High-resolution cryo-EM structure of urease from the pathogen *Yersinia enterocolitica*

Ricardo D. Righetto[1*], Leonie Anton[2*], Ricardo Adaixo[1*], Roman P. Jakob[2], Jasenko Zivanov[1], Mohamed-Ali Mahi[2,3], Philippe Ringler[1], Torsten Schwede[2,3], Timm Maier[2†] and Henning Stahlberg[1†]

[1] Center for Cellular Imaging and NanoAnalytics, Biozentrum, University of Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland.

[2] Biozentrum, University of Basel, Klingelbergstrasse 50/70, CH-4056 Basel, Switzerland.

[3] SIB Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Klingelbergstrasse 50/70, CH-4056 Basel, Switzerland.

[*] *these authors contributed equally to this work.*

[†] Corresponding authors:

Timm Maier
Biozentrum, University of Basel
Klingelbergstrasse 50/70
CH-4056 Basel, Switzerland
Phone: +41 61 207 21 76
E-mail: timm.maier@unibas.ch

Henning Stahlberg
Center for Cellular Imaging and NanoAnalytics (C-CINA)
Biozentrum, University of Basel
WRO-1058, Mattenstrasse 26
CH-4058 Basel, Switzerland
Phone: +41 61 387 32 62
E-mail: henning.stahlberg@unibas.ch

Running title: Cryo-EM structure of *Y. enterocolitica* urease

Keywords: urease; enzyme; metalloprotein; radiation damage; cryo-electron microscopy;

Abstract

Urease converts urea into ammonia and carbon dioxide and makes urea available as a nitrogen source for all forms of life except animals. In human bacterial pathogens, ureases also aid in the invasion of acidic environments such as the stomach by raising the surrounding pH. Here, we report the structure of urease from the pathogen *Yersinia enterocolitica* at better than 2 Å resolution from electron cryo-microscopy. *Y. enterocolitica* urease is a dodecameric assembly of a trimer of three protein chains, ureA, ureB and ureC. The high data quality enables detailed visualization of the urease bimetal active site and of the impact of radiation damage. Our data are of sufficient quality to support drug development efforts.

42 ## Introduction

43 Ureases are nickel-metalloenzymes produced in plants, fungi and bacteria, but not in
44 animals. They facilitate nitrogen fixation by metabolizing urea, but in some pathogenic
45 bacteria they serve a dual function and consume ammonia to promote survival in acidic
46 environments. This function is vital during host infection where the bacteria have to survive
47 the low pH of the stomach (Gripenberg-Lerche *et al.*, 2000), and of phagosomes in host
48 cells (Young *et al.*, 1996; Maroney & Ciurli, 2014). The relevance of ureases in early stages
49 of infection renders them attractive targets for novel anti-microbials.

50 Ureases catalyze the breakdown of urea into ammonia and carbamate at a rate 10^{14} to 10^{15}
51 times faster than the non-catalyzed reaction **(Fig. 1a)**, and are arguably the most efficient
52 hydrolases (Maroney & Ciurli, 2014). In a second non-catalyzed step, carbamate is
53 spontaneously hydrolyzed to yield another molecule of ammonia as well as one molecule of
54 bicarbonate (Mazzei *et al.*, 2017). Jack bean urease was the first enzyme to be crystallized,
55 offering evidence that enzymes are proteins (Sumner, 1926) and was the first
56 metalloenzyme to be shown to use nickel in its active site (Dixon *et al.*, 1975). All ureases
57 characterized to date share the architecture of their active site (Kappaun *et al.*, 2018). Two
58 Ni^{2+} ions in the active site are coordinated by a carbamylated lysine, four histidines and one
59 aspartate. Urea first interacts with Ni(1) through its carbonyl oxygen, making urea more
60 available for nucleophilic attacks. The amino nitrogen then binds to Ni(2) and a proton is
61 transferred from a water molecule to the amine in a nucleophilic attack on the carbonyl
62 carbon of urea. The active site is closed off during the reaction by a conformationally variable
63 helix-turn-helix motif, referred to as the mobile flap. Despite their common catalytic
64 mechanism, ureases display different chain topologies and higher order oligomeric
65 assemblies **(Fig. 1b,c)**. Jack bean urease assembles into an oligomer of a single type of
66 polypeptide chain with D3 symmetry ($[[\alpha]_3]_2$ D3) **(Fig. 1b,c)** (Maroney & Ciurli, 2014). The

urease of the human pathogen *Helicobacter pylori* is composed of two types of polypeptide chains (ureA, ureB) and assembles into a dodecamer (tetramer-of-trimers) with tetrahedral symmetry ($[[\alpha\beta]_3]_4$ T) **(Fig. 1b,c)**. The ability of ureases to raise the pH of their environment benefits *H. pylori* in colonizing the stomach and downstream gut, causing gastric ulcers (Kappaun *et al.*, 2018). The urease of the opportunistic pathogen *Klebsiella aerogenes* (Podschun & Ullmann, 1998) assembles into a heterotrimer of three proteins ureA, ureB and ureC, which in turn oligomerizes into a trimer ($[\alpha\beta\gamma]_3$ C3) (Jabri *et al.*, 1995) **(Fig 1b,c)**. This oligomeric assembly is common also to most other structurally characterized bacterial ureases (Maroney & Ciurli, 2014).

Yersinia enterocolitica is the causative agent of yersiniosis, a gastrointestinal infection, reactive arthritis and erythema nodosum. The infection spreads to humans through consumption of contaminated food with pigs being one of the largest reservoirs of *Y. enterocolitica*. The symptoms of yersiniosis are fever, abdominal pain, diarrhea and/or vomiting and is one of the most reported enteritis in some countries, although outbreaks are rare (Drummond *et al.*, 2012). *Y. enterocolitica* is a facultative intracellular bacterium and can survive in very different environments. In the presence of urea *Y. enterocolitica* can tolerate extremely acidic conditions (Young *et al.*, 1996). *Y. enterocolitica* urease comprises three polypeptide chains (ureA, ureB and ureC), an architecture similar to that of *K. aerogenes* urease **(Fig. 1b)**.

Here we present the structure of *Y. enterocolitica* urease at an overall resolution of 1.98 Å, which was achieved using recent advances in cryo-EM data collection and processing. The structure shows that *Y. enterocolitica* urease assembles into a dodecameric hollow sphere with $[[\alpha\beta\gamma]_3]_4$ T oligomeric assembly structure **(Fig. 2)**. A tightly embedded kinked loop is interacting with neighboring domains and is potentially responsible for the assembly of the oligomer. The data allows model building of the active site carbamylated lysine, visualization of radiation damage to the nickel-metal center, and hydration networks throughout the

93 protein. Our work highlights the potential of cryo-EM for structure-based drug discovery,
94 specifically for large enzyme assemblies, which have been historically difficult to analyze by
95 X-ray crystallography.

96

97 Results & Discussion

98 Structure determination of *Y. enterocolitica* urease by cryo-EM

99 We have used single particle cryo-EM to determine the structure of the fully assembled *Y.*
100 *enterocolitica* urease. We acquired 4,494 movies of urease particles using a Titan Krios
101 transmission electron microscope (TEM) equipped with a K2 direct electron detector and an
102 energy filter (see **Methods** for details). Approximately half of the movies (2,243) were
103 acquired by illuminating three locations ("shots") per grid hole using beam-image shift in
104 order to speed up the data collection (Cheng *et al.*, 2018), whereas the remaining movies
105 were recorded without this feature i.e. just a single shot at the center of the hole. This allowed
106 us to measure and assess the extent of beam tilt and other optical aberrations, as well as
107 the behavior of sample drift between each condition and beam-image shift position. Typical
108 micrographs from the imaged grids are shown in **Supp. Fig. 1a** and a summary of data
109 collection information is given in **Supp. Tab. 1**.
110 Each dataset was processed separately for 3D reconstruction following the strategy
111 depicted in **Supp. Fig. 2**. The first obtained 3D map, at an overall resolution of 2.6 Å,
112 revealed that this urease assembly is a dodecamer of tetrahedral (T) symmetry with a
113 diameter of approximately 170 Å. The separate processing of each dataset yielded refined
114 3D maps at nominal resolutions of 2.10 Å and 2.20 Å for the multi-shot and single-shot
115 cases, respectively (see **Methods**).
116 For comparison, we also processed the merged set of particles from both datasets
117 altogether. We observed on the 2D class averages a preferential orientation for the three-
118 fold symmetric view of urease and also the presence of isolated monomers and broken

119 assemblies **(Supp. Fig. 1b)**. The presence of such incomplete assemblies was further
120 confirmed by performing 3D classification without imposing symmetry, as shown in **Supp.**
121 **Fig. 1c**. The 3D class corresponding to the complete dodecameric assembly of urease
122 contained 119,020 particles, of which 69,512 (58.4%) came from the multi-shot and 49,518
123 (41.6%) from the single-shot dataset. With respect to the number of particles picked from
124 each dataset, 64.7% of the particles from the multi-shot and 56.8% from the single-shot
125 datasets were retained at this stage and throughout the final reconstruction. While coma-
126 free alignment was performed and active beam-tilt compensation in SerialEM (Schorb *et al.*,
127 2019) was used on our data collections, after performing beam tilt refinement in RELION-3
128 (Zivanov *et al.*, 2018) we observed that the single-shot case has a residual beam-tilt higher
129 than the smallest residual observed in the multi-shot case **(Supp. Tab. 2)**. These two values
130 are however very close to zero and are possibly within the error margin of the *post hoc* beam
131 tilt refinement procedure.
132 The reduced need to move the specimen stage in beam-image shift mode not only speeds
133 up data collection but also minimizes stage drift. The second and third shots from the multi-
134 shot dataset have comparatively less drift than both the first multi-shot and the single shot,
135 as suggested by the parameter values obtained from the Bayesian polishing training
136 (Zivanov *et al.*, 2019*a*) on each beam-tilt class separately **(Supp. Tab. 3)**. As all the three
137 multi-shots are taken in nearby areas within the same foil hole, this observation is consistent
138 with the annealing of the vitreous ice layer and its carbon support after pre-irradiating the
139 specimen as reported previously (Brilot *et al.*, 2012).
140 At this point, the nominal resolution of the map after 3D refinement was 2.05 Å. Finally,
141 correcting for residual higher-order aberrations in CTF refinement (Zivanov *et al.*, 2019*b*)
142 **(Supp. Fig. 3)** yielded a map at a global resolution of 1.98 Å **(Supp. Fig. 4a)**. Local
143 resolution estimation reveals that the core of the map is indeed at this resolution level or
144 better **(Fig. 2a and Supp. Fig. 4b)**, and the local resolution-filtered map was then used for

145 model building as explained in the next section. Despite the twelve-fold symmetry of the
146 urease assembly, a limiting factor in the resolution of the map is the strong presence of
147 preferential orientation, as confirmed by the plot of the final orientation assignments (**Supp.**
148 **Fig. 4c**). An overview of the cryo-EM map and its main features are depicted in **Supp. Mov.**
149 **1.**
150

151 ***Y. enterocolitica* urease assembles as a tetramer of trimers**

152 Model building was initiated from available crystallographic models with subsequent fitting
153 and refinement against the cryo-EM map. The model was built and refined for one
154 asymmetric unit containing one copy of the ureA, ureB and ureC protein each. The model
155 was then expanded using NCS (see **Methods**). The complete model covering the whole
156 oligomeric assembly contains 9,552 residues, 3,672 waters and 24 nickel ions (two per
157 active site, twelve active sites) **(Table 1)**. The quality of the model was assessed with the
158 cryo-EM validation tools in the PHENIX package (Afonine, Klaholz *et al.*, 2018). The map
159 allowed for the building of all residues of ureA (1-100), and residues 31-162 of ureB and 2-
160 327/335-572 of ureC **(Supp. Fig. 5)**. The hetero-trimer formed by the three protein chains
161 (ureA, ureB, ureC) **(Fig. 1b)** oligomerizes into a homo-trimer. The homo-trimer is arranged
162 in a tetramer-of-trimers making the full complex a dodecamer of the hetero-trimer **(Fig. 2b)**.
163 There are four large oval shaped holes between the trimers (64 Å long, 12 Å wide, high
164 electrostatic potential) and four smaller holes at the center of the trimer with a diameter of 6
165 Å (low electrostatic potential), as shown in **Fig. 2b**, and the center of the enzyme assembly
166 is hollow **(Fig. 2).** The holes provide ample opportunity for diffusion of the uncharged
167 substrate and product, and the hollow inside potentially leads to local increase of reaction
168 product. The assembly has the same symmetry as the urease homologue in *H. pylori*, which
169 was postulated to increase stability and/or resistance to acidic environments (Ha *et al.*,
170 2001).

