Structural and enzymatic analyses of *Anabaena* heterocyst-specific alkaline invertase InvB

Jin Xie, Hai-Xi Hu, Kun Cai, Ling-Yun Xia, Feng Yang, Yong-Liang Jiang, Yuxing Chen*, Cong-Zhao Zhou*

From the Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China, Hefei Anhui 230027, People's Republic of China.

*To whom correspondence should be addressed: Telephone and Fax: 0086-551-63600406, E-mail: zcz@ustc.edu.cn (C.Z.Z); cyxing@ustc.edu.cn (Y.C.)

Key words: alkaline/neutral invertases, cyanobacteria, crystal structure, catalytic activity, nitrogen fixation, phylogenetic analysis

ABSTRACT

Anabaena sp. PCC 7120 encodes two alkaline/neutral invertases, namely InvA and InvB. Following our recently reported InvA structure, here we report the crystal structure of the heterocyst-specific InvB. Despite sharing an overall structure similar to InvA, InvB possesses a much higher catalytic activity. Structural comparisons of the catalytic pockets revealed that Arg430 of InvB adopts a different conformation, which may facilitate the deprotonation of the catalytic residue Glu415. We propose that the higher activity may be responsible for the vital role of InvB in the heterocyst development and nitrogen fixation. Further phylogenetic analysis combined with activity assays also suggested the role of this highly conserved arginine in plants and cyanobacteria, as well as some proteobacteria living in highly extreme environments.

Sucrose, the major product of photosynthesis in plants and cyanobacteria [1], could be degraded into hexoses or their derivates to provide carbon and energy or to act as signaling molecules, for the growth, development and defense [2-5]. Besides, the hydrolysis of sucrose was reported to be involved in nitrogen fixation in diazotrophic cyanobacteria [6, 7]. More recently,

33 genomic and experimental evidence suggested the existence of sucrose-related genes in
34 proteobacteria [4, 8-12], where sucrose is used as a compatible solute, and its osmotic effect can
35 be doubled by sucrose degradation [5, 11, 13, 14]. Sucrose degradation could be catalyzed by two
36 different enzymes. Sucrose synthase (SUS, *EC 2.4.1.13*) cleaves sucrose into fructose and
37 UDP-glucose or ADP-glucose reversibly, whereas invertase (Inv, *EC 3.2.1.26*) catalyzes the
38 irreversible hydrolysis of sucrose into glucose and fructose [15]. Current understanding indicates
39 that these two enzymes have distinct roles. SUS-catalyzed sucrose degradation is mainly involved
40 in the biosynthesis of structural and storage polysaccharide by providing sugar nucleotides as the
41 precursors [2, 16]. By contrast, Invs play a critical role when there is a demand for carbon and
42 energy [2, 4].

43 There are two classes of invertases, which are initially categorized according to their
44 optimum pH: the acid invertases (Ac-Invs) with an optimum pH of 4.0–5.5, and the
45 alkaline/neutral invertases (A/N-Invs) with an optimum pH of 6.5–8.0 [17, 18]. Notably,
46 A/N-Invs and Ac-Invs differ from each other in the primary sequence. As the members of
47 glycoside hydrolase (GH) 32 family, Ac-Invs adopt an all β structure with a fivefold β-propeller
48 catalytic domain, and adopt a so-called double displacement mechanism for hydrolysis [19-22].
49 Ac-Invs can hydrolyze sucrose and other β-fructose-containing oligosaccharides, therefore they
50 are also termed β-fructofuranosidases. Recently, the first structure of an A/N-Inv, *Anabaena*
51 alkaline invertase InvA was solved, which revealed that A/N-Invs share an overall structure of
52 $(\alpha/\alpha)_6$ barrel differing from Ac-Invs [23]. Structural analysis and enzymatic assays identified the
53 catalytic residues, Asp188 and Glu414, and the general acid-base mechanism for hydrolysis [23].
54 Furthermore, the stringent substrate specificity toward sucrose enabled us to propose that
55 A/N-Invs, which exclusively constitute GH100 family, represent a novel family of glucosidase
56 [23]. Besides, A/N-Invs and Ac-Invs differ a lot in biochemical properties, species distribution
57 and subcellular location [1, 24].

58 Recent studies revealed that sucrose translocated from vegetative cells to heterocysts as the
59 major carbon carrier molecule plays a central role in connecting carbon metabolism and nitrogen

60 fixation in diazotrophic cyanobacteria [7, 15, 25]. Curatti et al. showed that SUS was specially
61 involved in the cleavage of sucrose in vegetative cells [7, 25]. Bioinformatics analysis revealed
62 that nearly all genome-sequenced cyanobacteria encode only A/N-Invs, but not Ac-Invs. Vargas et
63 al. identified two A/N-Invs, namely InvA and InvB, in the filamentous nitrogen-fixing
64 cyanobacterium *Anabaena* sp. PCC 7120 [24]. As the detected optimum pHs were 7.8 and 6.7,
65 they further classified InvA and InvB into alkaline and neutral invertase, respectively. Both InvA
66 and InvB are expressed in vegetative cells; however, only InvB is specifically expressed in
67 heterocysts [6, 26]. Moreover, knockout of *invB* (*alr0819*) but not *invA* (*alr1521*) blocked the
68 growth of cyanobacteria at diazotrophic conditions, which suggested that InvB is essential for
69 diazotrophic growth [6, 26]. However, why InvB is specifically involved in the diazotrophic
70 growth remains not clear.

71 It is recognized that A/N-Invs widely spread in plants and cyanobacteria, and plant A/N-Inv
72 might evolve from cyanobacteria after the endosymbiotic origin of chloroplasts. [1, 24] Most
73 recently, Wan et al. performed detailed phylogenetic analysis of Ac-Invs and A/N-Invs from
74 cyanobacteria and especially green algae, lower and higher plants, and concluded that A/N-Invs
75 are evolutionarily and functionally more stable than Ac-Invs, possibly due to their roles in
76 maintaining cytosolic sugar homeostasis for cellular function [13]. However, there is no report
77 concerning the A/N-Inv activity or detailed phylogenetic analysis in other bacteria except for
78 cyanobacteria so far.

