Supporting Information

Coupled-mutations enabled glycerol transportation in an aquaporin Z mutant

Zhi Ping¹, Feng Zhou¹, Xin Lin¹, Haibin Su¹,²*

¹ Institute of Advanced Studies, Nanyang Technological University, 60 Nanyang View, Singapore 639673
² Department of Chemistry, The Hong Kong University of Science and Technology, Hong Kong, China
* Corresponding Author

E-mail: haibinsu@ust.hk
**Section 1. Cross reference table**

The aquaporin sequences for the SCA analysis were obtained using BLAST searches against a non-redundant protein database. 14 aquaporin sequences representing unique aquaporin types: AQP0 (NCBI protein ID: NP_036196), AQP1 (NCBI protein ID: NP_932766), AQP2 (NCBI protein ID: CAG46821), AQP3 (NCBI protein ID: CAG46822), AQP4 (NCBI protein ID: NP_001641), AQP5 (NCBI protein ID: CAG46819), AQP6 (NCBI protein ID: NP_001643), AQP7 (NCBI protein ID: NP_001161), AQP8 (NCBI protein ID: NP_001160), AQP9 (NCBI protein ID: CAG46824), AQP10 (NCBI protein ID: CAH70483), AQPM (NCBI protein ID: NP_275246), AQPZ (NCBI protein ID: NP_752939), GLPF (NCBI protein ID: NP_290556), were selected from the database. a PSI-BLAST\(^1\) (e < 0.001) was run for each of these 14 sequences to generate groups of more than 3000 homologous sequences for each type. By removing sequence with high identity (>90%) and selecting only bacteria proteins with annotation of either “water transporter” or “glycerol facilitator”, 305 sequences were obtained. A multiple sequence alignment was conducted on the bacteria aquaporin sequences. The aligned sequences had 192 columns (positions) after removing columns containing more than 30% of gaps. For convenience, equivalent residues of each position in AQGP and AQP were represented by their corresponding residues in Glp F and Aqp Z.

Table S1. Cross reference table of positions of multiple sequence alignment treated data set and their corresponding residues in Glp F and Aqp Z. For example, for position 38, the corresponding residues in Glp F and Aqp Z are W48 and F43, respectively.