171

172 For analysis of the protein sequences, the ConSurf Server (Berezin *et al.*, 2004; Ashkenazy
173 *et al.*, 2016) was used with the sample list of homologs option to get a diverse set of 150
174 sequences. The protein chains of *Y. enterocolitica* urease are highly conserved across
175 different organisms. The ureA chain is split after a LVTXXXP motif and is 99-100 amino
176 acids long in most cases, with a sequence identity of 55.7%. The ureB chain of *Y.*
177 *enterocolitica* has between 20 and 30 N-terminal amino acids more compared to the other
178 sequences (except *Kaistia sp.* SCN 65-12), which share an identity of 51.5%. This N-
179 terminal extension is located on the outside of the holoenzyme where ureA and ureB chain
180 split occurred and are too disordered to be modeled in the structure **(Supp. Fig. 5)**. They
181 could be remnants from when ureA and ureB were connected. The charges and properties
182 of this stretch of amino acids vary and if they still serve a function remains unclear. The last
183 20 amino acids of the C-terminus of ureB are only represented in half of the compared
184 sequences and accurate sequence conservation could not be determined in this part. This
185 stretch contains a loop and a C-terminal helix **(Supp. Fig. 5)**. The ureC protein of the
186 compared sequences has a shared sequence identity of 60.3%. All amino acids involved in
187 catalysis are highly conserved **(Supp. Fig. 5)**. The ureA and ureB chains show lower
188 conservation compared to ureC. They are not involved in catalysis but in scaffolding, so the
189 differences could stem from their role in different types of oligomeric assembly **(Fig. 3)**.
190 To investigate this aspect further, we compared the presented structure to the ureases of *H.*
191 *pylori*, *S. pasteurii* and *K. aerogenes*. Sequence identity scores among these ureases are
192 provided in **Supp. Tab. 4**. The *H. pylori* urease is made up of two protein chains ureA (that
193 contains the equivalent of ureA and ureB in *Y. enterocolitica*) and ureB (that is the equivalent
194 of ureC in *Y. enterocolitica*) **(Fig. 1c and Fig. 3c)**. It assembles into a T-symmetric oligomer
195 like in *Y. enterocolitica* and the crystal structure was solved to 3 Å. *S. pasteurii* and *K.*
196 *aerogenes* ureases both assemble into a trimer from the hetero-trimeric unit **(Fig. 1c and**

197 **Fig. 3b)**. *S. pasteurii* was solved in the presence of the inhibitor N-(n-Butyl)thiophosphoric
198 Triamid (NBPT) to 1.28 Å (Mazzei *et al.*, 2017). The crystal structure of *Klebsiella aerogenes*
199 was solved to 1.9 Å as in absence of substrate or inhibitors (Jabri *et al.*, 1995).
200 There are two main regions with high root mean square deviations (RMSDs) when
201 comparing these three ureases to the *Y. enterocolitica* model **(Supp. Fig. 6 and Supp. Tab.**
202 **5)**. The first region with high deviation is the mobile flap, which opens and closes over the
203 active site (residues 312 to 355 of ureC). The residues of its connecting loop could not be
204 built with confidence in the cryo-EM model (residues 326 to 333 of ureC). The other region
205 with large differences is on the edges of ureA and ureB where the interactions with the next
206 protomer occur. The *H. pylori* assembly contains an additional C-terminal loop (residues
207 224-238 of ureA) after the top alpha helix (residues 206-223 of ureA). This helix (central
208 helix) forms the three-fold axis of three neighboring trimers and the loop binds in a head-to-
209 tail fashion to the next trimer forming the tetramer **(Fig. 3a,c)** (Ha *et al.*, 2001). The core of
210 the assembly is identical in its structure. For whole-chain superposition scores and RMSD
211 values between the compared models please see **Supp. Tab. 5**.
212 In the dodecameric assembly seven different interfaces are formed between the hetero-
213 trimers **(Fig. 4a,b and Supp. Fig. 7)**. Intra-trimer interactions occur between the three basic
214 hetero-trimers in one assembled trimer, forming a three-fold symmetry axis **(Fig. 4a)**. The
215 interactions between these trimers to form the tetramer then make up a different three-fold
216 symmetry axis **(Fig. 4b)**. The three largest interfaces (interfaces 1-3) are formed intra-
217 trimeric between ureC of one hetero-trimer and ureC, ureA and ureB of the next trimer **(Fig.**
218 **4a)**. The three ureA proteins make up the intra-trimer-core (first three-fold axis) with interface
219 4 **(Fig. 4c and Supp. Fig. 7)**. Comparison of the interface areas formed in the trimer
220 assembly shows no substantial differences between the four organisms **(Supp. Fig. 7)**.
221 Inter-trimer interfaces (interfaces 5, 6, 7) formed in the dodecameric *Y. enterocolitica* and *H.*
222 *pylori* ureases have similar areas **(Fig. 4b and Supp. Fig. 7)**. Part of the interactions occur

223 between ureB and ureC forming interfaces with each other (interface 4, 6). The other
224 interaction is between the three ureB proteins and forms interface 7 and the inter-trimer-
225 core (second three-fold axis) with their central helices **(Fig. 4b)**. *Y. enterocolitica* does not
226 have the same oligomerization loop after the central helix proposed for *H. pylori*. However,
227 there is a short loop before the central helix, which is extended in *Y. enterocolitica*. It binds
228 into a pocket of ureC of the neighboring trimer in interface 6 **(Fig. 4e)**. These types of loops
229 or extensions are missing from *S. pasteurii* and *K. aerogenes* ureB proteins. *S. pasteurii* has
230 the central helix, but there is no extended loop before or after it **(Supp. Fig. 6b)**. *K.*
231 *aerogenes* urease does not have a helix nor a loop in this region **(Fig. 3b)**. This suggests
232 that the presence of oligomerization loops in ureB is crucial for determining the oligomeric
233 state of the enzyme.
234 The dodecameric holoenzyme structure of ureases might aid in stabilizing the protein at
235 acidic pH, and in combination with 12 active sites producing ammonia enables the formation
236 of a pH-neutralizing microenvironment around the assembly (Ha *et al.*, 2001). This ensures
237 the continued function of the enzyme and makes this type of oligomeric assembly essential
238 to survival of *Y. enterocolitica* in the host.
239

240 **The empty active site is filled with water**

241 At the global resolution of 1.98 Å, detailed structural features can be observed. All
242 throughout the highly resolved areas of the protein, salt bridges, backbone and side chain
243 hydration and alternative side chain conformations can be visualized **(Supp. Fig. 8a-c)**.
244 Furthermore, the high resolution allows for a detailed description of the nickel-metallo-center
245 and the active site. The active site is located on the ureC protein at the edge of the hetero-
246 trimer and is wedged in between the ureA and ureB proteins of the next hetero-trimer in the
247 homo-trimeric assembly **(Fig. 5a)**.
248 The catalysis of ammonia and carbamate from urea occurs in two steps **(Fig. 1a)**. Urea first

249 interacts with the nickel ions through its carbonyl oxygen and amino nitrogens. The active
250 site contains two Ni^{2+} ions which are coordinated by six different amino acids **(Fig. 5b)**. Both
251 Ni^{2+} ions are coordinated by the carbamylated LYS222*. Ni(1) is additionally coordinated by
252 HIS224, HIS251 and HIS277 and Ni(2) by HIS139, HIS141 and ASP365. Close to the active
253 site is a methionine (MET369), which can be modelled in different alternative conformations.
254 One conformation could potentially reach the active site. There is no described function for
255 this amino acid **(Fig. 5c and Supp. Fig. 9)**.
256 The active site is protected by a helix-turn-helix motif, called the mobile-flap. Its function is
257 to coordinate the access of substrate to the catalytic site and the release of the product from
258 it (Maroney & Ciurli, 2014). After closing of the mobile flap, a proton is transferred from a
259 water molecule to the other amine in a nucleophilic attack on the carbonyl carbon of urea.
260 A conserved histidine on the mobile flap (HIS325) is essential for catalysis by possibly acting
261 as a general acid and aiding in deprotonation **(Fig. 5a)** (Maroney & Ciurli, 2014; Kappaun *et*
262 *al.*, 2018). By closing of the mobile flap the HIS325 moves closer to the active site, also
263 stabilizing urea in the active site pocket (Maroney & Ciurli, 2014). Flap opening then releases
264 ammonia and carbamate, where the latter spontaneously hydrolyses into another molecule
265 of ammonia and bicarbonate. The mobile flap of the cryo-EM structure presented here is in
266 an open position, which is explained by the absence of substrate or inhibitors in the sample
267 **(Fig. 5a)**. Coordinated water molecules can be seen in the empty pocket of the active site,
268 which do not only form hydrogen bonds with side chains or the protein backbone, but also
269 with each other constituting a hydration network **(Supp. Fig. 8d)**.
270 The resolution in the active site is sufficient for complete atomic description of the
271 coordinated Ni^{2+} ions, including the carbamylated lysine. The protonation states of the active
272 site residues are also represented in the map **(Fig. 5b)**. One of the hydroxide molecules in
273 the active site is essential as it performs the nucleophilic attack on urea while other
274 molecules are displaced by urea and the closing of the mobile flap (Kappaun *et al.*, 2018).

275 Comparison to the crystal structure of *K. aerogenes* of similar nominal resolution (1.9 Å)
276 shows differences in the visualization of these features. This crystal structure was solved in
277 absence of inhibitors or substrate such that the active site is also empty and the mobile flap
278 in an open conformation **(Fig. 5a)**. The details of the map are better resolved around the
279 Ni^{2+} ions in the cryo-EM map than in the crystallographic data. The protonation of the
280 histidines is clearly visible in the cryo-EM density **(Fig. 5b)**. The positions of the side chains
281 and the Ni^{2+} ions in the active site are very similar to the *Y. enterocolitica* urease structure
282 with a RMSD of 0.270 Å **(Supp. Tab. 5)**. The *S. pasteurii* crystal structure was solved in
283 presence of the inhibitor NBPT which displaces the essential water molecules needed for
284 the reaction from the active site. The closing of the mobile flap displaces the rest of the
285 waters and brings the catalytic HIS323 closer to the active site. The tight packing of side
286 chains prevents urea from entering the active site, efficiently blocking it **(Fig. 5b,e)**. The
287 active site residues and Ni^{2+} ions have a RMSD of 0.293 Å between *S. pasteurii* and *Y.*
288 *enterocolitica*.
289

290 **Nickel atoms come closer together**

291 The distance between the Ni^{2+} ions is 3.7 Å in X-ray structures of *K. aerogenes* and *S.*
292 *pasteurii*, but only 3.2 Å in the *Y. enterocolitica* cryo-EM model **(Fig. 5b,d)**. Short distances
293 of 3.1-3.3 Å were described for *S. pasteurii* and *K. aerogenes* at high resolutions for
294 structures in presence of β-Mercaptoethanol (β-ME) (Benini *et al.*, 1998). Knowing that
295 metallic cores are particularly sensitive to radiation (Yano *et al.*, 2005), we tried to determine
296 the extent to which radiation damage can explain the shorter distance between the Ni^{2+} ions.
297 For this purpose, we generated per-frame reconstructions for the first 25 frames of our data
298 collection and refined the model on each of them (see **Methods**) and measured the
299 distances between the residues involved in ion coordination, shown in **Fig. 6**. At the
300 beginning of the exposure, in which the frames contribute more to the full reconstruction due

301 to dose-weighting (Grant & Grigorieff, 2015; Zivanov *et al.*, 2019*a*), there is a trend of the
302 ions coming closer together (**Fig. 6a-i**). While we cannot determine exactly how this arises
303 from radiation damage, it is likely a result of several interactions in the active site changing
304 simultaneously along the exposure. For example, both Ni(1) and Ni(2) tend to come closer
305 to the carbamylated LYS222 (**Fig. 6a-iv,v**) as ASP365 vanishes (**Fig. 6a-vi**), which can be
306 seen in the **Supp. Mov. 2**. Aspartic acid is known to have its side chain damaged very early
307 on (Hattne *et al.*, 2018). The dynamic interplay between residues along the exposure (**Fig.**
308 **6b**) is likely due to the different rates at which specific types of bonds and residues are
309 damaged (Fromm *et al.*, 2015): first negatively charged residues, then positively charged
310 ones followed by aromatic side chains, as also observed in **Supp. Mov. 2**. However, the
311 later part of the exposure must be interpreted with caution, as atomic coordinates become
312 less reliable, which is verified by the overall increase in B-factors in **Fig. 6c** and the error
313 bars in **Fig. 6a**.