79 Here we report the crystal structures of *Anabaena* InvB and its complex with sucrose.
80 Despite sharing an overall structure similar to the subunit of InvA, the loop$_{\beta2-\beta3}$ close to the
81 substrate binding pocket, in addition to the residue Arg430 of InvB differs a lot. Activity assays
82 revealed that the optimum pH of InvB is 7.7, suggesting that InvB is also an alkaline invertase
83 like InvA, but not a neutral invertase as previously reported [24]. Compared to InvA, InvB has a
84 higher affinity towards sucrose and a higher catalytic efficiency (k_{cat}/K_m). Structural analyses
85 combined with activity assays revealed that Arg430 contributes to the high catalytic efficiency of
86 InvB, which is necessary for heterocyst development and nitrogen fixation. Moreover,

87 phylogenetic analysis enabled us to extend the species distribution of A/N-Invs to proteobacteria
88 and further confirmed the role of the conserved arginine.
89
90 **Materials and Methods**
91
92 *Cloning, expression, and purification*—The gene of InvB (NCBI accession number
93 CAC85155.1, 483 residues) was amplified from the genomic DNA of *Anabaena* sp. PCC 7120.
94 The full-length *invB* and their mutants were individually cloned into a pET29a-derived vector
95 with an N-terminal His_6 tag. Likewise, the truncated InvB covering residues Thr9–Ser457 was
96 constructed for crystallization. Both the wild-type and mutant proteins were overexpressed in
97 *Escherichia coli* strain BL21 (DE3) (Novagen). Cells were grown in LB culture medium (10 g of
98 tryptone, 5 g of yeast extract and 10 g of NaCl per liter) containing 30 μg/ml kanamycin at 37°C
99 until the $A_{600\ nm}$ reached ~0.6. Then protein expression was induced with 0.2 mM isopropyl
100 β-D-1-thiogalactopyranoside at 37°C for another 4 hr. Cells were collected and resuspended in 40
101 ml of lysis buffer (20 mM HEPES, pH 8.0, 100 mM NaCl). After 12 min of sonication and
102 centrifugation at 12,000 × g for 25 min, the supernatant containing the target protein was loaded
103 onto a nickel-nitrilotriacetic acid column (Qiagen) equilibrated with the binding buffer (20 mM
104 HEPES, pH 8.0, 100 mM NaCl). The target protein was eluted with 300 mM imidazole, and
105 further applied to a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated with
106 the binding buffer. Samples for enzymatic activity assays were collected at the highest peak
107 fractions without concentration and stored at -80°C with 50% glycerol. The purity of protein was
108 assessed by gel electrophoresis.
109 Analytical gel filtration chromatography was used to determine the molecular weight of
110 InvB in solution by a Superdex 75 10/300 GL column (GE Healthcare). The following standard
111 molecular markers were used for calibration: ribonuclease A (13.7 kD), ovalbumin (43.0 kD),
112 conalbumin (75.0 kD), aldolase (158.0 kD), ferritin (440.0 kD) and thyroglobulin (669.0 kD).
113 The selenomethionine (SeMet)-substituted full-length and truncated InvB were

114 overexpressed in *E. coli* B834 (DE3) (Novagen). Transformed cells were first cultured in LB
115 medium at 37°C overnight, then harvested and washed twice with the M9 medium [27]. Then the
116 cells were cultured in SeMet medium (M9 medium with 50 mg/l SeMet and other essential amino
117 acids) to an $A_{600\ nm}$ of 0.6–0.8. The following steps in protein expression and purification were the
118 same as those for the native protein in addition to adding 5 mM β-mercaptoethanol during
119 purification.

120 *Crystallization, data collection, and processing*—Both native and SeMet-substituted InvB
121 were concentrated to 5–10 mg/ml by ultrafiltration (Millipore) for crystallization. Crystals were
122 grown at 16°C using the hanging drop vapor diffusion method, with a drop of 1μl protein solution
123 mixed with an equal volume of reservoir solution. Microseeding was adopted in crystal
124 optimization. Crystals were obtained against the reservoir solution of 10% polyethylene glycol
125 6000, 0.1 M sodium citrate, pH 5.6 for the native and SeMet-substituted full-length protein, and
126 0.7 M lithium chloride, 6% polyethylene glycol 6000 and 0.1 M MES, pH 6.0 for the
127 SeMet-substituted truncated InvB. The crystals were transferred to cryoprotectant (reservoir
128 solution supplemented with 30% glycerol) and flash-cooled with liquid nitrogen. For soaking
129 experiments, crystals were transferred to 2 μl of mother liquor containing 30% saturated sucrose
130 and then 3 min later flash-cooled directly, with sucrose as the cryoprotectant. X-ray diffraction
131 data were collected at 100 K using beamline BL17U with an ADSC Q315r CCD detector at the
132 Shanghai Synchrotron Radiation Facility (SSRF). All of the diffraction data were integrated and
133 scaled with the program HKL2000 [28].

134 *Structure determination and refinement*—The crystal structure of SeMet-substituted
135 truncated InvB was determined using the single-wavelength anomalous dispersion (SAD) method
136 [29]. The AutoSol program of PHENIX [30] was used to search the selenium atoms and to
137 calculate the phase. Then automatic model building was carried out using Autobuild in PHENIX.
138 The initial model was refined using the maximum likelihood method implemented in
139 REFMAC5[31] of CCP4i program suite [32] and rebuilt interactively using the program COOT
140 [33]. The crystal data of InvB complexed with sucrose were refined against the apo-from InvB

141 structure. The final models were evaluated with the web service MolProbity
142 (http://molprobity.biochem.duke.edu) [34]. Crystallographic parameters were listed in Table 1.
143 The $|F_o|-|F_c|$ omit electron density map of the ligand contoured at 3.0 σ was calculated by
144 PHENIX. The interface areas were calculated by PDBePISA [35]. All structure figures were
145 prepared with PyMOL [36].

146 *Enzymatic assays*—The pH dependence of invertase activity of recombinant InvB was
147 measured in 40 mM MES (pH 5.5–6.5), HEPES (pH 7.0–8.0), Bicine (pH 8.5–9.0) and
148 glycine-NaOH (pH 9.5) buffer containing 100 mM NaCl. Subsequent enzymatic assays were
149 carried out at pH 7.9 [23] and pH 7.7 for InvA and InvB, respectively. A 50 µl reaction mixture
150 containing 100 mM sucrose and an appropriate volume of wild-type enzymes or the mutants was
151 incubated at 30°C for different times, and then stopped by heating at 95°C for 10 min. After
152 centrifuged at 12,000 × g for 10 min, the supernatant was diluted 5-fold for HPLC analysis by
153 Agilent 1200 Series coupled with an evaporative light scattering detector (ELSD, Alltech
154 2000ES). The samples were injected in volumes of 10 µl onto a Prevail™ Carbohydrate ES
155 column (4.6 x 250mm, 5µm, GRACE) and the column temperature was kept at 25°C. An
156 acetonitrile/water (70:30, v/v) solution was used as the mobile phase at 1 ml/min. For ELSD, the
157 temperature of the detector nebulizer was set to 85°C, and the gas flow was 2.5 l/min. The kinetic
158 determinations of wild-type enzymes and the mutants were performed at different sucrose
159 concentrations, and the K_m and k_{cat} values were calculated by nonlinear fitting to the
160 Michaelis-Menten equation using the program Origin 8. Three independent determinations were
161 made to calculate the means and standard deviations.

