODS

The SDP prediction algorithm SDPred [1, 2] was used to determine positions responsible for specific ligand recognition. The algorithm is based on statistical analysis of multiple alignment, in which groups of proteins having the same specificity are defined. The training alignment was done by ClustalX and included 17 proteins: 10 glyceroaquaporins with average sequence identity 48%, and 7 aquaporins with average identity 59%. The average identity between the groups was 45%.

The set of SDPs was compared with the GlpF structure [6] from PDB, identifier 1fx8, using the visual representation and contact analysis by RasMol v. 2.7.2.1.

RESULTS

SDPred identified in the MIP alignment 21 SDPs corresponding to the following residues in GlpF from *E. coli*:: 20Leu, 22Ile, 24Phe, 43Glu, 48Trp, 108Tyr, 135Phe, 136Ser, 137Thr, 159Leu, 187Ile, 191Gly, 193Ser, 194Met, 195Gly, 199Gly, 200Phe, 201Ala, 207Asp, 211Lys, 236Pro (Fig. 1b). In [1] we demonstrated that of these, 16 residues either are in tight contact with glycerol, or belong to the channel-forming helices on the internal (channel) side (gray spheres in Fig. 1). Five residues, 20Leu, 24Ile, 43Glu, 108Tyr, 193Ser, lie on the external side of the monomer (white spheres on Fig. 1). One more

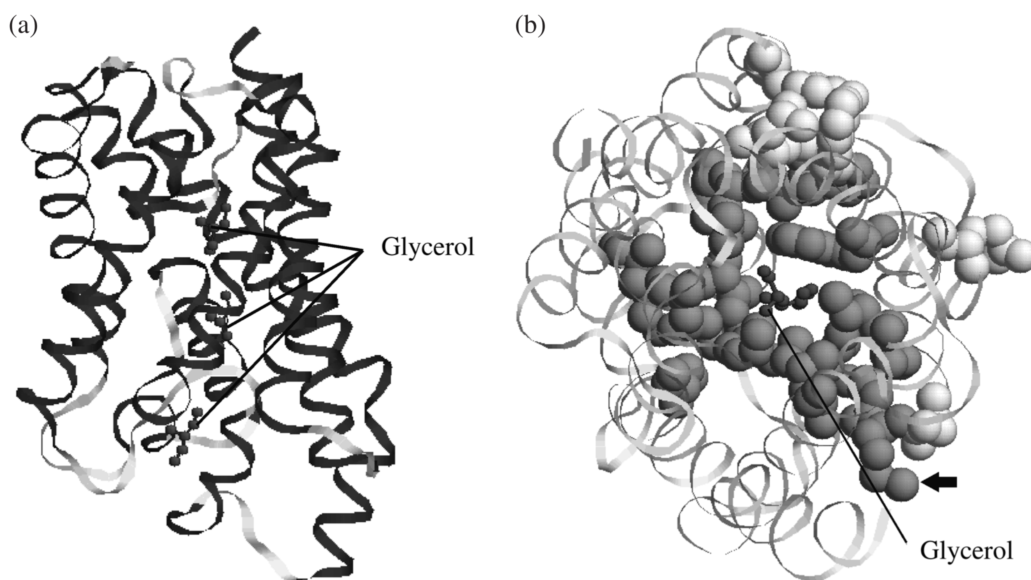

Fig. 1. (a) The structure of the GlpF monomer from *E. coli* (1fx8). (b) Predicted SDPs in the GlpF monomer from *E. coli* (1fx8). Atoms of SDPs contacting glycerol or situated on the cjannel side are shown as gray spheres. Atoms of SDPs lying on the external side are shown as white spheres. 194Met is marked by an arrow.