314 Conclusion

315 Large urease assemblies have been historically difficult to study by X-ray crystallography
316 (Ha *et al.*, 2001). We have determined the structure of a dodecameric urease assembly, a
317 metalloenzyme from the pathogen *Y. enterocolitica* at an overall resolution of 1.98 Å using
318 cryo-EM. The collection of datasets with and without beam-image shift demonstrates the
319 advantages of using this feature of modern TEMs and invites further investigations on the
320 behavior of optical aberrations and specimen drift.

321 Our results demonstrate the feasibility of cryo-EM as a technique for obtaining structures of
322 clinically relevant enzymes with sufficient quality for *de novo* model building and drug design.
323 The cryo-EM map has allowed a detailed description of the active site and the oligomeric
324 assembly. More specifically, we could observe the putative oligomerization loop that enables
325 the dodecameric assembly, which was hypothesized to be responsible for the enhanced

survival of *Y. enterocolitica* in highly acidic environments (Young *et al.*, 1996).

Furthermore, in comparison to the *K. aerogenes* structure, which is at approximately the same nominal resolution, the cryo-EM map offers an improved representation of protons and Ni^{2+} ions. A possible explanation is that the error in the phases derived in the X-ray structure determination grows faster towards the limit of observed diffraction. Another aspect to be considered is that X-rays and electrons probe different properties of matter, respectively the electron density and the integrated Coulomb potential. Our results prompt a more detailed investigation of these effects and how they affect the representation of features at high resolution.

Finally, we noticed that radiation damage can partially explain the shorter distance observed between the nickel atoms in the active site. Given that ions and charged residues are damaged very early on in the exposure (Yano *et al.*, 2005; Hattne *et al.*, 2018), this effect cannot be neglected in structures derived from cryo-EM reconstructions. Novel direct electron detectors with higher frame rates may allow time-resolved experiments to investigate these effects in more detail.

Data availability

The model has been deposited at the PDB under accession code 6YL3. The map has been deposited at the EMDB under accession code EMD-10835. Raw electron microscopy data is deposited in EMPIAR, accession code EMPIAR-10389.

Acknowledgments

The authors would like to thank L. Kovacik and K. Goldie for assistance in data collection. R.D.R. acknowledges S. Klumpe and A. Nunes-Alves for discussions on tracking atomic coordinates. Cryo-EM data processing calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing center at the University of Basel. R.D.R. and

L.A. acknowledge funding from the Fellowships for Excellence program sponsored by the Werner-Siemens Foundation and the University of Basel. This work was in part supported by the Swiss National Science Foundation (grants 177195 and 185544, NCCR TransCure).

Author Contributions

R.D.R. and R.A. performed the cryo-EM experiments and data analysis. L.A. built and analyzed the atomic model. R.P.J. performed X-ray diffraction experiments. M.A.M. expressed and purified the protein. P.R. prepared and screened EM samples. J.Z. performed the higher-order aberration corrections and analysis. T.S., T.M. and H.S. initiated and supervised the project. R.D.R, L.A., R.A., T.M. and H.S. wrote the manuscript with assistance from all authors.

Methods

Protein expression and purification

The *Y. enterocolitica* urease was purified for cryo-electron microscopy according to the protocol of (Rokita *et al.*, 2000). The strain was precultured overnight at 37°C for 18 hours in a medium containing 37 g/l of brain/heart infusion (Oxoid, CM0225), 50 µg/ml streptomycin sulfate (Applichem, A1852.0100), 35 µg/ml nalidixic acid (Applichem, A1894.0025), 50 µg/ml meso-diaminopimelic acid (Sigma, D1377) and 100 µM nickel(II) chloride hexahydrate (Sigma, N6136). 6 x 600ml of expression cultures were inoculated at OD of 1 at 28°C for 23 hours. Cells were harvested by centrifugation and the cell pellet resuspended in 0.15M NaCl, 50mM Tris pH 8.0. The cell lysate was applied directly to a Sephacryl S-300 HR 26/60 column equilibrated with 150 mM NaCl, 50mM TrisHCl pH 8.0. The active fractions as identified by a phenol-hypochlorite assay (Weatherburn, 1967) were buffer-exchanged to 50mM Tris pH 7.0 within a centrifugal filter unit (Sartorius, Vivaspin MWCO 50kDa) and applied to a Mono Q HR 5/5 column pre-equilibrated with 50mM

374 TrisHCl, pH 7.0. The protein was eluted in 50mM Tris pH 7.0 by a gradient to 1M NaCl,
375 concentrated on a centrifugal filter unit (Sartorius, Vivaspin MWCO 50 kDa) and purified by
376 SEC as before. The purity of the urease sample of the two preparations was verified on a
377 4%/12% SDS-PAGE and by mass spectroscopy.

378 *Sample preparation*

379 Approximately 3 µl of the 0.39 mg/ml urease solution were applied to glow-discharged
380 Quantifoil holey carbon grids. After 3-second blotting, the grids were flash-frozen in liquid
381 ethane, using a FEI Vitrobot IV (Thermo Fisher Scientific) with the environmental chamber
382 set at 90% humidity and 20 °C temperature.

383 *Data acquisition*

384 Cryo-EM data were collected on a FEI Titan Krios (Thermo Fisher Scientific) transmission
385 electron microscope, operated at 300 kV and equipped with a Quantum-LS imaging energy
386 filter (GIF, 20 eV zero loss energy window; Gatan Inc.) and a K2 Summit direct electron
387 detector (Gatan Inc.) operated in dose fractionation mode. Data acquisition was controlled
388 by the SerialEM (Schorb *et al.*, 2019) software, performed in counting mode, with a 42 $e^-/Å^2$
389 total exposure fractioned into 40 frames over 8 seconds. The physical pixel size was 0.639
390 Å at the sample level. The data was pre-processed via the FOCUS package (Biyani *et al.*,
391 2017), including drift-correction and dose-weighting using MotionCor2 (Zheng *et al.*, 2017)
392 (grouping every 5 frames and using 3x3 tiles) and CTF estimation using CTFFIND4 (Rohou
393 & Grigorieff, 2015) (using information between 30 Å and 5 Å from the movie stacks). With
394 these settings, we collected two datasets: one using beam-image shift (Cheng *et al.*, 2018),
395 with three shots per grid hole, comprising 2,243 movies, and a second one taking a single
396 shot per hole, with 2,252 movies. A summary of data collection information is given in the
397 **Supp. Tab. 1**.

398 *Image processing*

399 The two datasets were initially processed separately as shown in the flowchart of **Supp. Fig.**
400 **2**. We excluded all movies whose resolution of CTF fitting was worse than 4 Å according to
401 CTFFIND4, leaving 2,197 movies in the multi-shot dataset or 2,115 in the single-shot dataset
402 for further processing. Using the template-free LoG-picker algorithm (Zivanov *et al.*, 2018)
403 we picked an initial set of 157,699 particle coordinates on the multi-shot dataset. These
404 particles were extracted and subjected to one round of 2D classification with the aim of
405 removing "bad" or false-positive particles. Best results in 2D classification were observed
406 when enabling the RELION option "Ignore CTFs until first peak?". Selecting only the classes
407 displaying views of urease with high resolution features, a new subset containing 60,271
408 particles was obtained. Using this subset, a first 3D map was obtained by the *ab initio*
409 stochastic gradient descent (SGD) algorithm (Zivanov *et al.*, 2018; Punjani *et al.*, 2017) with
410 and without tetrahedral symmetry imposed. The symmetric map was consistent with
411 previously determined structures of ureases (Arnold *et al.*, 2016; Ha *et al.*, 2001). The
412 particles were then subjected to 3D refinement using the map from the *ab initio* procedure
413 as starting reference, resulting in a map at 2.6 Å resolution. We then generated new
414 templates for particle picking by low-pass filtering the unsharpened map from this first 3D
415 refinement to 20 Å and calculating evenly oriented 2D projections from it. These templates
416 were then used for picking with Gautomatch (K. Zhang, http://www.mrc-
417 lmb.cam.ac.uk/kzhang/), detecting 107,399 particle coordinates on the multi-shot dataset or
418 87,204 particle coordinates on the single-shot dataset. Visual inspection of randomly
419 selected micrographs indicated this set of coordinates was better than that previously found
420 by the LoG-picker, in the sense that it contained less false positives and more true particles.
421 The newly extracted particles were then subjected to 2D and 3D classification procedures
422 to get rid of false positive, damaged or broken particles, which yielded cleaner subsets with
423 62,884 (multi-shot) or 51,173 (single-shot) particles. Using the current best map from the

424 multi-shot dataset as a starting reference, we then performed masked 3D refinements on
425 the two datasets separately, interleaved with rounds of CTF refinement and Bayesian
426 particle polishing (Zivanov *et al.*, 2019*a*). More specifically, we refined defocus *per particle*,
427 astigmatism *per micrograph* and beam tilt in CTF refinement. In the multi-shot dataset, each
428 of the three relative "shooting targets" were assigned a different class for separate beam tilt
429 refinement. The parameters for Bayesian particle polishing were trained separately on
430 ~5,000 particles from each dataset at this stage. Each dataset yielded refined maps at 2.10
431 Å (multi-shot) and 2.20 Å (single-shot) resolution. Best results, however, were obtained
432 when merging the particles picked by template-matching on each dataset (194,603 particles
433 in total) and processing them altogether. After 2D classification 141,069 particles remained
434 **(Supp. Fig. 1b)** and after 3D classification there were 119,020 particles **(Supp. Fig. 1c)**.
435 CTF refinement was now performed using four beam tilt classes, with the particles from the
436 single-shot dataset belonging to a new, fourth class **(Supp. Tab. 2)**. Defocus and
437 astigmatism were both refined *per particle* this time resulting in a map at resolution of 2.05
438 Å. We experimented polishing the full merged dataset at once, or each beam tilt class
439 separately, to verify if there were differences in the patterns of particle motion. For training
440 the polishing parameters, this time ~10,000 particles were used in each case. Although we
441 did observe different statistics of particle motion **(Supp. Tab. 3)**, resolution and overall
442 quality of the map did not improve further by doing either merged or separate polishing of
443 the different "shots". Finally, correction of third-order aberrations in RELION-3.1 (Zivanov *et*
444 *al.*, 2019*b*) **(Supp. Fig. 3)** followed by local 3D refinement resulted in a map at global
445 resolution of 1.98 Å.

446 All resolution estimates reported were obtained by considering the 0.143 threshold
447 (Rosenthal & Henderson, 2003) on the Fourier shell correlation (FSC) curve (Harauz & van
448 Heel, 1986) between independently refined half-maps (Scheres & Chen, 2012). A solvent-
449 excluding mask was generated by low-pass filtering the maps to 12 Å, binarizing the filtered

450 map and adding a soft edge consisting of cosine-shaped falloff to zero. The FSC curve was
451 corrected for artificial correlations introduced by the mask (Chen *et al.*, 2013). Local
452 resolution was estimated using the approach implemented in RELION (Cardone *et al.*,
453 2013).