162 *Phylogenetic analysis*—The sequence of InvB was used in the BLAST search against the
163 non-redundant protein sequence database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). In total 60
164 A/N-Inv homologues from plants including green algae, cyanobacteria and proteobacteria were
165 selected to do multiple sequence alignment by the program MultAlin [37] and the result was
166 visualized by ESPript 3.0 [38]. Then the rooted phylogenetic tree was constructed using the
167 neighbor-joining method or maximum likelihood method by MEGA 6 [39] with the α-1,6

168 glucosidase domain of glycogen debranching enzymye of *Giardia lamblia* ATCC 50803 as the
169 outgroup. Confidence limits were assessed by bootstrap analysis using 1,000 replicates.
170
171 **Results**
172
173 *Overall structure of InvB*—We obtained crystals of the full-length native and
174 SeMet-substituted InvB, but failed in optimizing because of the twinned crystals. Partial
175 proteolysis combined with multiple sequence alignment indicated that both N- and C-termini are
176 not conserved and most likely unstructured. Thus a truncated version of InvB (residues
177 Thr9−Ser457) was constructed in a similar way to that of InvA [23], and eventually the
178 SeMet-substituted crystal structure of the apo-form was successfully determined at 1.93 Å
179 resolution in the space group *C2*. Afterwards, we obtained the sucrose-complexed crystals by
180 soaking crystals of the truncated InvB with sucrose. The 1.95 Å complex structure of InvB with
181 sucrose was solved by using the apo-form SeMet-substituted InvB structure as the initial model.
182 The parameters for data collection and structure determination are listed in Table 1.
183 In the sucrose-complexed structure (InvB-Suc), each asymmetric unit contains two
184 molecules of InvB, which form a stable dimer (Fig. 1a) with a buried interface area of ~5500 $Å^2$
185 as calculated by PDBePISA. The dimerization in crystal structure is consistent with the result
186 obtained from the size exclusion chromatography (Fig. 1b). In fact, InvA exists as a hexamer
187 composed of three dimers [23]. Besides, the plant A/N-Invs also assemble into oligomers, either
188 tetramers [40-43] or octamers [44]. Therefore, we propose that the dimer may be the basic
189 structural unit of these A/N-Invs. The InvB monomer has an overall structure similar to InvA,
190 both of which possess an $(\alpha/\alpha)_6$-barrel core structure composed of 12 α-helices and an insertion
191 structure containing helices α8, α9, η1 and η2 (Fig. 1c) [23].
192 The apo-form structure of InvB shares an overall structure almost identical to that of the
193 sucrose complex, with a root mean square deviation (RMSD) of 0.52 Å over 856 Cα atoms of the
194 dimer. Notably, a glycerol molecule, which was used as the cryoprotectant, occupies the substrate

195 binding pocket.

196 InvB has a high structural similarity to InvA with a RMSD of 0.98 Å over 414 Cα atoms. 197 The loop between β2 and β3 strands (designated as loop$_{\beta2-\beta3}$), which is close to the substrate 198 binding pocket, in InvB is different from that in InvA (Fig. 1d). Superposition of the dimeric InvB 199 against the corresponding dimer of the hexameric InvA revealed a difference of about 10° for the 200 inter-subunit angles (Fig. S1a). In consequence, the insertion moieties that contribute to the 201 majority of the inter-dimer interactions of InvA hexamer [23] shift away, resulting in a dimeric 202 form of InvB. Besides, sequence analysis revealed that the residues at the inter-dimer interface of 203 InvA show a relatively lower sequence conservation among different A/N-Invs (Fig. S1b). 204 Notably, the charged Arg251 and Glu286 that form a couple of interactions with Arg422 205 from the neighboring dimer in InvA (Fig. S1c) are missing in InvB.

206 *The active-site pocket*—In the sucrose-complexed structure InvB-Suc, each subunit binds 207 one sucrose molecule (Fig. 2) which adopts a same conformation as that of subunit C of InvA-Suc 208 structure [23]. InvB shares a nearly identical active-site pocket as InvA except for Arg430, Asn47 209 and Tyr48 (Fig. 3a, S2). Site-directed mutagenesis combined with activity assays showed that 210 Asp189 and Glu415, corresponding to Asp188 and Glu414 of InvA, are catalytic residues. As 211 stated above, superposition revealed that InvA and InvB mainly differ in the loop$_{\beta2-\beta3}$ close to the 212 substrate binding pocket (Fig. 1d). In subunit A of InvA-Suc structure, the loop$_{\beta2-\beta3}$ which can be 213 fully traced in the electron density map inserts into the pocket through Asn46 and Tyr47 to 214 participate in the binding of sucrose, and Tyr47 is also stabilized by π-stacking with Arg429 (Fig. 215 3a). But in the corresponding loop of InvB-Suc structure, residues from Ala41 to Leu46 are 216 missing in the electron density map. Besides, residues Asn47 and Tyr48 of InvB, which are 217 corresponding to Asn46 and Tyr47 of InvA, shift outwards from the catalytic pocket. Structural 218 comparison combined with sequence alignment revealed several other factors that determine the 219 different conformations of loop$_{\beta2-\beta3}$ in InvA and InvB (Fig. 3a, 3b). InvB possesses longer β3 and 220 β4 strands and a preceding β-turn between Asn47 and Asn50, whereas polar interactions of Ala43 221 with Arg429, in addition to Gln38 with Ser22 and Leu19 could be found in InvA. Moreover, the

222 residue Pro40 makes the loop of InvA adopt a relatively rigid conformation. In contrast, InvB has
223 a more flexible loop$_{\beta2-\beta3}$, a part of which is invisible in the electron density map.

224 *Optimum pH of InvB*—To compare the catalytic properties of InvB with InvA, we first
225 performed a series of enzymatic assays. We detected that InvB has an optimum pH at 7.7 (Fig. 4),
226 which is close to the optimum pH of InvA [23], but different from the result reported by Vargas et
227 al [24]. In fact, an earlier report of Schilling et al. identified an alkaline invertase with an
228 optimum pH of 7.5 to 7.8 in *Anabaena* sp. ATCC 29413, which is equivalent to InvB for their
229 involvement with heterocysts [45]. Therefore InvB should be also an alkaline invertase like InvA,
230 but not the neutral invertase reported before [24].