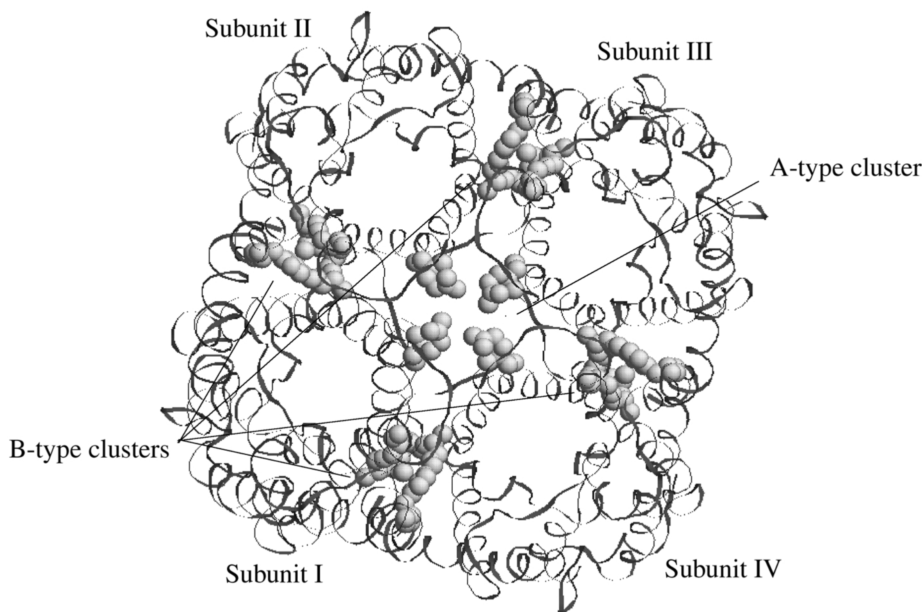

Fig. 2. Predicted SDPs in the GlpF tetramer from *E. coli* (1fx8, biological unit). Atoms of SDPs lying on the intersubunit contact surface are shown as white spheres.

residue, 194Met, has main chain atoms in the channel side of the helix, and side chain atoms on the external side (arrow on Fig. 1). Here we checked whether the external side residues 20Leu, 24Ile, 43Glu, 108Tyr, 193Ser participate in contacts responsible for formation of the GlpF tetramer.

Each GlpF subunit forms contacts (defined as having at least a pair of atoms at the distance not exceeding 4.5Å) with 69 residues from the other three subunits. All five external SDPs belong to this list. Moreover, of 6 tightest contacts (at least one interatomic distance less than <3.0Å), two are the predicted 108Tyr and 193Ser. It should be noted that the oligomerization of GlpF leads to a decrease in accessible surface, and this decrease is highest at 193Ser that is almost completely shielded from the solvent [12].