454 *Model building, refinement and analysis*

455 After refinement of the map to high resolution it had to be flipped in UCSF Chimera
456 (Pettersen *et al.*, 2004) to match the correct handedness. The biological assembly from the
457 crystal structure of *Y. enterocolitica* urease **(Supp. Note 1)** was rigid-body fitted to the map
458 in Chimera. The non-crystallographic symmetry (NCS) was calculated with PHENIX v1.17
459 (Adams *et al.*, 2002) from the crystal structure. Only using the hetero-trimer of the three
460 proteins, backbone and side chains were built, corrected or confirmed in Coot (Emsley &
461 Cowtan, 2004). After several rounds of manual refinement of the model in Coot, applying
462 NCS and real-space refinement in PHENIX, the model comprised side chains of residues 1-
463 100 of ureA, 31-162 of ureB and 2-327, 335-572 of ureC. Residues 328-334 are disordered
464 and could not be modeled with confidence. NCS constraints were not used during final
465 refinements as to include alternative side chain conformations. Waters were built manually
466 and refined in PHENIX and were added to the closest chain by the program
467 phenix.sort_hetatoms. The quality of the refinement was assessed by cryo-EM Validation
468 tool **(Table 1)** (Afonine, Klaholz *et al.*, 2018).

469

470 *Structure analysis*

471 The electrostatic potential was calculated using the APBS plugin in PyMOL (Jurrus et al,
472 2018). The ConSurf server (Berezin *et al.*, 2004; Ashkenazy *et al.*, 2016) was used to find
473 150 sequences per urease protein for alignment with ClustalW (Madeira *et al.*, 2019) and
474 calculate conservation per residue. The “sample the list of homologs” option was used to
475 get a diverse representation across all species. The sequence identity of the 150 sequences

476 was determined using BLSM62 in Geneious (Kearse *et al.*, 2012). The PDBePISA server
477 (https://www.ebi.ac.uk/pdbe/pisa/) was used to find and calculate interface areas (Krissinel
478 & Henrick, 2007). Figures were created using UCSF Chimera (Pettersen *et al.*, 2004),
479 ChimeraX (Goddard *et al.*, 2018), PyMol (Schrödinger), Inkscape, Adobe Illustrator and
480 Adobe Photoshop.

481

482 *Radiation damage analysis*

483 Bayesian particle polishing in RELION (Zivanov *et al.*, 2019*a*, 2018) was carried out on a
484 sliding-window basis along the exposure, including 5 frames at a time, starting from frame
485 1 up to frame 25. Half-set reconstructions were then created from each polished particle
486 stack and post-processed using the same mask as that applied on the reconstruction from
487 all frames. On each post-processed reconstruction, real space refinement of chain C (α-
488 subunit containing the active site) from the full reconstruction was carried out in PHENIX
489 (Afonine, Poon *et al.*, 2018) for 5 macro-cycles. This procedure was repeated 5 times with
490 different random seeds. Distance between residues in the resulting refined models were
491 calculated using BioPython (Cock *et al.*, 2009) and plotted using the NumPy
492 (https://www.numpy.org) and Matplotlib (https://www.matplotlib.org) Python modules.

493

494

495

496 **References**

497 Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R.
498 J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Crystallogr. Sect. D*. **58**, 1948–1954.

499 Afonine, P. V, Klaholz, B. P., Moriarty, N. W., Poon, B. K., Sobolev, O. V, Terwilliger, T. C., Adams, P. D. &
500 Urzhumtsev, A. (2018). *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 814–840.

501 Afonine, P. V, Poon, B. K., Read, R. J., Sobolev, O. V, Terwilliger, T. C., Urzhumtsev, A. & Adams, P. D.
502 (2018). *Acta Crystallogr. Sect. D*. **74**, 531–544.

503 Arnold, S. A., Albiez, S., Bieri, A., Syntychaki, A., Adaixo, R., Mcleod, R. A., Goldie, K. N., Stahlberg, H. &
504 Braun, T. (2016). *J. Struct. Biol.* **197**, 0–6.

505 Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. & Ben-Tal, N. (2016). *Nucleic Acids Res.*
506 **44**, W344–W350.

507 Benini, S., Rypniewski, W. R., Wilson, K. S., Ciurli, S. & Mangani, S. (1998). *JBIC J. Biol. Inorg. Chem.* **3**, 268–
508 273.

509 Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P., Casadio, R. & Ben-Tal, N. (2004).
510 *Bioinformatics*. **20**, 1322–1324.

511 Biyani, N., Righetto, R. D., McLeod, R., Caujolle-Bert, D., Castano-Diez, D., Goldie, K. N. & Stahlberg, H.
512 (2017). *J. Struct. Biol.* **198**, 124–133.

513 Brilot, A. F., Chen, J. Z., Cheng, A., Pan, J., Harrison, S. C., Potter, C. S., Carragher, B., Henderson, R. &
514 Grigorieff, N. (2012). *J. Struct. Biol.* **177**, 630–637.

515 Cardone, G., Heymann, J. B. & Steven, A. C. (2013). *J. Struct. Biol.* **184**, 226–236.

516 Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H. W. & Henderson, R.
517 (2013). *Ultramicroscopy*. **135**, 24–35.

518 Cheng, A., Eng, E. T., Alink, L., Rice, W. J., Jordan, K. D., Kim, L. Y., Potter, C. S. & Carragher, B. (2018). *J.*
519 *Struct. Biol.* **204**, 270–275.

520 Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T.,
521 Kauff, F., Wilczynski, B. & de Hoon, M. J. L. (2009). *Bioinformatics*. **25**, 1422–1423.

522 Dixon, N. E., Gazzola, C., Blakeley, R. L. & Zerner, B. (1975). *J. Am. Chem. Soc.* **97**, 4131–4133.

523 Drummond, N., Murphy, B. P., Ringwood, T., Prentice, M. B., Buckley, J. F. & Fanning, S. (2012). *Foodborne*
524 *Pathog. Dis.* **9**, 179–189.

525 Emsley, P. & Cowtan, K. (2004). *Acta Crystallogr. Sect. D*. **60**, 2126–2132.

526 Fromm, S. A., Bharat, T. A. M., Jakobi, A. J., Hagen, W. J. H. & Sachse, C. (2015). *J. Struct. Biol.* **189**, 87–
527 97.

528 Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H. & Ferrin, T. E. (2018).
529 *Protein Sci.* **27**, 14–25.

530 Grant, T. & Grigorieff, N. (2015). *Elife*. **4**,.

531 Gripenberg-Lerche, C., Zhang, L., Ahtonen, P., Toivanen, P. & Skurnik, M. (2000). *Infect. Immun.* **68**, 942–
532 947.

533 Ha, N.-C., Oh, S.-T., Sung, J. Y., Cha, K. A., Lee, M. H. & Oh, B.-H. (2001). *Nat. Struct. Biol.* **8**, 505.

534 Harauz, G. & van Heel, M. (1986). *Optik (Stuttg).* **78**, 146–156.

535 Hattne, J., Shi, D., Glynn, C., Zee, C., Gallagher-Jones, M., Martynowycz, M. W., Rodriguez, J. A. & Gonen,
536 T. (2018). *Structure*. **26**, 759-766.e4.

537 Jabri, E., Carr, M., Hausinger, R. & Karplus, P. (1995). *Science (80-.).* **268**, 998–1004.

538 Kappaun, K., Regina, A., Regina, C. & Ligabue-braun, R. (2018). *J. Adv. Res.* **13**, 3–17.

539 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A.,
540 Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012). *Bioinformatics*.

28, 1647–1649.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D. & Lopez, R. (2019). *Nucleic Acids Res.* **47**, W636–W641.

Maroney, M. J. & Ciurli, S. (2014). *Chem. Rev.* **114**, 4206–4228.

Mazzei, L., Cianci, M., Contaldo, U., Musiani, F. & Ciurli, S. (2017). *Biochemistry*. **56**, 5391–5404.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.

Podschun, R. & Ullmann, U. (1998). *Clin. Microbiol. Rev.* **11**, 589–603.

Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. (2017). *Nat. Methods*. **14**, 290–296.

Rohou, A. & Grigorieff, N. (2015). *J. Struct. Biol.* **192**, 216–221.

Rokita, E., Makristathis, A., Hirschl, A. M. & Rotter, M. L. (2000). *J. Chromatogr. B Biomed. Sci. Appl.* **737**, 203–212.

Rosenthal, P. B. & Henderson, R. (2003). *J. Mol. Biol.* **333**, 721–745.

Scheres, S. H. W. W. & Chen, S. (2012). *Nat. Methods*. **9**, 853–854.

Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N. (2019). *Nat. Methods*.

Sumner, J. B. (1926). *J. Biol. Chem.* **69**, 435–441.

Weatherburn, M. W. (1967). *Anal. Chem.* **39**, 971–974.

Yano, J., Kern, J., Irrgang, K.-D., Latimer, M. J., Bergmann, U., Glatzel, P., Pushkar, Y., Biesiadka, J., Loll, B., Sauer, K., Messinger, J., Zouni, A. & Yachandra, V. K. (2005). *Proc. Natl. Acad. Sci.* **102**, 12047–12052.

Young, G. M., Amid, D. & Miller, V. L. (1996). *J. Bacteriol.* **178**, 6487–6495.

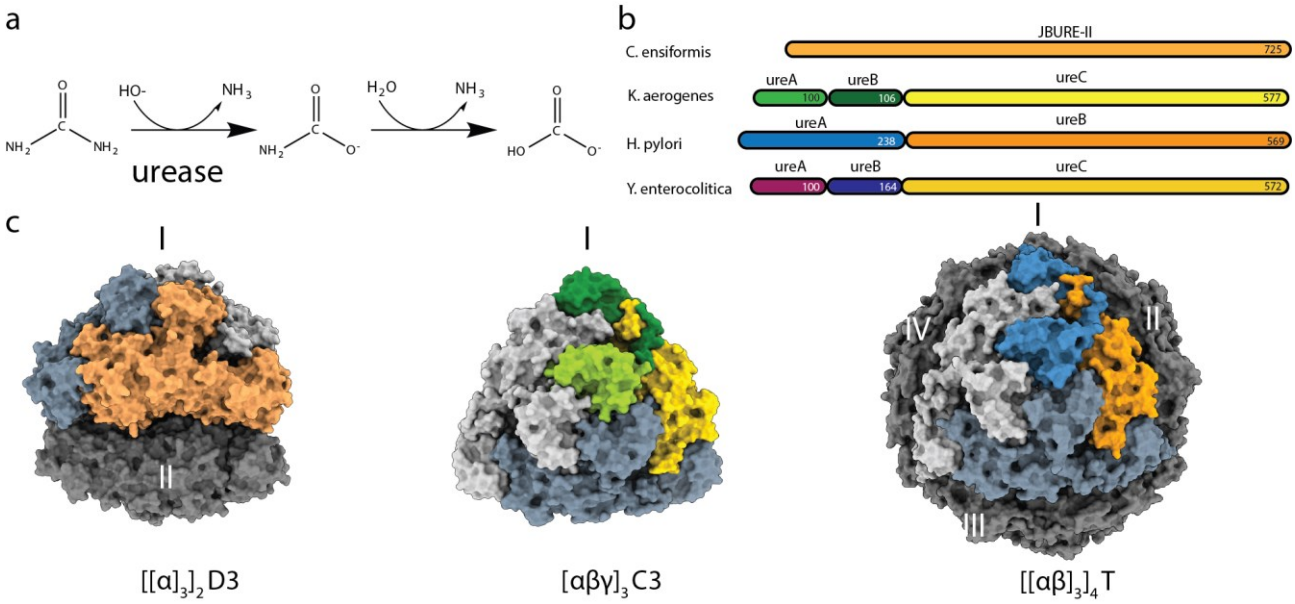
Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y. & Agard, D. A. (2017). *Nat. Methods*. **14**, 331–332.

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E. & Scheres, S. H. (2018). *Elife*. **7**, e42166.

Zivanov, J., Nakane, T. & Scheres, S. H. W. (2019*a*). *IUCrJ*. **6**,.