231 *Key residue contributing to the enzymatic properties of InvB*—Although InvA and InvB have
232 a similar optimum pH, InvB shows an obviously higher catalytic activity than InvA between pH
233 6.5 and pH 9.0 (Fig. 4). Further kinetic determinations revealed that InvB has a higher affinity
234 towards sucrose and a higher catalytic efficiency (k_{cat}/K_m) of about five times to that of InvA
235 (Table 2). As stated above, Arg430 of InvB close to the catalytic glutamate residues has a
236 different orientation, compared to its counterpart residue Arg429 of InvA. From the InvB
237 structure complexed with sucrose, Arg430 forms a salt bridge with the catalytic residue Glu415,
238 and also forms five hydrogen bonds with Glu415, Gln433 and the sucrose molecule (Fig. 5a). In
239 the apo-form InvB structure, Arg430 possesses the same orientation (Fig. 5b). Similar to the
240 previous proposals [46, 47], we presume that Arg430 could provide the positive charge to
241 decrease the pK_a of the catalytic residue Glu415, which is more susceptible for the deprotonation
242 to trigger the nucleophilic attack. But in the two InvA structures complexed with sucrose
243 molecules adopting different conformation [23], the corresponding residue Arg429 deviates away
244 from Glu414 and is fixed by polar interactions with Ala43 and a water molecule, in addition to
245 π-stacking against Tyr47 (Fig. 5c, 5d). Consequently, the activation of Glu414 in InvA is driven
246 by the water molecule stabilized by Arg429 and Lys364. As expected, the mutant R430A of InvB
247 completely loses the hydrolysis activity, whereas the mutant R429A of InvA maintains a residual
248 activity of about 9.3% to the wild-type protein (Table 2). Notably, upon mutation of the arginine

249 to asparagine, the enzymatic activities of InvB and InvA decreased to about 1.6% and 19.8%,
250 respectively (Table 2). All together, the residue Arg430 is important for the activity of InvA and
251 InvB; and moreover, its conformation may also make InvB differ from InvA in catalytic
252 efficiency.

253

254 **Discussion**

255

256 *The higher activity of InvB might be necessary for the heterocyst development and nitrogen*
257 *fixation*—Recent studies reported that *Anabaena* sp. PCC 7120 *invA* is only expressesed in
258 vegetative cells, whereas *invB* is expressesed in both vegetative cells and heterocysts [6]. The
259 spatiotemporal expression assay showed that the amount of InvB increased higher than that of
260 InvA and particularly accumulated in heterocysts upon deprivation of combined nitrogen [26].
261 Interestingly, Vargas et al. found a possible DNA binding site of NtcA, a global nitrogen
262 responsive regulator, in the promoter region of *invB* [6]. Besides, the expression increase of *invB*
263 upon combined nitrogen deprivation was impaired in the *ntcA* mutant strain [26]. These findings
264 implied that the expression of InvB may be regulated by nitrogen starvation. Moreover, knock-out
265 of *invB* impaired diazotrophic growth, but no affect was detected in *invA* mutant strain [6, 26].

266 Nitrogen fixation requires lots of ATP, reductant, and carbon skeletons, which come from the
267 carbohydrate metabolism, most likely starting with sucrose imported from vegetative cells [7, 48].
268 As there is no SUS or Ac-Inv activity in heterocysts, A/N-Inv, namely InvB becomes vital for
269 nitrogen fixation. What's more, rather than InvA, InvB of higher activity could fulfill the
270 necessity for large amount of carbon and energy via sucrose degradation. Moreover, the hexoses
271 produced from sucrose could also be processed through the oxidative pentose pathway to yield
272 α-ketoglutarate and NADPH. Notably, the reductant NADPH is necessary for the nitrogenase, and
273 also needed for the synthesis of cell-envelope components, such as polysaccharide and
274 glycolipids of heterocysts [26, 48].

275 *Extend the species distribution of A/N-Invs to proteobacteria*—To date, the prokaryotic

276 A/N-Invs are only found in cyanobacteria. To further decipher the evolutionary hints of A/N-Invs,
277 we performed the phylogenetic analysis, which classified A/N-Invs into five major groups,
278 including two groups in cyanobacteria, two groups in proteobacteria, and one group in plants (Fig.
279 6a). Notably, in addition to the previously identified clades in cyanobacteria and plants [24, 49],
280 we identified two new clades of A/N-Invs in proteobacteria. Sequence analysis suggested that
281 proteobacteria A/N-Invs share 41%–54% sequence identities with InvB and harbor conserved
282 catalytic and substrate binding residues (Fig. S3). Two distant cyanobacteria groups representing
283 unicellular and filamentous cyanobacteria, respectively, are separated by proteobacteria A/N-Invs
284 from chemolithoautotrophic bacteria or anoxygenic photosynthetic bacteria. In addition, these
285 proteobacteria A/N-Invs can be divided into two sister clades. The inconsistence of the topology
286 between A/N-Invs phylogeny and the organismal phylogeny suggested that proteobacteria are
287 likely to have acquired A/N-Inv genes from cyanobacteria through horizontal gene transfer,
288 similar to other sucrose-metabolizing enzymes [4, 12, 50, 51].

289 In fact, cyanobacteria and proteobacteria are almost the only prokaryotes known to
290 synthesize and metabolize sucrose [4, 8-12]. In agreement with it, here we identified putative
291 A/N-Invs in proteobacteria. Specifically, multiple sequence alignment revealed that all A/N-Invs
292 share a conserved arginine residue corresponding to Arg430 of InvB except for the first clade of
293 proteobacteria A/N-Invs substituted by asparagine (Fig. 6b). Our results showed that this arginine
294 is important for the invertase activity, and the asparagine mutant only retains weak activity. Base
295 on the phylogenetic analysis, we propose that the acquisition of the A/N-Inv genes by
296 proteobacteria probably occurred in a single event and in the early period during the evolution of
297 A/N-Invs among cyanobacteria. Thus most proteobacteria A/N-Invs (clade I) kept a proposed
298 ancient asparagine at Arg430 site with a relatively lower activity. However, the proteobacteria
299 from the second clade (clade II) that live in highly extreme environments, such as hypersalinity
300 (up to 4 M NaCl) [52, 53], extremely alkaline pH (optimum at 10.0) [53] or cold temperature (as
301 low as -5°C) [54, 55], changed the asparagine into arginine. This substitution might enhance
302 sucrose hydrolysis to provide osmoprotective substances and energy in order to overcome the

303 harsh environments, and it was also adopted in cyanobacteria and plants with high demand of
304 sucrose metabolism.
305
306 **Acknowledgments:**

307 This work was supported by the National Natural Science Foundation of China (Grants No.
308 31630001, 31370757 and 31500598). We appreciate the staff at the Shanghai Synchrotron
309 Radiation Facility (SSRF) and the Core Facility Center for Life Sciences in University of Science
310 and Technology of China for technical assistance. We are grateful to all of the developers of the
311 CCP4 Suite, PyMOL, PHENIX, ESPript and MEGA.
312
313 **Conflict of interest:**

314 The authors declare that they have no conflicts of interest with the contents of this article.
315
316 **Author Contributions:**

317 J.X., H.X.H., Y.C. and C.Z.Z. designed the study; J.X., K.C., L.Y.X. and F.Y. performed the
318 experiments; J.X., H.X.H., Y.L.J., Y.C. and C.Z.Z. analyzed the data; J.X., Y.C. and C.Z.Z. wrote
319 the paper.
320
321 **REFERENCES**
322

323 1. Vargas, W. A. and Salerno, G. L. (2010) The Cinderella story of sucrose hydrolysis:
324 Alkaline/neutral invertases, from cyanobacteria to unforeseen roles in plant cytosol and organelles.
325 *Plant Sci* **178**, 1-8.
326 2. Ruan, Y.-L., Jin, Y., Yang, Y.-J., Li, G.-J. and Boyer, J. S. (2010) Sugar input, metabolism,
327 and signaling mediated by invertase: roles in development, yield potential, and response to
328 drought and heat. *Mol Plant* **3**, 942-955.
329 3. Koch, K. (2004) Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar
330 sensing and plant development. *Curr Opin Plant Biol* **7**, 235-246.