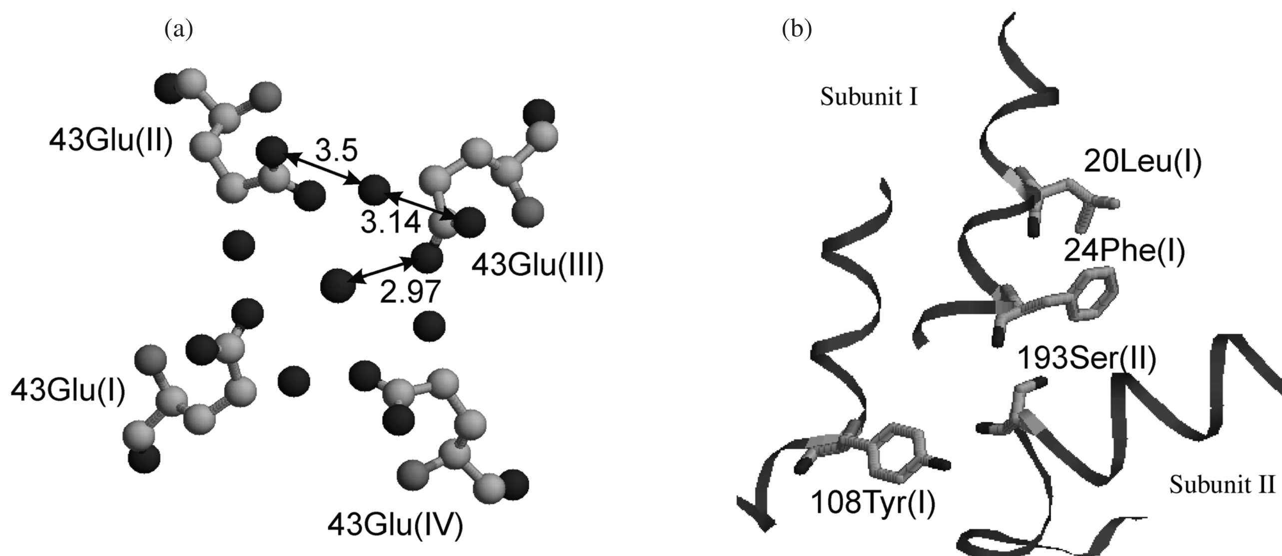

Fig. 3. (a) SDPs in the A-type cluster. Oxygen atoms are shown by gray spheres. The distances are given in Angstroms. (b) SDPs in the B-type cluster.

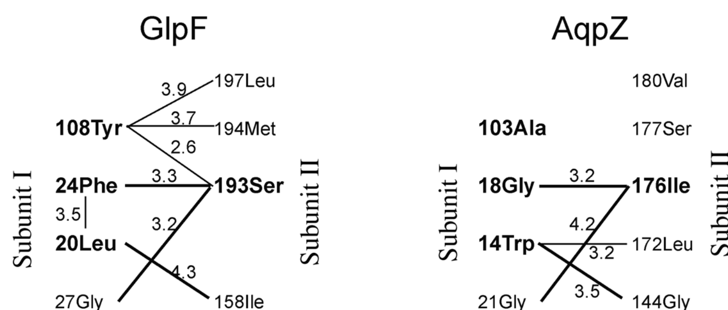

Fig. 4. Comparison of contacts in B-type clusters of the GlpF tetramers [6] and AqpZ tetramers [5] from *E. coli*. Aligned residues are shown in corresponding positions (e.g. 108Tyr in GlpF corresponds to 103Ala in AqpZ). Two residues are joined by a line if the distance between them allows for the contact. The distances are given in Angstroms. Hydrogen bonds supporting alpha-helices are not taken into account. Residues corresponding to predicted SDPs are set in bold.

Figure 2 shows the positions of 20Leu, 24Ile, 43Glu, 108Tyr, 193Ser in the tetramer structure. These residues form compact clusters of two types. One A-type cluster is formed by 43Glu from all four subunits (Fig. 3a), whereas each of four B-type clusters is formed by 20Leu, 24Ile, 108Tyr from one subnit and 193Ser from another subnit (Fig. 3b). The contacts of these residues with residues from other subunits are listed in Table 1a (A-type cluster) and 1b (B-type cluster).

The A-type cluster is formed by four glutamic residues contacting through carboxyl groups, which is unusual. One possible explanation could be the existence of magnesium ion coordinated by these residues and molecules of water [6]. Another possibility could be formation of a network of hydrogen bonds between the NH group of 40Gly, protonated forms of carboxyl groups of 43Glu, and water molecules clustered in this area (Fig. 3a). The carbon atoms of 43Glu(I) also form contacts with the carbon atoms of 42Trp(IV). Interestingly, in the glyceroaquaporin subfamily, Trp in position 42 is observed only in proteins with Glu in position 43; in other cases the combination 43V + 42I is observed.

The B-type clusters are formed by three residues, 20Leu(I), 24Phe(I), 108Tyr(I), of one subunit and one residue, 193Ser(II), of another subunit (Fig. 3b). One more close residue is 194Met(II) that forms a contact with 108Tyr(I). 20Leu(I) does not form contacts with SDP from the other subunit, but it forms contacts with other residues from the second subunit and with 24Phe(I) (not shown). The hydroxyl of 108Tyr(I) and the carbonyl of 193Ser(II) could form a hydrogen bond.