Zivanov, J., Nakane, T. & Scheres, S. H. W. (2019*b*). *BioRxiv*. 798066.

569 ## Figures and legends



570

571 **Figure 1** Protein architectures and oligomeric assemblies of ureases. **a)** Schematic of
572 biochemical reaction catalyzed by urease. **b)** Protein architecture of urease functional unit
573 of *C. ensiformis* (Jack bean urease), *K. aerogenes, H. pylori* and *Y. enterocolitica.* **c)** Surface
574 representation of oligomeric assembly of urease in *C. ensiformis* (PDB: 3LA4), *K. aerogenes*
575 (PDB: 1EJW), *H. pylori* (PDB: 1E9Z), respectively. Proteins are color-coded as in **b)**.
576 Oligomeric state of urease assembly is indicated at the bottom and the trimeric assemblies
577 are indicated in roman numerals.

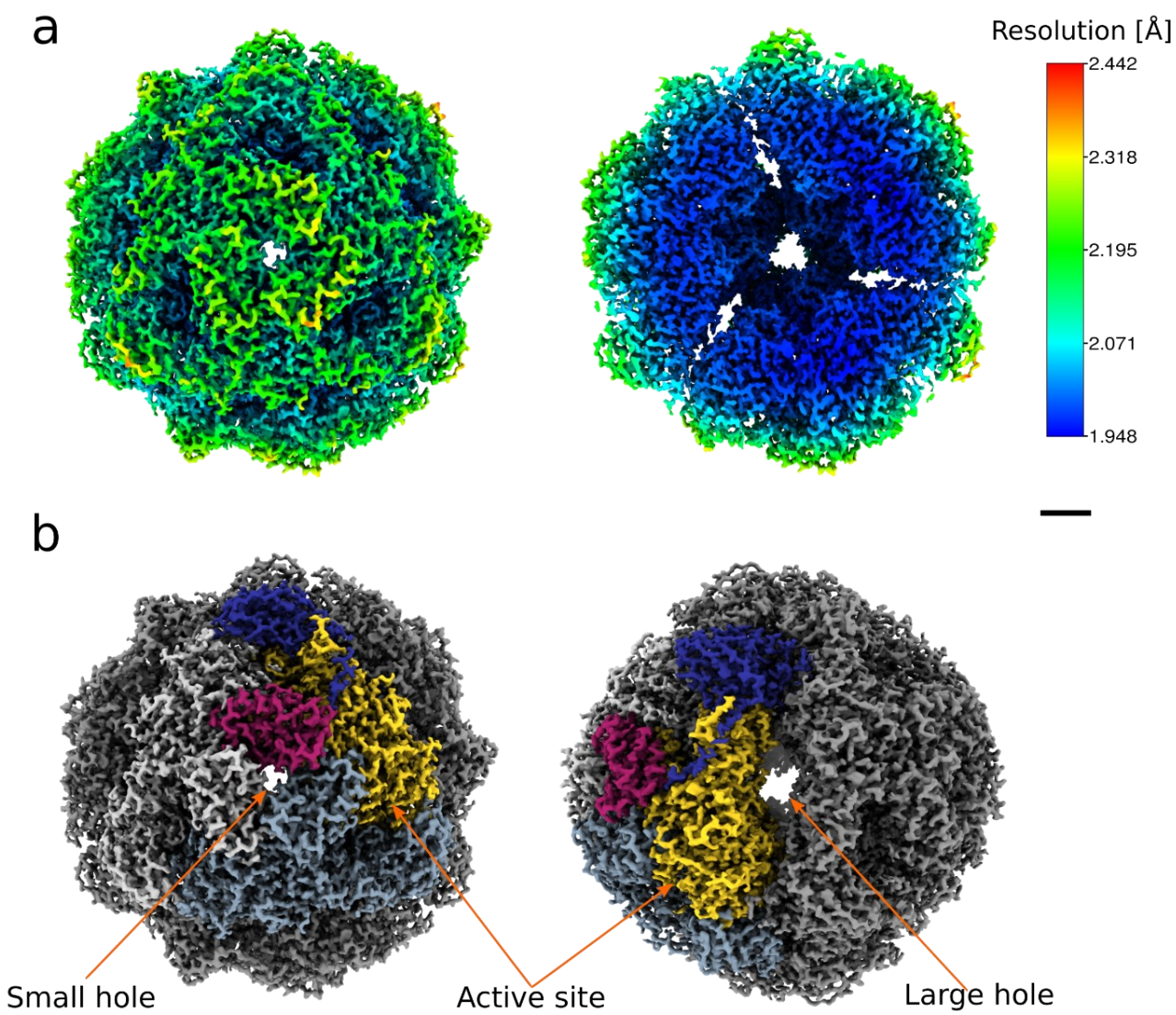
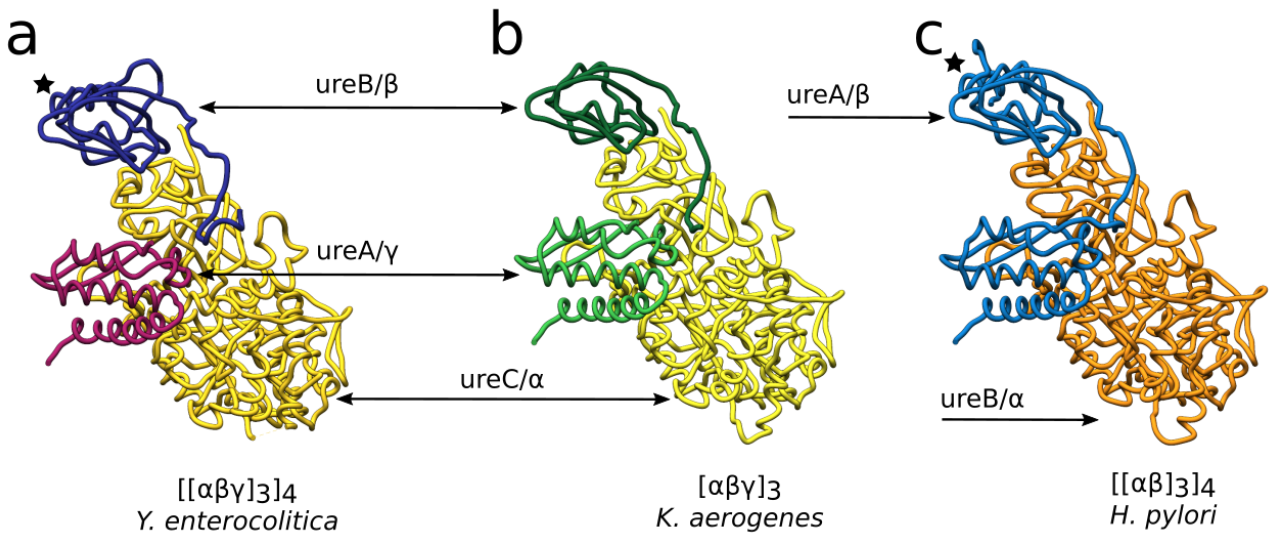


Figure 2 Cryo-EM analysis of the *Y. enterocolitica* dodecameric urease assembly. **a)** The cryo-EM map filtered and colored by local resolution (left), and a slice cut through the map to show the internal details (right). **b)** The assembly architecture highlighted on the map. The three chains that form the basic hetero-trimer are shown in different colors, with the other hetero-trimers shown in shades of gray. Two different views are shown to indicate the location of the small and larger holes at the interfaces, as well as the active site. Scale bars: 20 Å.



586

587

588 **Figure 3** Comparison of *Y. enterocolitica* urease chain architecture with ureases with
589 different modes of assembly from other pathogens. Hetero-trimers are shown in tube
590 representation with each chain in the same colors of the sequences in **Fig. 1b**. The black
591 star indicates the central helix of the β subunit. The *Y. enterocolitica* and *H. pylori* ureases
592 form the same dodecameric assembly despite having different types of chain splitting, while
593 *K. aerogenes* urease has the same type of chain splitting as in *Y. enterocolitica* but forms
594 only a trimeric assembly.

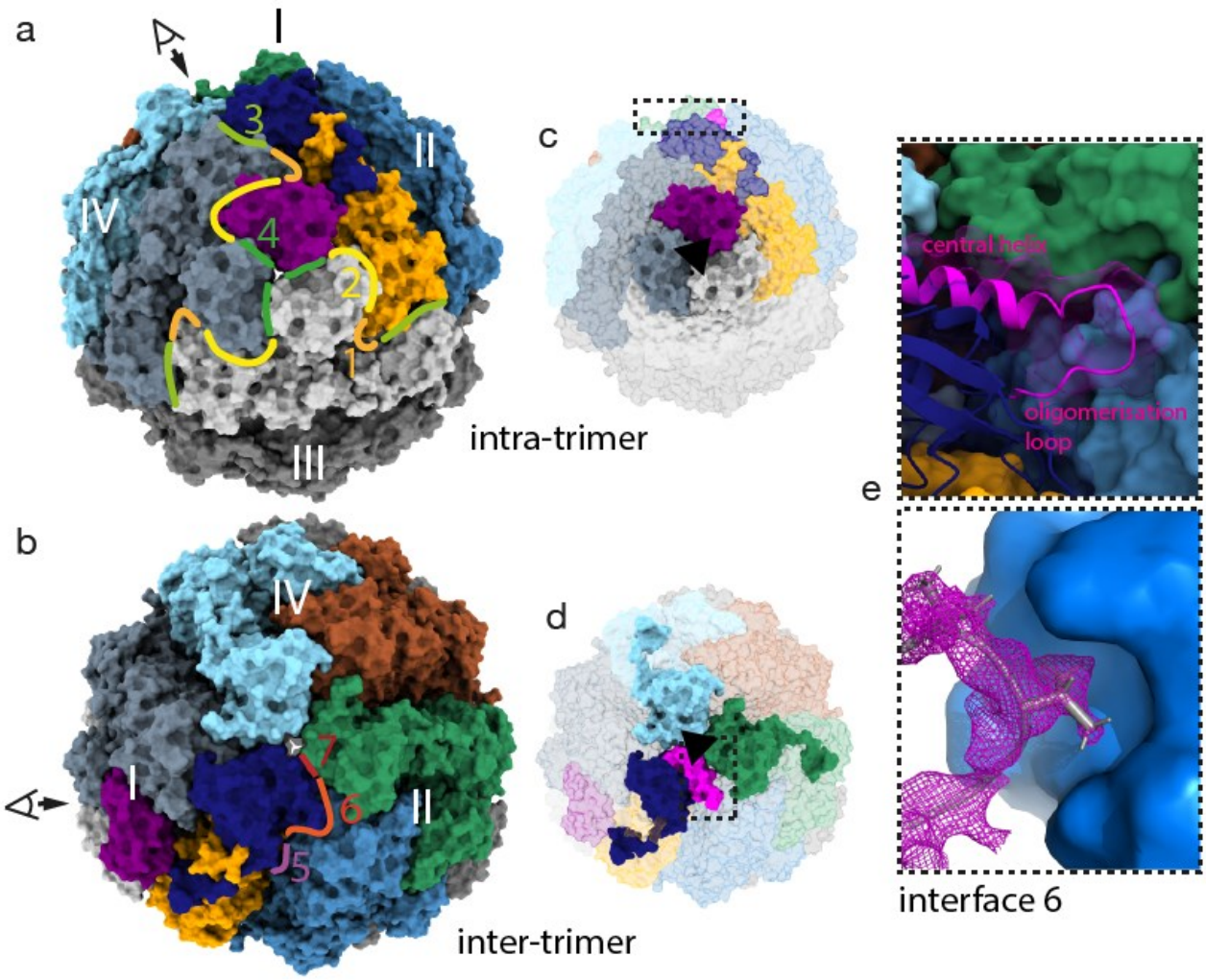


Figure 4 Interfaces in dodecameric assembly of *Y. enterocolitica*. **a)** surface model in front view of trimer of *Y. enterocolitica* urease with intra-trimeric interfaces 1-4 indicated with color-coded lines and numbers. **b)** Same model shown from the top (view indicated with eye) and the inter-trimeric interfaces. **c)** Front view with intra-trimeric-core highlighted and three-fold axis indicated with black triangle. Inset for e) in dashed box. **d)** same as c) but from the top view. **e)** Interface 6 with loop from ureB (magenta) binding into pocket of ureC of neighboring trimer. Upper inset shows ureB in cartoon and transparent surface and ureC in surface representation. Lower panel shows ureC as surface and ureB loop as cartoon with density.