331 4. Salerno, G. L. and Curatti, L. (2003) Origin of sucrose metabolism in higher plants: when,
332 how and why?. *Trends Plant Sci* **8**, 63-69.
333 5. Ruan, Y.-L. (2014) Sucrose metabolism: gateway to diverse carbon use and sugar signaling.
334 *Annu Rev Plant Biol* **65**, 33-67.
335 6. Vargas, W. A., Nishi, C. N., Giarrocco, L. E. and Salerno, G. L. (2011) Differential roles of
336 alkaline/neutral invertases in *Nostoc* sp. PCC 7120: Inv-B isoform is essential for diazotrophic
337 growth. *Planta* **233**, 153-162.
338 7. Curatti, L., Flores, E. and Salerno, G. (2002) Sucrose is involved in the diazotrophic
339 metabolism of the heterocyst-forming cyanobacterium *Anabaena* sp. *FEBS Lett* **513**, 175-178.
340 8. Lunn, J. E. (2002) Evolution of sucrose synthesis. *Plant Physiol* **128**, 1490-500.
341 9. Diricks, M., De Bruyn, F., Van Daele, P., Walmagh, M. and Desmet, T. (2015) Identification
342 of sucrose synthase in nonphotosynthetic bacteria and characterization of the recombinant
343 enzymes. *Appl Microbiol Biotechnol* **99**, 8465-74.
344 10. Syamaladevi, D. P., Jayaraman, N. and Subramonian, N. (2013) Evolutionary relationship
345 between available homologous sequences of sucrose phosphate phosphatase (SPP) enzyme. *Sugar*
346 *Tech* **15**, 136-144.
347 11. Empadinhas, N. and da Costa, M. S. (2008) Osmoadaptation mechanisms in prokaryotes:
348 distribution of compatible solutes. *Int Microbiol* **11**, 151-61.
349 12. Wu, R., Asención Diez, M. D., Figueroa, C. M., Machtey, M., Iglesias, A. A., Ballicora, M.
350 A. and Liu, D. (2015) The crystal structure of *Nitrosomonas europaea* sucrose synthase reveals
351 critical conformational changes and insights into sucrose metabolism in prokaryotes. *J Bacteriol*
352 **197**, 2734-2746.
353 13. Wan, H., Wu, L., Yang, Y., Zhou, G. and Ruan, Y. L. (2017) Evolution of sucrose
354 metabolism: the dichotomy of invertases and beyond. *Trends Plant Sci* **23**:163-177
355 14. Khmelenina, V. N., Kalyuzhnaya, M. G., Sakharovsky, V. G., Suzina, N. E., Trotsenko, Y. A.
356 and Gottschalk, G. (1999) Osmoadaptation in halophilic and alkaliphilic methanotrophs. *Arch*
357 *Microbiol* **172**, 321-329.

358 15. Cumino, A. C., Marcozzi, C., Barreiro, R. and Salerno, G. L. (2007) Carbon cycling in
359 *Anabaena* sp. PCC 7120. Sucrose synthesis in the heterocysts and possible role in nitrogen
360 fixation. *Plant Physiol* **143**, 1385-1397.
361 16. Coleman, H. D., Yan, J. and Mansfield, S. D. (2009) Sucrose synthase affects carbon
362 partitioning to increase cellulose production and altered cell wall ultrastructure. *Proc Natl Acad*
363 *Sci U S A* **106**, 13118-13123.
364 17. Sturm, A. (1999) Invertases. Primary structures, functions, and roles in plant development
365 and sucrose partitioning. *Plant Physiol* **121**, 1-8.
366 18. Liu, S., Lan, J., Zhou, B., Qin, Y., Zhou, Y., Xiao, X., Yang, J., Gou, J., Qi, J., Huang, Y. and
367 Tang, C. (2015) HbNIN2, a cytosolic alkaline/neutral-invertase, is responsible for sucrose
368 catabolism in rubber-producing laticifers of Hevea brasiliensis (para rubber tree). *New Phytol* **206**,
369 709-725.
370 19. Lammens, W., Le Roy, K., Van Laere, A., Rabijns, A. and Van den Ende, W. (2008) Crystal
371 structures of *Arabidopsis thaliana* cell-wall invertase mutants in complex with sucrose. *J Mol*
372 *Biol* **377**, 378-385.
373 20. Alberto, F., Jordi, E., Henrissat, B. and Czjzek, M. (2006) Crystal structure of inactivated
374 *Thermotoga maritima* invertase in complex with the trisaccharide substrate raffinose. *Biochem J*
375 **395**, 457-462.
376 21. Alberto, F., Bignon, C., Sulzenbacher, G., Henrissat, B. and Czjzek, M. (2004) The
377 three-dimensional structure of invertase (beta-fructosidase) from *Thermotoga maritima* reveals a
378 bimodular arrangement and an evolutionary relationship between retaining and inverting
379 glycosidases. *J Biol Chem* **279**, 18903-18910.
380 22. Reddy, A. and Maley, F. (1996) Studies on identifying the catalytic role of Glu-204 in the
381 active site of yeast invertase. *J Biol Chem* **271**, 13953-13958.
382 23. Xie, J., Cai, K., Hu, H.-X., Jiang, Y.-L., Yang, F., Hu, P.-F., Cao, D.-D., Li, W.-F., Chen, Y.
383 and Zhou, C.-Z. (2016) Structural analysis of the catalytic mechanism and substrate specificity of
384 *Anabaena* alkaline invertase InvA reveals a novel glucosidase. *J Biol Chem* **291**, 25667-25677.