| Position | Glp F | Aqp Z |
|----------|-------|-------|
|          | Sequence No. | Residue | Sequence No. | Residue |
| 1        | 8 | K | 2 | F |
| 2        | 9 | G | 3 | R |
| 3        | 10 | Q | 4 | K |
| 4        | 11 | C | 5 | L |
| 5        | 12 | I | 6 | A |
| 6        | 13 | A | 7 | A |
| 7        | 14 | E | 8 | E |
| 8        | 15 | F | 9 | C |
| 9        | 16 | L | 10 | F |
| 10 | 17 | G | 11 | G |
|---|---|---|---|---|
| 11 | 18 | T | 12 | T |
| 12 | 19 | G | 13 | F |
| 13 | 20 | L | 14 | W |
| 14 | 21 | L | 15 | L |
| 15 | 22 | I | 16 | V |
| 16 | 23 | F | 17 | F |
| 17 | 24 | F | 18 | G |
| 18 | 25 | G | 19 | G |
| 19 | 26 | V | 20 | C |
| 20 | 27 | G | 21 | G |
| 21 | 28 | C | 22 | S |
| 22 | 29 | V | 23 | A |
| 23 | 30 | A | 24 | V |
| 24 | 31 | A | 25 | L |
| 25 | 32 | L | 26 | A |
| 26 | 33 | K | 27 | A |
| 27 | 34 | V | 28 | G |
| 28 | 35 | A | GAP | - |
| 29 | 36 | G | 29 | F |
| 30 | GAP | - | 35 | G |
| 31 | GAP | - | 36 | F |
| 32 | 42 | W | 37 | A |
| 33 | 43 | E | 38 | G |
| 34 | 44 | I | 39 | V |
| 35 | 45 | S | 40 | A |
| 36 | 46 | V | 41 | L |
| 37 | 47 | I | 42 | A |
| 38 | 48 | W | 43 | F |
| 39 | 49 | G | 44 | G |
| 40 | 50 | L | 45 | L |
| 41 | 51 | G | 46 | T |
| 42 | 52 | V | 47 | V |
| 43 | 53 | A | 48 | L |
| 44 | 54 | M | 49 | T |
| 45 | 55 | A | 50 | M |
|   |   |   |   |   |
|---|---|---|---|---|
|46|56|I|51|A|
|47|57|Y|52|F|
|48|58|L|53|A|
|49|59|T|54|V|
|50|60|A|55|G|
|51|61|G|56|H|
|52|62|V|57|I|
|53|63|S|58|S|
|54|64|G|59|G|
|55|65|A|60|G|
|56|66|H|61|H|
|57|67|L|62|F|
|58|68|N|63|N|
|59|69|P|64|P|
|60|70|A|65|A|
|61|71|V|66|V|
|62|72|T|67|T|
|63|73|I|68|I|
|64|74|A|69|G|
|65|75|L|70|L|
|66|76|W|71|W|
|67|77|L|72|A|
|68|78|F|73|G|
|69|82|D|75|R|
|70|83|K|76|F|
|71|84|R|77|P|
|72|85|K|80|E|
|73|86|V|81|V|
|74|87|I|82|V|
|75|88|P|83|G|
|76|89|F|84|Y|
|77|90|I|85|V|
|78|91|V|86|I|
|79|92|S|87|A|
|80|93|Q|88|Q|
|81|94|V|89|V|
|    |     |     |     |     |
|----|-----|-----|-----|-----|
| 82 | 95  | A   | 90  | V   |
| 83 | 96  | G   | 91  | G   |
| 84 | 97  | A   | 92  | G   |
| 85 | 98  | F   | 93  | I   |
| 86 | 99  | C   | 94  | V   |
| 87 | 100 | A   | 95  | A   |
| 88 | 101 | A   | 96  | A   |
| 89 | 102 | A   | 97  | A   |
| 90 | 103 | L   | 98  | L   |
| 91 | 104 | V   | 99  | L   |
| 92 | 105 | Y   | 100 | Y   |
| 93 | 106 | G   | 101 | L   |
| 94 | 107 | L   | 102 | I   |
| 95 | 129 | V   | 116 | F   |
| 96 | 130 | D   | 117 | A   |
| 97 | 131 | L   | 118 | S   |
| 98 | 132 | A   | 119 | N   |
| 99 | 133 | G   | 120 | G   |
| 100| 134 | T   | 121 | Y   |
| 101| 135 | F   | 122 | G   |
| 102| 136 | S   | 123 | E   |
| 103| 137 | T   | 124 | H   |
| 104| 138 | Y   | 125 | S   |
| 105| 139 | P   | 126 | P   |
| 106| 146 | V   | 132 | L   |
| 107| 147 | Q   | 133 | S   |
| 108| 148 | A   | 134 | A   |
| 109| 149 | F   | 135 | L   |
| 110| 150 | A   | 136 | V   |
| 111| 151 | V   | 137 | V   |
| 112| 152 | E   | 138 | E   |
| 113| 153 | M   | 139 | L   |
| 114| 154 | V   | 140 | V   |
| 115| 155 | I   | 141 | L   |
| 116| 156 | T   | 142 | S   |
| 117| 157 | A   | 143 | A   |
|    |    |    |    |    |
|----|----|----|----|----|
| 118| 158| G  |    |    |
| 119| 159| F  |    |    |
| 120| 160| L  |    |    |
| 121| 161| L  |    |    |
| 122| 162| V  |    |    |
| 123| 163| I  |    |    |
| 124| 164| H  |    |    |
| 125| 165| G  |    |    |
| 126| 166| A  |    |    |
| 127| 167| T  |    |    |
| 128| 168| D  |    |    |
| 129| 169| K  |    |    |
| 130| 176| F  |    |    |
| 131| 177| G  |    |    |
| 132| 178| A  |    |    |
| 133| 179| I  |    |    |
| 134| 180| P  |    |    |
| 135| 181| I  |    |    |
| 136| 182| A  |    |    |
| 137| 183| I  |    |    |
| 138| 184| G  |    |    |
| 139| 185| L  |    |    |
| 140| 186| A  |    |    |
| 141| 187| L  |    |    |
| 142| 188| T  |    |    |
| 143| 189| L  |    |    |
| 144| 190| I  |    |    |
| 145| 191| H  |    |    |
| 146| 192| L  |    |    |
| 147| 193| I  |    |    |
| 148| 194| S  |    |    |
| 149| 195| I  |    |    |
| 150| 196| P  |    |    |
| 151| 197| V  |    |    |
| 152| 198| T  |    |    |
| 153| 199| N  |    |    |
|   |   |   |   |   |
|---|---|---|---|---|
| 154 | 200 | F | 183 | T |
| 155 | 201 | A | 184 | S |
| 156 | 202 | M | 185 | V |
| 157 | 203 | N | 186 | N |
| 158 | 204 | P | 187 | P |
| 159 | 205 | A | 188 | A |
| 160 | 206 | R | 189 | R |
| 161 | 207 | D | 190 | S |
| 162 | 208 | F | 191 | T |
| 163 | 209 | G | 192 | A |
| 164 | 210 | P | 193 | V |
| 165 | 211 | K | 194 | A |
| 166 | 212 | V | 195 | I |
| 167 | 213 | F | 196 | F |
| 168 | 214 | A | 197 | Q |
| 169 | 215 | W | 198 | G |
| 170 | 233 | F | 205 | L |
| 171 | 234 | L | 206 | W |
| 172 | 235 | V | 207 | F |
| 173 | 236 | P | 208 | F |
| 174 | 237 | L | 209 | W |
| 175 | 238 | F | 210 | V |
| 176 | 239 | G | 211 | V |
| 177 | 240 | P | 212 | P |
| 178 | 241 | I | 213 | I |
| 179 | 242 | V | 214 | V |
| 180 | 243 | G | 215 | G |
| 181 | 244 | A | 216 | G |
| 182 | 245 | I | 217 | I |
| 183 | 246 | V | 218 | I |
| 184 | 247 | G | 219 | G |
| 185 | 248 | A | 220 | G |
| 186 | 249 | F | 221 | L |
| 187 | 250 | A | 222 | I |
| 188 | 251 | Y | 223 | Y |
| 189 | 252 | R | 224 | R |
Section 2. General information of multiple sequence alignment treated AQP/AQGP sequences