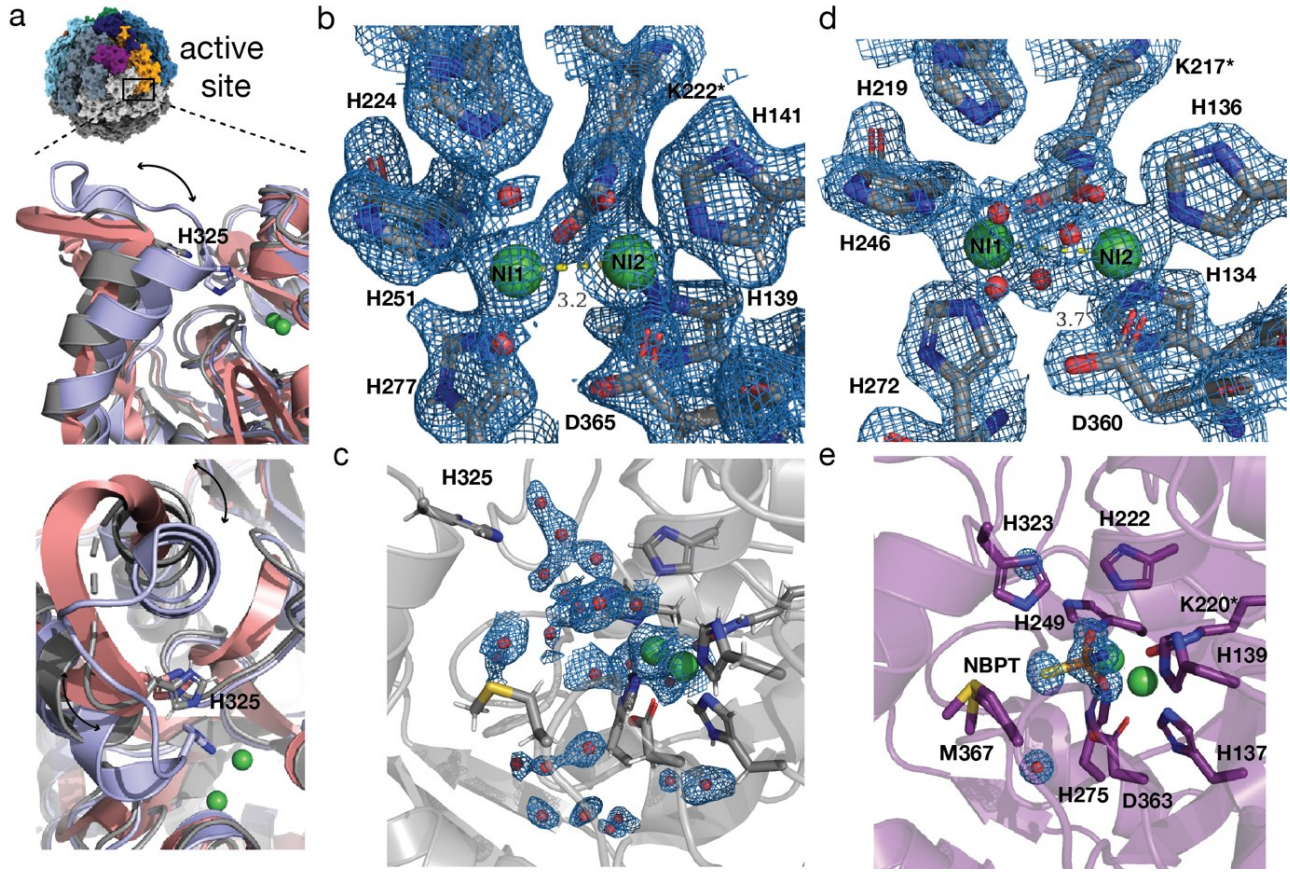


Figure 5 Active site of *Y. enterocolitica* urease. **a)** Overview of urease assembly with the active site location indicated. Inset shows in top panel side view of urease crystal structures from *K. aerogenes* mobile flap shown in open conformation in salmon (PDB: 2UBP) and in closed position as light purple (PDB: 3UBP). In gray the cryo-EM structure of *Y. enterocolitica* is overlaid and the green spheres represent the Ni^{2+} ions of the active site. Bottom panel shows top view of the three structures. Arrows indicate movement of helix and catalytic HIS325 is shown as stick. **b)** Model of active site residues and Ni^{2+} ions with the cryo-EM map of *Y. enterocolitica* at 1.98 Å nominal resolution. Yellow line indicates distance between Ni^{2+} ions in Å; **c)** shows the water molecules in the active site. **d)** Crystal structure of *K. aerogenes* urease at 1.9 Å resolution (PDB: 1EJW). Yellow line indicates distance between Ni^{2+} ions in Å. **e)** Crystal structure of *S. pasteurii* at 1.28 Å with inhibitor NBPT (PDB: 5OL4).

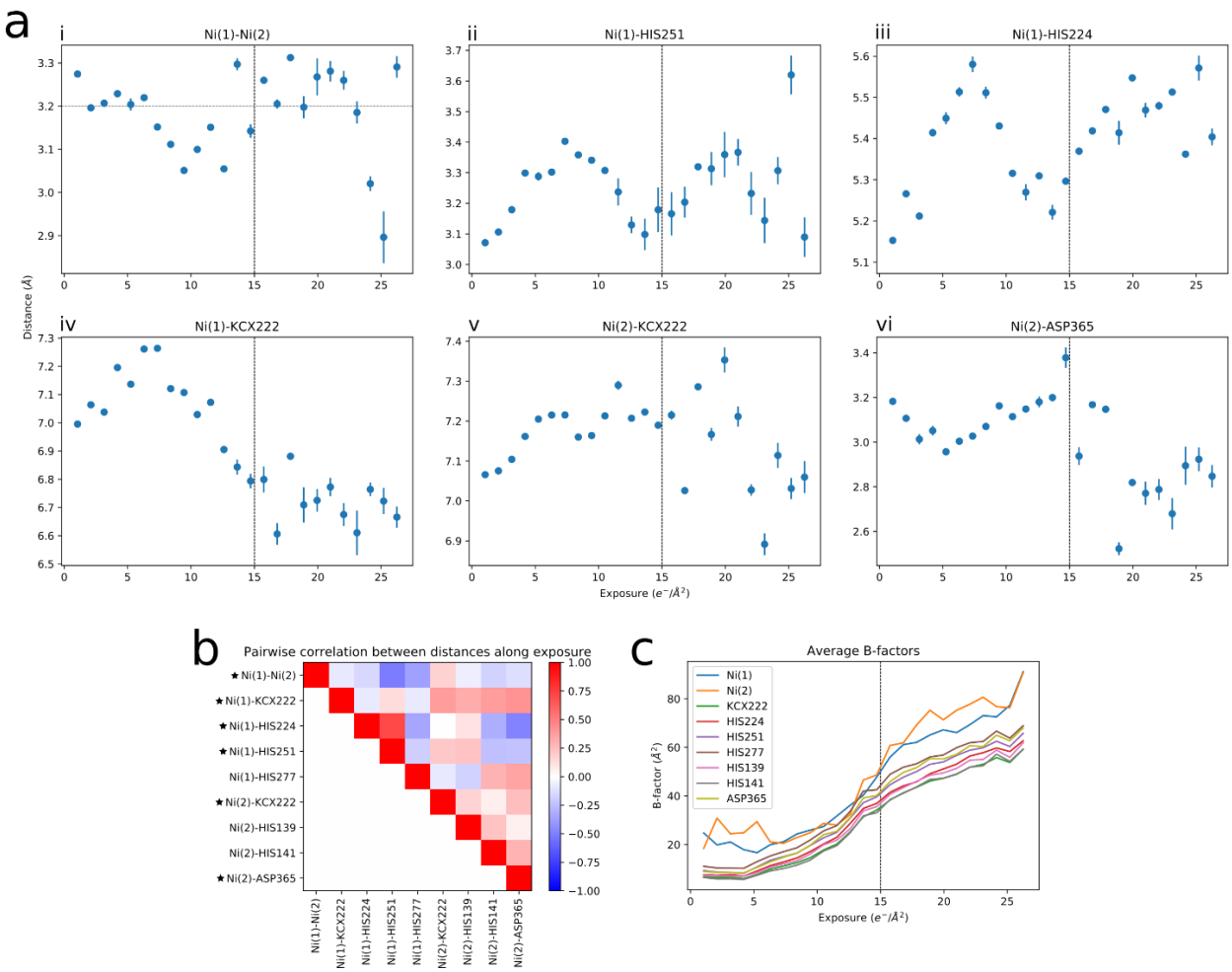


Figure 6 Radiation damage affects the distance between residues in the active site. **a)** Distances between the Ni^{2+} ions and selected residues involved in their coordination are plotted against the accumulated exposure. For each reconstruction calculated along the exposure, the model was refined, and distances measured. Dots indicate the average and error bars show +/- one standard deviation across five refinement runs with different random seeds. Horizontal dashed line in panel a-i) shows the distance in the model obtained from the full reconstruction with all frames. Vertical dashed lines show approximately the exposure at which the density for charged residues completely vanishes (see **Supp. Mov. 2**). **b)** Correlation coefficients between distance changes along the exposure for selected residues involved in ion coordination. Distance plots shown in a) are indicated with a star. **c)** Average B-factors of selected residues plotted against the accumulated exposure.

631 **Table 1** Model building and refinement. Statistics shown for full assembly calculated from
632 the asymmetric unit using NCS.

Model composition	
Chains	36
Non-hydrogen atoms	77076
Protein residues	9552
Waters	3672
Ligands (Ni^{2+})	24
Refinement	
Resolution (Å)	1.98
Average B factor ($Å^2$)	14.45
Rms deviations	
Bonds (Å)	0.066
Angles (°)	3.665
Validation (protein)	
Molprobity score	1.93
Clashscore	4.37
Good rotamers	97.26
Ramachandran plot	
Favored (%)	94.35
Allowed (%)	5.27
Outliers (%)	0.38