385 24. Vargas, W., Cumino, A. and Salerno, G. L. (2003) Cyanobacterial alkaline/neutral invertases.
386 Origin of sucrose hydrolysis in the plant cytosol?. *Planta* **216**, 951-960.
387 25. Curatti, L., Giarrocco, L. and Salerno, G. L. (2006) Sucrose synthase and RuBisCo
388 expression is similarly regulated by the nitrogen source in the nitrogen-fixing cyanobacterium
389 Anabaena sp. *Planta* **223**, 891-900.
390 26. López-Igual, R., Flores, E. and Herrero, A. (2010) Inactivation of a heterocyst-specific
391 invertase indicates a principal role of sucrose catabolism in heterocysts of *Anabaena* sp. *J*
392 *Bacteriol* **192**, 5526-5533.
393 27. Walden, H. (2010) Selenium incorporation using recombinant techniques. *Acta Crystallogr*
394 *D Biol Crystallogr* **66**, 352-357.
395 28. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in
396 oscillation mode. *Methods Enzymol* **276**, 307-326.
397 29. Brodersen, D. E., de La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J. and Kjeldgaard,
398 M. (2000) Applications of single-wavelength anomalous dispersion at high and atomic resolution.
399 *Acta Crystallogr D Biol Crystallogr* **56**, 431-441.
400 30. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J.
401 J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner,
402 R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010)
403 PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta*
404 *Crystallogr D Biol Crystallogr* **66**, 213-221.
405 31. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A.,
406 Winn, M. D., Long, F. and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular
407 crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367.
408 32. Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein
409 crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763.
410 33. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta*
411 *Crystallogr D Biol Crystallogr* **60**, 2126-2132.

412 34. Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G.
413 J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) MolProbity: all-atom structure
414 validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.
415 35. Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from
416 crystalline state. *J Mol Biol* **372**, 774-797.
417 36. Delano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San
418 Carlos, CA, USA.
419 37. Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids*
420 *Res* **16**, 10881-10890.
421 38. Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new
422 ENDscript server. *Nucleic Acids Res* **42**, W320-W324.
423 39. Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: molecular
424 evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729.
425 40. Lin, C.-L., Lin, H.-C., Wang, A.-Y. and Sung, H.-Y. (1999) Purification and characterization
426 of an alkaline invertase from shoots of etiolated rice seedlings. *New Phytol* **142**, 427-434.
427 41. Chen, J. Q. and Black, C. C. (1992) Biochemical and immunological properties of alkaline
428 invertase isolated from sprouting soybean hypocotyls. *Arch Biochem Biophys* **295**, 61-69.
429 42. Liu, C.-C., Huang, L.-C., Chang, C.-T. and Sung, H.-Y. (2006) Purification and
430 characterization of soluble invertases from suspension-cultured bamboo (*Bambusa edulis*) cells.
431 *Food Chem* **96**, 621-631.
432 43. Van den Ende, W. and Van Laere, A. (1995) Purification and properties of a neutral invertase
433 from the roots of *Cichorium intybus*. *Physiol Plant* **93**, 241-248.
434 44. Lee, H. S. and Sturm, A. (1996) Purification and characterization of neutral and alkaline
435 invertase from carrot. *Plant Physiol* **112**, 1513-1522.
436 45. Schilling, N. and Ehrnsperger, K. (1985) Cellular differentiation of sucrose metabolism in
437 *Anabaena* variabilis. *Z Naturforsch C* **40**, 776-779.
438 46. Guillén Schlippe, Y. V. and Hedstrom, L. (2005) A twisted base? The role of arginine in

439 enzyme-catalyzed proton abstractions. *Arch Biochem Biophys* **433**, 266-278.

440 47. Harris, T. K. and Turner, G. J. (2002) Structural basis of perturbed p*K*a values of catalytic
441 groups in enzyme active sites. *IUBMB life* **53**, 85-98.

442 48. Haselkorn, R. (2007) Heterocyst differentiation and nitrogen fixation in cyanobacteria. In
443 *Associative and endophytic nitrogen-fixing bacteria and cyanobacterial associations*, pp. 233-255.
444 Edited by Elmerich C., Newton W. E.. Dordrecht, The Netherlands: Springer.

445 49. Vargas, W. A., Pontis, H. G. and Salerno, G. L. (2008) New insights on sucrose metabolism:
446 evidence for an active A/N-Inv in chloroplasts uncovers a novel component of the intracellular
447 carbon trafficking. *Planta* **227**, 795-807.

448 50. Kolman, M. A., Torres, L. L., Martin, M. L. and Salerno, G. L. (2012) Sucrose synthase in
449 unicellular cyanobacteria and its relationship with salt and hypoxic stress. *Planta* **235**, 955-964.

450 51. Chua, T. K., Bujnicki, J. M., Tan, T. C., Huynh, F., Patel, B. K. and Sivaraman, J. (2008)
451 The structure of sucrose phosphate synthase from *Halothermothrix orenii* reveals its mechanism
452 of action and binding mode. *Plant Cell* **20**, 1059-1072.

453 52. Sorokin, D. Y., Tourova, T. P., Galinski, E. A., Muyzer, G. and Kuenen, J. G. (2008)
454 *Thiohalorhabdus denitrificans* gen. nov., sp. nov., an extremely halophilic, sulfur-oxidizing,
455 deep-lineage gammaproteobacterium from hypersaline habitats. *Int J Syst Evol Microbiol* **58**,
456 2890-2897.

457 53. Sorokin, D. Y., Tourova, T. P., Lysenko, A. M., Mityushina, L. L. and Kuenen, J. G. (2002)
458 *Thioalkalivibrio thiocyanoxidans* sp. nov. and *Thioalkalivibrio paradoxus* sp. nov., novel
459 alkaliphilic, obligately autotrophic, sulfur-oxidizing bacteria capable of growth on thiocyanate,
460 from soda lakes. *Int J Syst Evol Microbiol* **52**, 657-664.

461 54. Jones, R. D., Morita, R. Y., Koops, H.-P. and Watson, S. W. (1988) A new marine
462 ammonium-oxidizing bacterium, *Nitrosomonas cryotolerans* sp. nov. *Can J Microbiol* **34**,
463 1122-1128.