Single site conservation in AQP/AQGP

In AQGPs, with threshold of conservation rate being 90%, there were only 8 highly conserved sites, including two NPA motif and two other sites (Gly96 in M4 & Arg206 in M7). AQPs appeared to have much more highly conserved sites. With the same threshold, 46 highly conserved positions were identified in AQPs, including all the 8 conserved positions mentioned in AQGPs. Clearly, AQP sequences contained more conserved residues when compared to AQGP sequences.

General network analysis of Aquaporin results

By connecting the positions with SCA scores, we generated a network (referred as SCA network) of bacteria aquaporin proteins of both AQPs and AQGPs. The multiple sequence alignment gave a set of sequences with 192 columns each. Thus, a 192x192 matrix was given by SCA with 18528 ((192x192+192)/2) non-redundant couples (Fig. S1), the SCA score were represented by gradient color from high to low.
Figure S1 192x192 matrix of SCA scores, the dot $i_{j,k}$ represents the SCA score of two positions $j$ and $k$. The value increases evenly as right shows.

Section 3. Comparison between SF residues’ bond angles in mAqpZ & Glp F

The average bond angle degree of each angle in SF residues in mAqpZ and Glp F were computed, elucidated that the local molecular configuration in SF region of mAqpZ was very similar to that of Glp F.
Figure S2 The average bond angle degree of each angle in SF residues in mAqpZ and Glp F

Section 4. Amino acids distributed on and near the channels of Glp F, Aqp Z and mAqpZ

As reported in a previous study\textsuperscript{2}, the electrostatic profile in ar/R region of AQGPs exhibits a more negative charge compared to that of AQP5s. The electrostatic status of the proteins is considered to be correlated with glycerol permeability. In this work, we use Pore-walker\textsuperscript{3} to determine the pore-lining residues of Glp F, Aqp Z and mAqp Z respectively. From the results, residues in vicinity of SF, or ar/R region are presented in Fig. S3. Aqp Z exhibits a more positive-charged profile in the SF region, with two positive-charged residues, Arg and His. In contrast, Glp F has a more negative-charged profile, with a positive-charged Arg and a negative-charged Asp. The Glu152, although not a pore-lining residue, is located very close to the lumen, which contributes to the negative-charged profile of Glp F. In mAqp Z, a negative-charged electrostatic profile is also observed. With a positive-charged Arg and two negative-charged residues, Glu and Asp, the electrostatic profile of mAqp Z is more negative than that of Glp F. Therefore, glycerol is more favorable to be transported through mAqp Z.
Figure S3 Amino acids distributed on and near channels of Glp F, mAqp Z and wtAqp Z. Acidic residues are colored in blue, basic residues in red, polar residues in purple, and non-polar ones in black. The Glu152, although not a pore-lining residue, is located very close to the lumen, as presented above Ala in Glp F’s sequence.

Section 5 Hydrogen bond acceptors in mAqp Z

Hydrogen bond interactions are considered to be important for both water and glycerol conduction\(^{4-5}\). Here, we analyze the structure of pore-lining residues in the vicinity of SF in Fig. S4. These residues act as H-bond acceptors to facilitate both water and glycerol transportation. For instance, Glu138 and Asp190 can form H-bond with glycerol with the carboxyl group on their side chains and make the channel more conductive to glycerol molecules.
Figure S4 Residues acting as hydrogen bond acceptors near SF region.

References

1. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *J Mol Biol* 1990, 215 (3), 403-10.
2. Oliva, R.; Calamita, G.; Thornton, J. M.; Pellegrini-Calace, M., Electrostatics of aquaporin and aquaglyceroporin channels correlates with their transport selectivity. *P Natl Acad Sci USA* 2010, 107 (9), 4135-4140.
3. Pellegrini-Calace, M.; Maiwald, T.; Thornton, J. M., PoreWalker: A Novel Tool for the Identification and Characterization of Channels in Transmembrane Proteins from Their Three-Dimensional Structure. *Plos Comput Biol* 2009, 5 (7).
4. Tajkhorshid, E.; Nollert, P.; Jensen, M. O.; Miercke, L. J. W.; O’Connell, J.; Stroud, R. M.; Schulten, K., Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* 2002, 296 (5567), 525-530.
5. Jensen, M. O.; Tajkhorshid, E.; Schulten, K., The mechanism of glycerol conduction in aquaglyceroporins. *Structure* 2001, 9 (11), 1083-93.