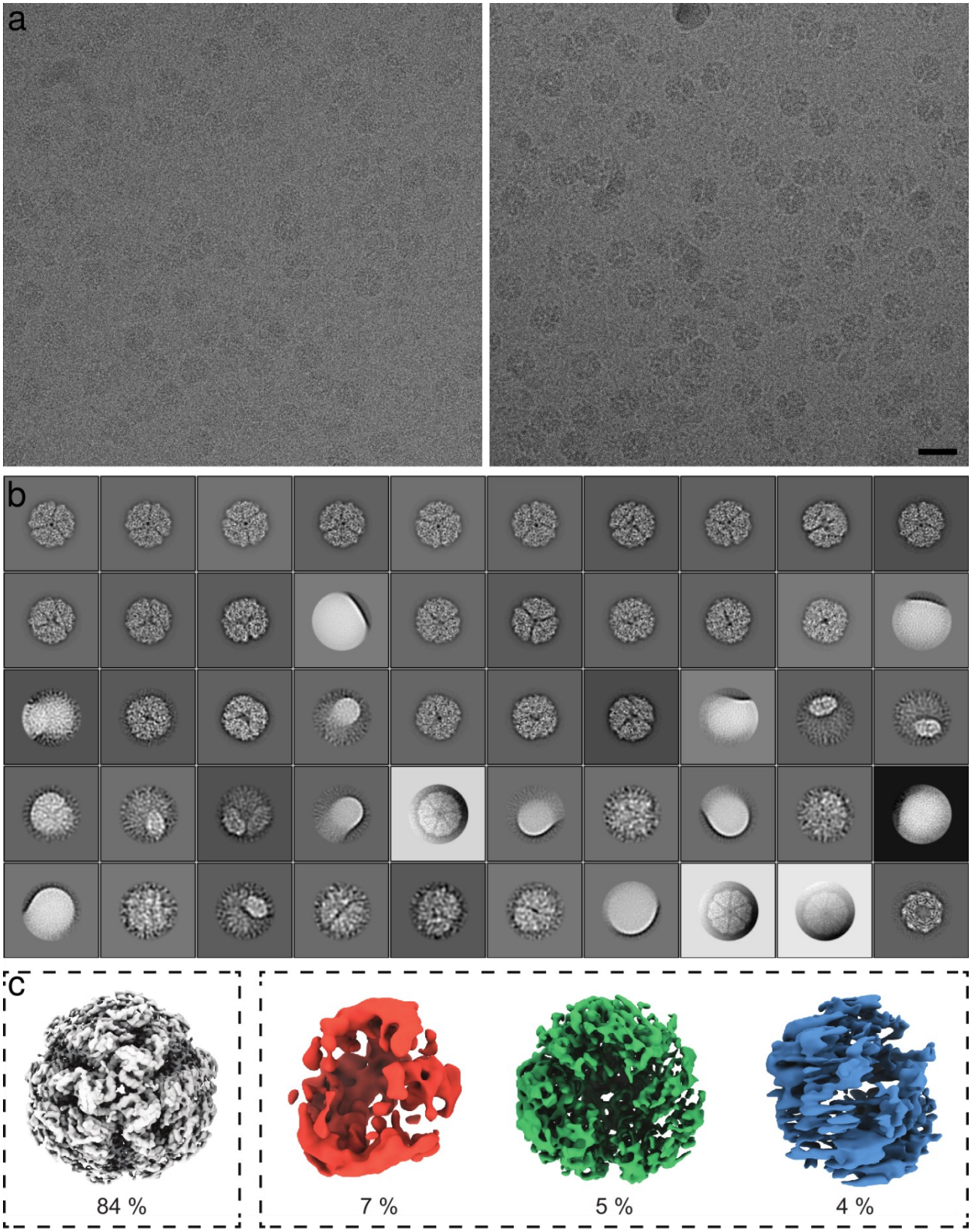
633

634 **High resolution cryo-EM structure of urease from the pathogen**

635 ***Yersinia enterocolitica*.**

636 ***Supplementary Information***

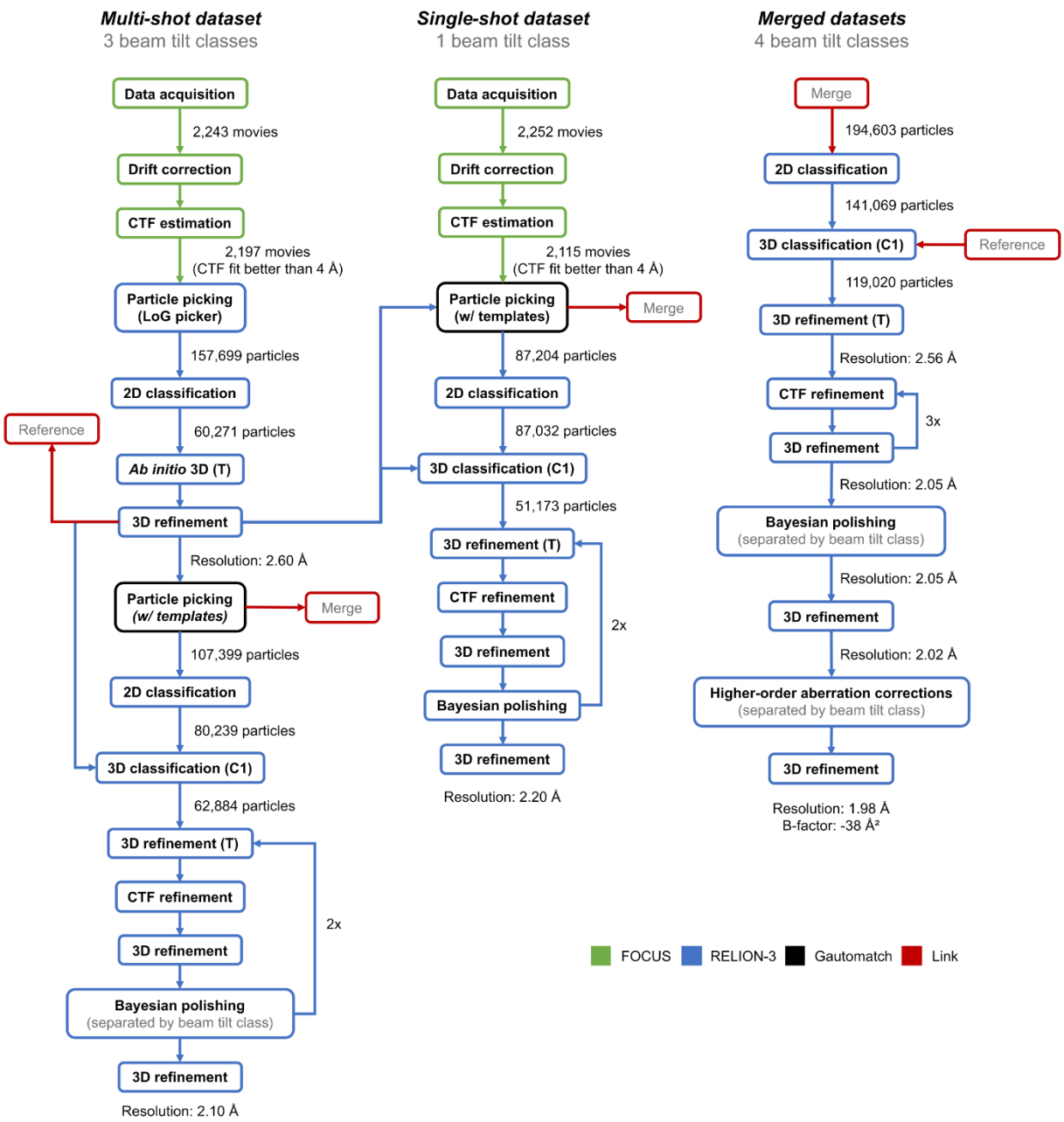
637 **Supplementary Figures**



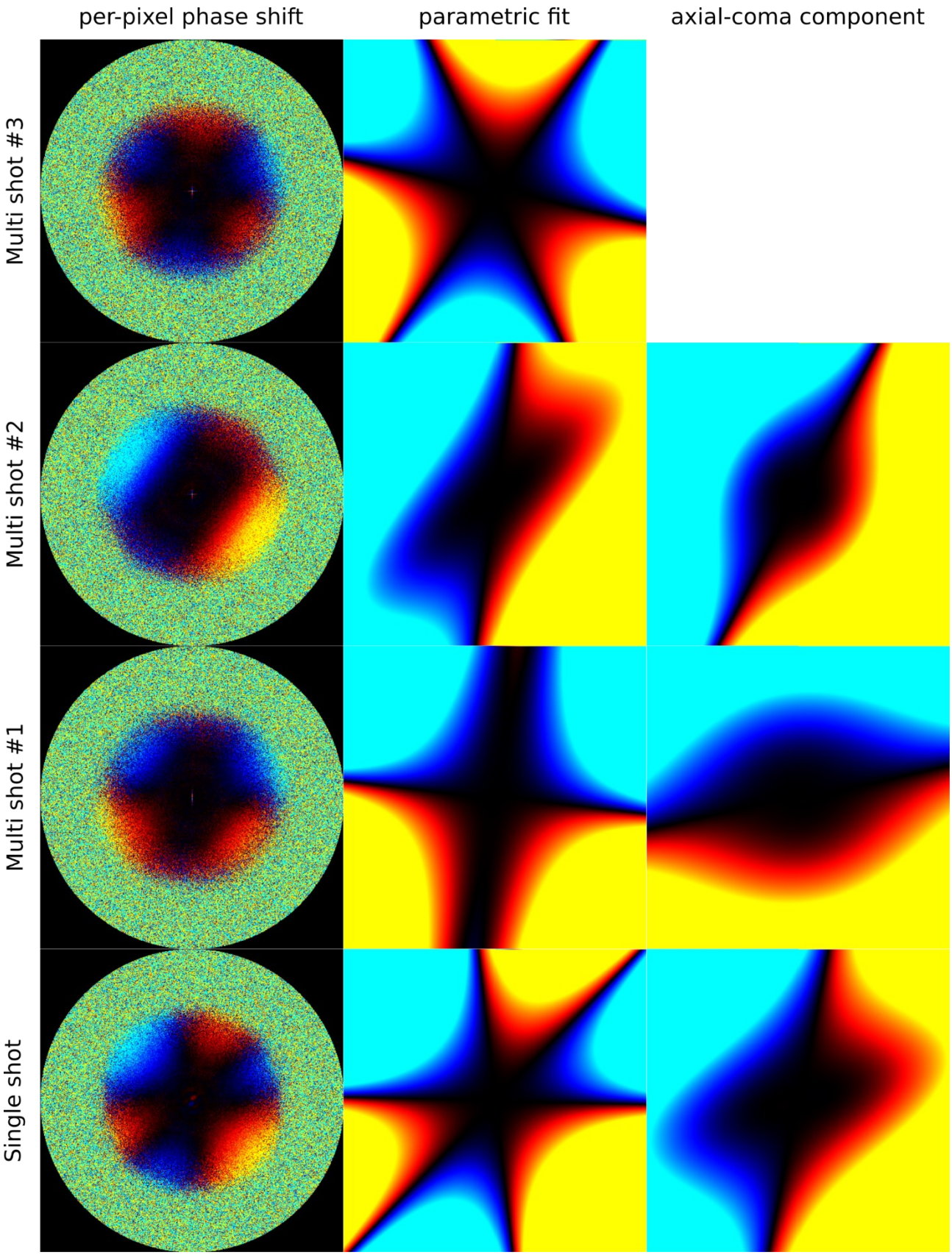
638 Complete Urease Broken Particles / Missing Subunits

639

640 **Supplementary Figure 1** Cryo-EM data of *Y. enterocolitica* urease. a) Two representative
641 micrographs from the dataset, acquired at -0.64 and -1.36 µm defocus, respectively. Scale
642 bar: 200 Å. b) 2D class averages obtained from 194'603 particles in the merged dataset.
643 These averages were obtained with RELION's "Ignore CTF until first peak" option enabled
644 and are sorted by decreasing order of number of particles in each class. Views of urease
645 with missing subunits are observed. The bottom right average shows a contamination by
646 GroEL. c) 3D class averages obtained from 141'069 particles in the merged dataset. These
647 averages were obtained without symmetry imposition in RELION. The 3D classes are sorted
648 by the indicated fraction of particles assigned to it. The first class is a complete dodecameric
649 urease assembly, while the other classes represent urease structures with at least one
650 trimer missing from the tetrahedron.