464 55. Rice, M. C., Norton, J. M., Stein, L. Y., Kozlowski, J., Bollmann, A., Klotz, M. G.,
465 Sayavedra-Soto, L., Shapiro, N., Goodwin, L. A., Huntemann, M., Clum, A., Pillay, M., Varghese,

466 N., Mikhailova, N., Palaniappan, K., Ivanova, N., Mukherjee, S., Reddy, T. B. K., Yee Ngan, C.,
467 Daum, C., Kyrpides, N. and Woyke, T. (2017) Complete genome sequence of *Nitrosomonas*
468 *cryotolerans* ATCC 49181, a phylogenetically distinct ammonia-oxidizing bacterium isolated
469 from arctic waters. *Genome Announc* **5**, e00011-17.
470

471 TABLE 1. Crystal parameters, data collection, and structure refinement

	SeInvB	**SeInvB-Suc**
Data collection		
Space group	***C2***	***C2***
Unit cell		
a, b, c (Å)	166.6, 82.4, 94.8	166.9, 82.2, 97.4
α, β, γ (°)	118.9	119.1
Resolution range (Å)	50.00–1.93 (2.00–1.93)[h]	50.00–1.95 (2.02–1.95)
Unique reflections	82,269 (8,274)	80,206 (8,173)
R_{merge}[a]	0.078 (0.505)	0.096 (0.537)
$I/\sigma I$	15.1 (2.3)	13.1 (2.4)
Completeness (%)	97.6 (98.9)	95.7 (98.1)
Average redundancy	2.7 (2.7)	3.5 (3.5)
Structure refinement		
Resolution range (Å)	31.41–1.93	31.35–1.95
R_{work}[b]/R_{free}[c]	0.206/0.228	0.210/0.239
Number of protein atoms	7,206	7,245
Number of water atoms	287	184
RMSD[d] bond lengths (Å)	0.005	0.005
RMSD bond angles (°)	0.914	1.114
Average B-factors ($Å^2$)		
Protein	41.3	37.0
Ligand	42.0 (GOL in subunit A only)	32.6/33.9 (Suc in both subunits)
Ramachandran plot[e] (residues, %)		
Most favored	98.58	98.96
Additional allowed	1.42	1.04
Outliers	0	0
MolProbity clashscore[f]/score percentile	0.87/100th percentile (N=742, 1.93±0.25 Å)	2.06/100th percentile (N=821, 1.95±0.25 Å)
MolProbity score[g]/score percentile	0.77/100th percentile (N=11840, 1.93±0.25 Å)	1.00/100th percentile (N=13349, 1.95±0.25 Å)
Protein Data Bank entry	5Z73	5Z74

472

473 [a] $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an
474 observation and $\langle I(hkl)\rangle$ is the mean value for its unique reflection. Summations are over all
475 reflections.

476 [b] $R_{work} = \sum_h|F_o(h)-F_c(h)|/\sum_h F_o(h)$, where F_o and F_c are the observed and calculated structure
477 factor amplitudes, respectively.

478 [c] R_{free} was calculated with 5% of the data excluded from the refinement.

479 [d] RMSD, root mean square deviation from ideal values.

480 [e] The categories were defined by Molprobity.

481 [f] The number of serious steric overlaps (> 0.4 Å) per 1000 atoms.

482 [g] MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a
483 single score, normalized to be on the same scale as X-ray resolution.

484 [h] The values in parentheses refer to statistics in the highest bin.

485 TABLE 2. Kinetic constants of wild-type and mutant InvA and InvB towards sucrose

InvA				InvB			
Enzyme	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($10^2\ M^{-1}s^{-1}$)	Enzyme	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($10^2\ M^{-1}s^{-1}$)
WT	20.8 ± 1.7	73.7 ± 1.9	35.4	WT	8.9 ± 0.5	163.8 ± 2.1	184.0
R429A	89.3 ± 5.4	29.1 ± 0.7	3.3	R430A	ND[a]	ND	ND
R429N	27.1 ± 2.3	18.9 ± 0.4	7.0	R430N	18.1 ± 1.6	5.3+0.1	2.9

486 [a] ND, no detectable activity.

487

488

489

490 **FIGURE LEGENDS**

491

492 **Fig. 1.** Overall structure of InvB. (a) Dimeric structure. Helices are shown as cylinders, and
493 sucrose molecules in the active site are shown as wheat sticks. The subunits A and B are colored
494 in cyan and hotpink, respectively. The disordered regions are shown as dotted lines. (b) Molecular
495 mass determination of InvB by analytical gel filtration chromatography. (c) A close view of InvB
496 monomer. Subunit A in InvB-Suc is used as an example. The core $(\alpha/\alpha)_6$-barrel and the insertion
497 part are colored in cyan and orange, respectively. The sucrose molecule is shown in sticks. (d)
498 Superposition of InvB monomer against InvA (PDB entry 5GOP) monomer, which is shown in
499 gray. The different loop close to the active site is highlighted in red rectangle.

500

501 **Fig. 2.** The substrate binding pocket of InvB. The involved residues are shown as cyan sticks, the
502 sucrose molecule is shown as wheat sticks, and the water molecules are shown as red spheres.
503 The polar interactions are indicated by dashed lines. The $|F_o|-|F_c|$ omit electron density map of the
504 ligand contoured at 3.0 σ is shown as blue mesh.

505

506 **Fig. 3.** The difference of the loop$_{\beta2-\beta3}$ close to the substrate binding pocket between in InvA and
507 InvB. (a) Stereo representation of loop$_{\beta2-\beta3}$. Subunit A in InvB-Suc structure and subunit A in
508 InvA-Suc structure (PDB entry 5GOP) are shown as cartoon in cyan and red, respectively. The
509 disordered region is shown as dotted lines, with the flanking residues labeled. The key residues
510 are shown as sticks. The polar interactions are indicated by dashed lines. (b) Sequence alignment
511 of the loop$_{\beta2-\beta3}$ in A/N-Invs. Secondary structure elements of InvB are shown at the top of the
512 alignment. The substrate binding residues and the residues probably involved in determining the
513 loop$_{\beta2-\beta3}$ structure are depicted by red triangles and red circles, respectively. The sequences of
514 A/N-Invs are from the following species: *Anabaena* sp. PCC 7120 (InvA, WP_010995690.1;
515 InvB, CAC85155.1), *Arabidopsis thaliana* (At-A/N-InvA, NP_176049.1; At-A/N-InvB,
516 NP_195212.1), *Cyanothece* sp. PCC 7822 (Cy-A/N-Inv, WP_013325329.1), *Daucus carota*
517 (Dc-N-Inv, CAA76145.1), *Fischerella* sp. JSC-11 (Fi-A/N-Inv, ZP_08987807.1), *Oryza sativa*

518 subsp. Japonica (Os-A/N-Inv1, NP_001049936), and *Triticum aestivum* (Ta-A-Inv, CAL26914.1).

519

520 **Fig. 4.** The optimum pH of InvA (red) and InvB (blue). The invertase activity was determined
521 with 100 mM sucrose in a pH range from 5.5 to 9.5. The relative activity of InvB at pH 7.7 was
522 set to 100%. The error bars denote standard deviations of the mean calculated from three
523 independent experiments.

524

525 **Fig. 5.** Conformational comparison of Arg430 of InvB and Arg429 of InvA. (a) In the structure of
526 InvB-Suc, Arg430 directly interacts with the catalytic residue Glu415. The involved residues are
527 shown as lines, the sucrose molecule is shown as sticks, and the water molecule is shown as
528 sphere. The polar interactions are indicated by dashed lines. (b) In the apo-from InvB structure,
529 Arg430 adopts an orientation similar to that in InvB-Suc, despite the substrate binding pocket is
530 occupied by a glycerol molecule. In (c) subunit A and (d) subunit C of InvA-Suc structure,
531 Arg429 interacts with Glu414 via a water molecule. The residues in InvA-Suc structure are
532 labeled with a prime.