Thus five external SDPs participate in intersubunit contacts of the GlpF tetramer. Moreover, they participate in tight contacts and form contacts with each other. Since, by the basic assumption, SDPs are responsible for specific interactions, it is interesting to compare the above-described clusters of GlpF with the corresponding clusters in AqpZ. In the latter, the region corresponding to the A-type cluster is different. Residue 43Glu of GlpF corresponds to residue 38Gly of AqpZ that forms no contacts with either 38Gly of other subunits, or 37Ala (corresponding to 42Trp of GlpF), but can form an intersubunit contact with 175Leu (the distance between them is 3.7 Å). On the other hand, the structure of the region corresponding to the B-type cluster is more or less conserved (Fig. 4), although it is formed by different residues. It should be noted that the residues corresponding to two of four SDPs in this cluster, 24Phe and 193Ser in GlpF, and 18Gly and 176Ile in AqpZ form tight contacts in both tetramers (3.3 Å and 3.2 Å respectively).

DISCUSSION

Prediction of SDPs in proteins functioning as oligomers often identifies positions that in known structures are involved in intersubunit contacts. Thus, the set of SDPs identified in the bacterial transcription factor family LacI included not only residues contacting DNA and ligand, as expected, but also tight intersubunit contacts [1, 2]. In the case of the MIP family considered in detail here, five SDP residues participate in tight intersubunit contacts.

This prompts a conjecture that selection in positions responsible for oligomerization is correlated with selection in positions determining the specificity of substrate binding. In the MIP family we actually see tight spatial clusters of SDPs involved in intersubunit contacts that could provide specific recognition patterns precluding formation of chimeric aquaporin–glyceroaquaporin heterooligomers.

ACKNOWLEDGMENTS

We are grateful to L. Mirny, A. Finkelstein, and S. Sunyaev for useful discussions.

This study was partially supported by grants from HHMI (55000309) and LICR (CRDF RB0-1268).

REFERENCES

1. Kalinina, O.V., Mironov, A.A., Gelfand, M.S., and Rakhmaninova, A.B., *Protein Sci.*, 2004, vol. 13, pp. 443–456.
2. Kalinina, O.V., Proceedings of the International Moscow conference on computational molecular biology, 2003, pp. 94–96.
3. Transport Protein Database, TC-DB, http://tcdb.ucsd.edu/tcdb/tcfamilybrowse.php?tcname=1.A.8#description
4. Zardoya, R. and Villalba, S., *J. Mol. Evol.*, 2001, vol. 52, pp. 391–404.
5. Savage, D.F., Egea, P.F., Robles-Colmenares, Y., O'Connell, J.D., Stroud, R.M., *PLoS Biol.*, 2003, vol. 1, no. 3, pp. 334–340.
6. Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski, J., and Stroud, R.M., *Science*, 2000, vol. 290, pp. 481–486.
7. Sui, H., Han, B.G., Lee, J.K., Walian, P., Jap, B.K., *Nature*, 2001, vol. 414, pp. 872–878.
8. Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J.B., Engel, A., and Fujiyoshi, Y., *Nature*, 2000, vol. 407, pp. 599–605.
9. Hannenhalli, S.S. and Russell, R.B., *J. Mol. Biol.*, 2000, vol. 303, pp. 61–76.
10. Mirny, L.A. and Gelfand, M.S., *J. Mol. Biol.*, 2002, vol. 321, pp. 7–20.
11. Gaucher, E.A., Gu, X., Miyamoto, M.M., and Benner, S.A., 2002, *Trends Biochem. Sci.*, vol. 27, no. 6, pp. 315–321.
12. EBI MSD Macromolecule File Server http://pqs.ebi.ac.uk/.