533

534 **Fig. 6.** Phylogenetic analysis of A/N-Invs and conservation analysis of Arg430. (a) The
535 neighbor-joining method was adopted to construct the rooted phylogenetic tree. The maximum
536 likelihood method produced identical topology. The α-1,6 glucosidase domain of glycogen
537 debranching enzymye of *Giardia lamblia* ATCC 50803 (Gl-GDE, XP_001705733.1) was used as
538 the outgroup. The five major groups are shaded in different colors. Bootstrap values higher than
539 70% are shown at the nodes. The scale bar represents 0.05 changes per amino acid. (b) The site
540 corresponding to Arg430 of InvB is depicted by a blue triangle, with conserved arginine and
541 asparagine residues shaded in red and yellow, respectively. The grouping of A/N-Invs sequences
542 is consistent with the phylogenetic tree in (a). The sequences of A/N-Invs are from the following
543 species: *Anabaena* sp. PCC 7120 (InvA, WP_010995690.1; InvB, CAC85155.1),
544 *Acidithiobacillus thiooxidans* ATCC 19377 （Act-1, WP_010639088.1）, *Acidihalobacter*

545 *ferrooxidans* (Af-1, WP_076837606.1), *Acidihalobacter prosperus* (Ap-1, WP_065089328.1),
546 *Arabidopsis thaliana* (At-1, NP_176049.1; At-2, NP_195212.1; At-3, NP_187302.2; At-4,
547 NP_564177.1; At-5, NP_197643.1; At-6, NP_177345.1; At-7, NP_174791.2; At-8,
548 NP_001326532.1; At-9, NP_567347.1), *Cylindrospermopsis raciborskii* CS-505 （Cr-1,
549 ZP_06306902.1）, *Chroococcidiopsis thermalis* (Ct-1, WP_015156189.1), *Cyanothece* sp. PCC
550 7822 (Cy-1, WP_013325329.1), *Daucus carota* (Dc-1, CAA76145.1), *Ectothiorhodospira* sp.
551 PHS-1 (Ec-1, WP_083838783.1), *Ectothiorhodospira haloalkaliphila* (Eh-1, WP_026623645.1),
552 *Fischerella* sp. JSC-11 (Fi-1, ZP_08987807.1), *Halothiobacillus* sp. LS2 (Ha-1,
553 WP_066101158.1), *Halothiobacillus neapolitanus* c2 (Hn-1, WP_012823125.1), *Magnetofaba*
554 *australis* IT-1 (Ma-1, WP_085442439.1), *Myxosarcina* sp. GI1 (My-1, WP_036482288.1; My-2,
555 WP_036476871.1), *Nitrosomonas cryotolerans* (Nc-1, SFQ17342.1), *Nostoc punctiforme* ATCC
556 29133 (Np-1, CAD37134.1; Np-2, CAD37133.1), *Nodularia spumigena* CCY9414 （Ns-1,
557 ZP_01631199.1）, *Oryza sativa* subsp. Japonica (Os-1, NP_001049936; Os-2, XP_015625957.1;
558 Os-3, XP_015627031.1), *Pleurocapsa* sp. PCC 7319 (Pl-1, WP_019507642.1), *Prochlorococcus*
559 *marinus* (Pm-1, WP_011131014.1; Pm-2, WP_036892687.1), *Physcomitrella patens* (Pp-1,
560 XP_001754878.1; Pp-2, XP_001758344.1) *Stanieria cyanosphaera* (Sc-1, WP_041619725.1;
561 Sc-2, WP_015195523.1), *Selaginella moellendorffii* (Sm-1, XP_002978791.1; Sm-2,
562 XP_002968256.1), *Spirogyra pratensis* (Sp-1, GBSM01000698.1; Sp-2, GBSM01023190.1;
563 obtained from the Transcriptome Shotgun Assembly database of NCBI and referred to reference
564 [13]), *Synechococcus* sp. WH 8020 (Sy-8020-1, WP_048348470.1), *Synechococcus* sp.
565 TMED187 (Sy-TMED187-2, OUW48797.1), *Synechocystis* sp. (PCC 6803, Syi-6803-1,
566 CAD33848.1; PCC 7509, Syi-7509-2, WP_009632367.1; PCC 6714, Syi-6714-3,
567 WP_028948477.1), *Triticum aestivum* (Ta-1, CAL26914.1), *Thiohalorhabdus denitrificans* (Td-1,
568 WP_054965954.1), *Thiohalospira halophila* DSM 15071 (Th-1, WP_093427069.1),
569 *Thiohalomonas denitrificans* (Tmd-1, WP_092994335.1), *Thioalkalivibrio nitratireducens* (Tn-1,
570 WP_043739311.1), *Thioalkalivibrio paradoxus* (Tp-1, WP_006747994), *Thioalkalivibrio*
571 *sulfidophilus* (Ts-1, WP_012637287.1; Ts-2, WP_026289890.1), and *Zea mays* (Zm-1,

572 XP_008661659.1; Zm-2, XP_008655058.1; Zm-3, NP_001130493.1).

573

574 **Fig. S1.** Different oligomerization patterns of InvA and InvB. (a) Superposition of InvA and InvB
575 dimer. Helices are shown as cylinders, and sucrose molecules in the active site are shown as
576 sticks. Two subunits of InvB are colored in cyan and hotpink, respectively, except for the two
577 insertion parts colored in orange, while both subunits of InvA are colored in gray. The rotation of
578 subunit B of InvB relative to that of InvA is shown by the rotated angle of helix α14. Two views
579 of the dimers are shown. (b) Sequence analysis of the inter-dimer interface of InvA. The segments
580 at the inter-dimer interface of InvA are depicted by red lines. Green triangles indicate the residues
581 involved in polar interactions in InvA but different in InvB. The sequences of A/N-Invs are the
582 same as those in Fig. 3b. (c) The residues Arg251 and Glu286 at the inter-dimer interface of InvA.
583 Subunit A and E in the full-length hexameric InvA structure (PDB entry 5GOR) are colored in
584 blue and green, respectively, except for the insertion part of subunit A colored in orange. The
585 polar interactions are indicated by dashed lines.

586

587 **Fig. S2.** The active site comparison of InvA (yellow) and InvB (cyan). The involved residues and
588 the sucrose molecules are shown as sticks. The water molecules are shown as spheres.

589

590 **Fig. S3.** Multiple sequence alignment of A/N-Invs. The catalytic residues and substrate binding
591 residues are depicted by red stars and red triangles, respectively, and the site corresponding to
592 Arg430 of InvB is depicted by a blue triangle. The sequences are the same as in phylogenetic
593 analysis.

594