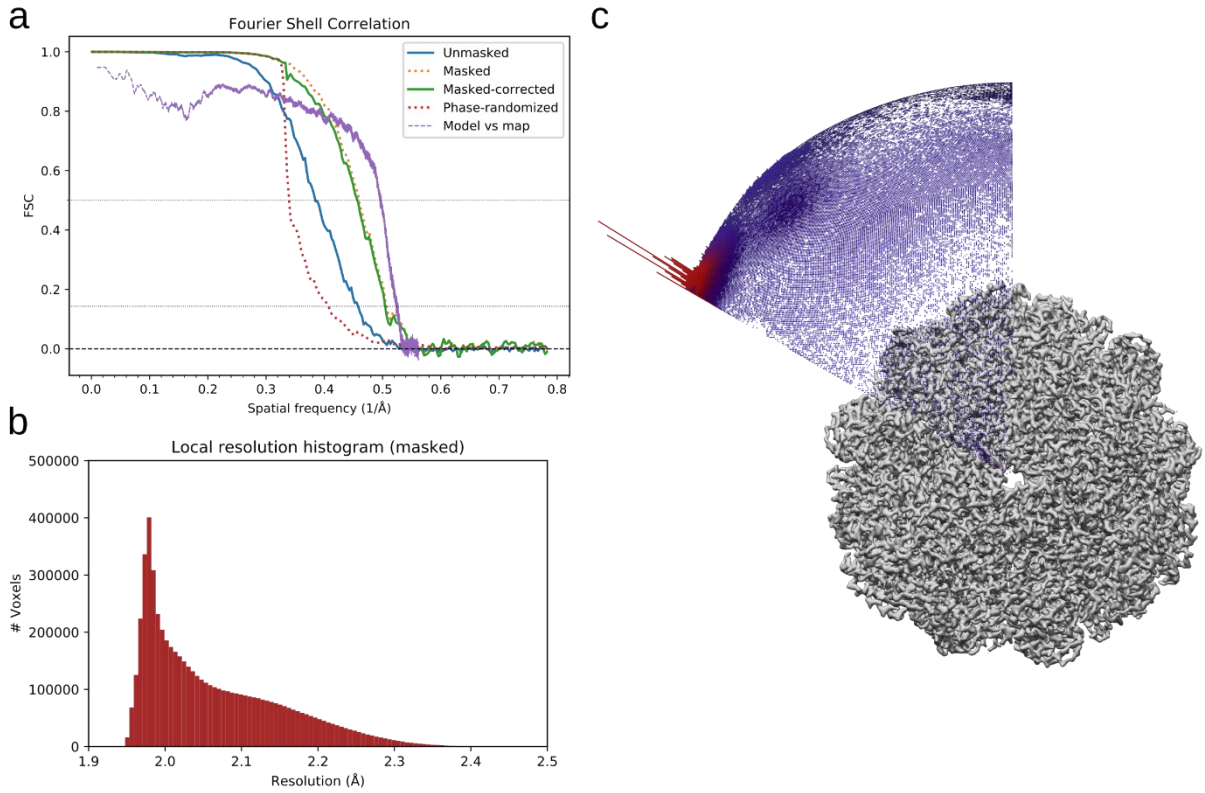
Supplementary Figure 2 Data processing flowchart for the urease cryo-EM map. Masking and postprocessing jobs have been omitted for clarity. All CTF refinement jobs included beam tilt and per-particle defocus refinement (see **Methods** for details). All resolution estimates given correspond to the corrected FSC curves between masked half-maps obtained from postprocessing jobs in RELION.


per-pixel phase shift
parametric fit
axial-coma component
Multi shot #3
Multi shot #2
Multi shot #1
Single shot

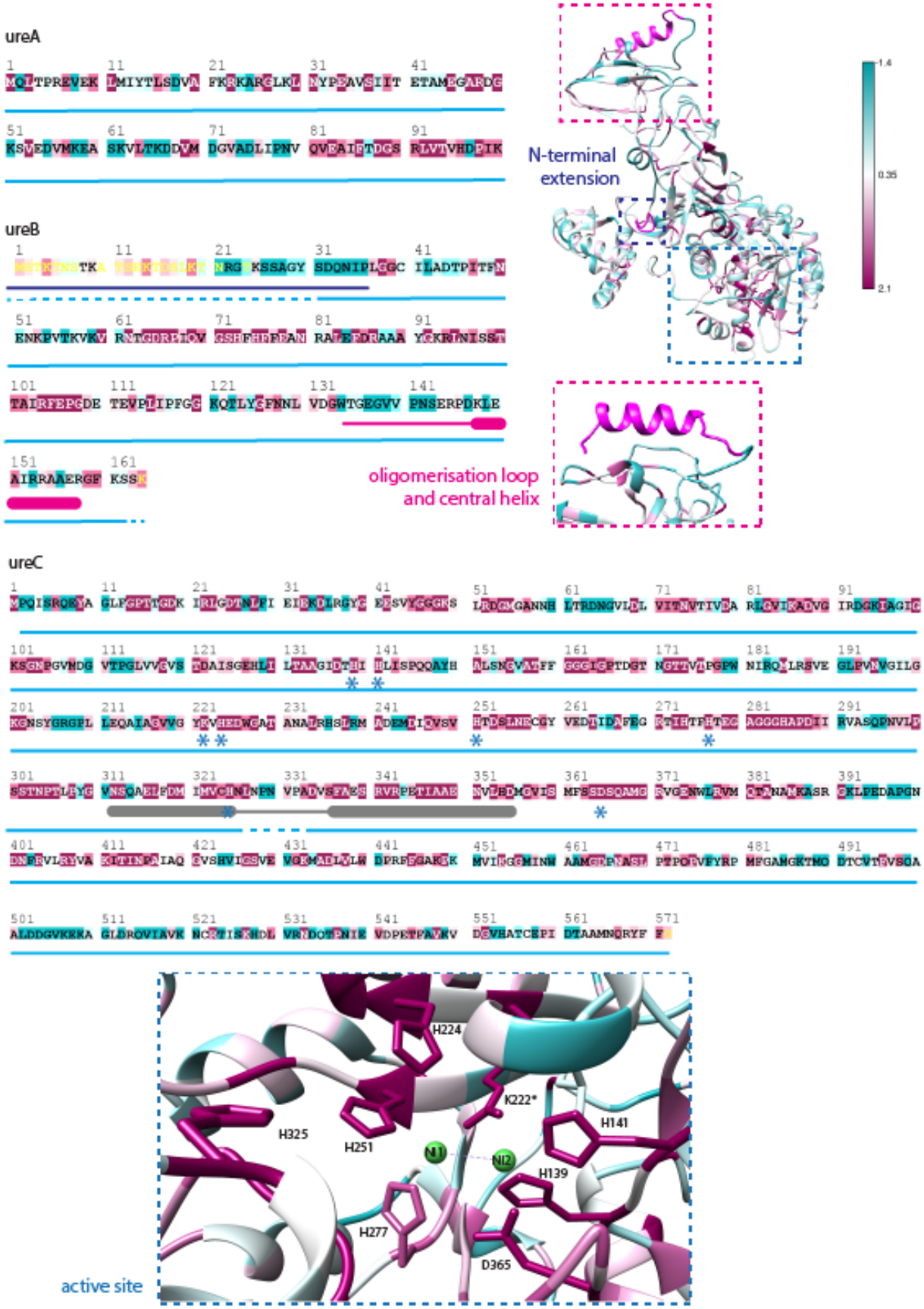
657

658

659 **Supplementary Figure 3** Fits of the anti-symmetrical aberrations arising at the four different
660 beam-shift positions. The left column shows phase shifts measured independently for each
661 Fourier pixel, while the center column shows their parametric fits using third-order Zernike
662 polynomials. The first position (top row, the third beam-shifted multi-shot) corresponds to an
663 essentially untilted beam **(Supp. Tab. 2)**, while the other two multishots and the single-shot
664 dataset exhibit tilts to different extents. Note that even the untilted dataset shows a
665 significant trefoil aberration. In the right column, the parametric fit of the first position has
666 been subtracted, yielding residuals roughly consistent with the axial coma produced by a
667 tilted beam.



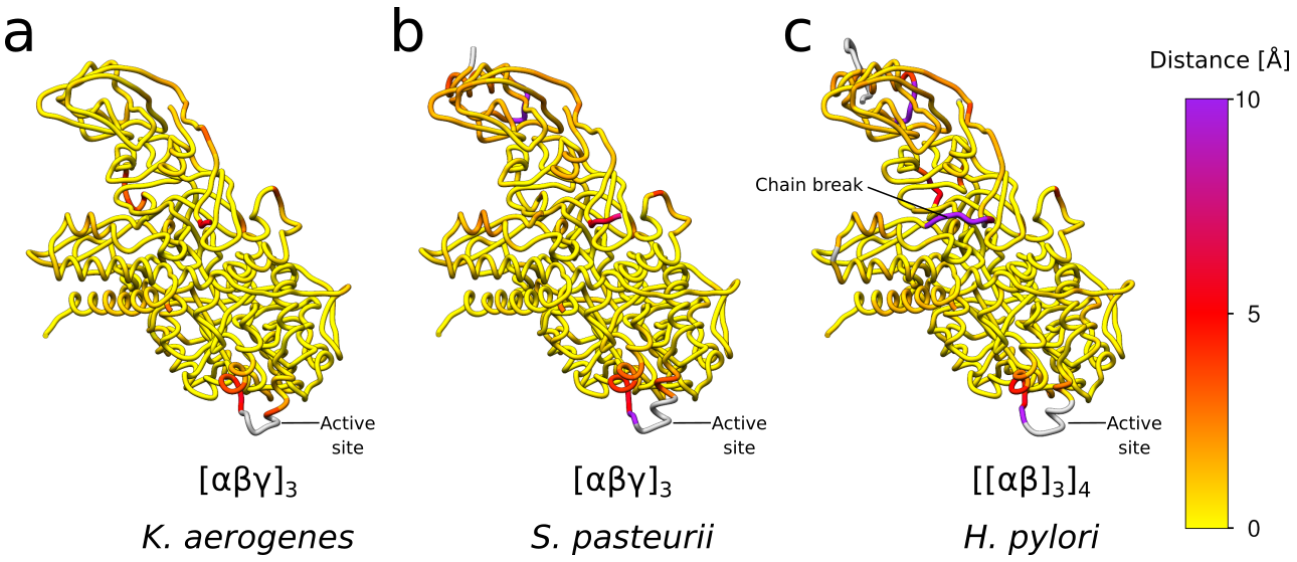
Supplementary Figure 4 Resolution estimates of the urease cryo-EM map. a) FSC curves between the half-maps when unmasked (solid blue line), masked (dashed orange), corrected by high-resolution noise substitution after masking (solid green), phase-randomized (dashed red) and between the atomic model and the full experimental map (dashed violet). b) Histogram of local resolution assigned to each voxel. c) Angular distribution of particles in the urease cryo-EM reconstruction overlaid on the unsharpened map.


ureA
ureB
ureC
N-terminal extension
oligomerisation loop and central helix
active site
H224
K222*
H141
H325
H251
Ni1
Ni2
H139
H277
D365

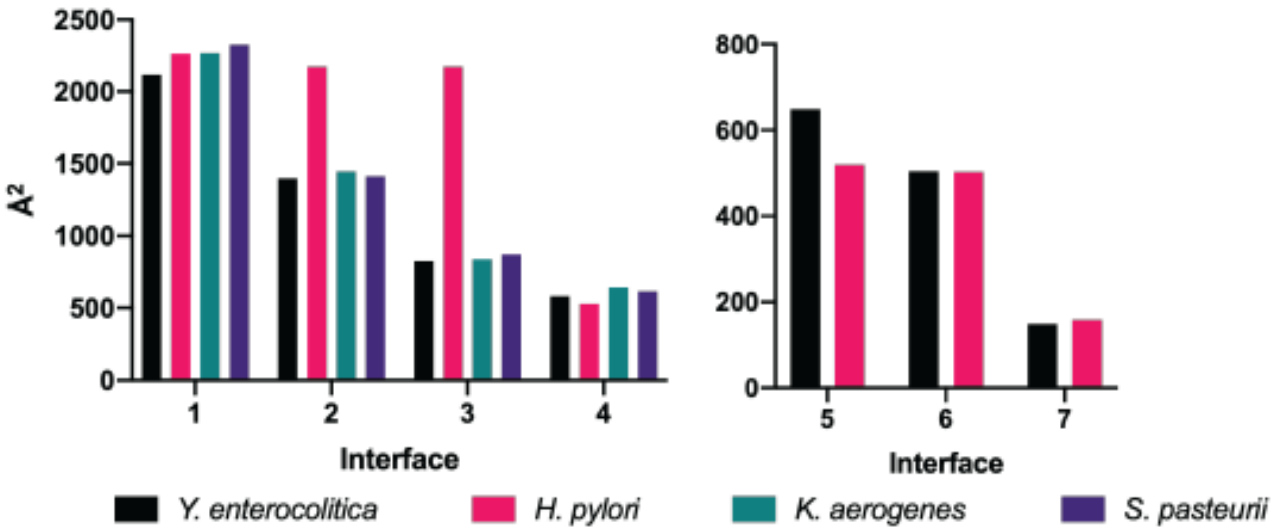
676

677

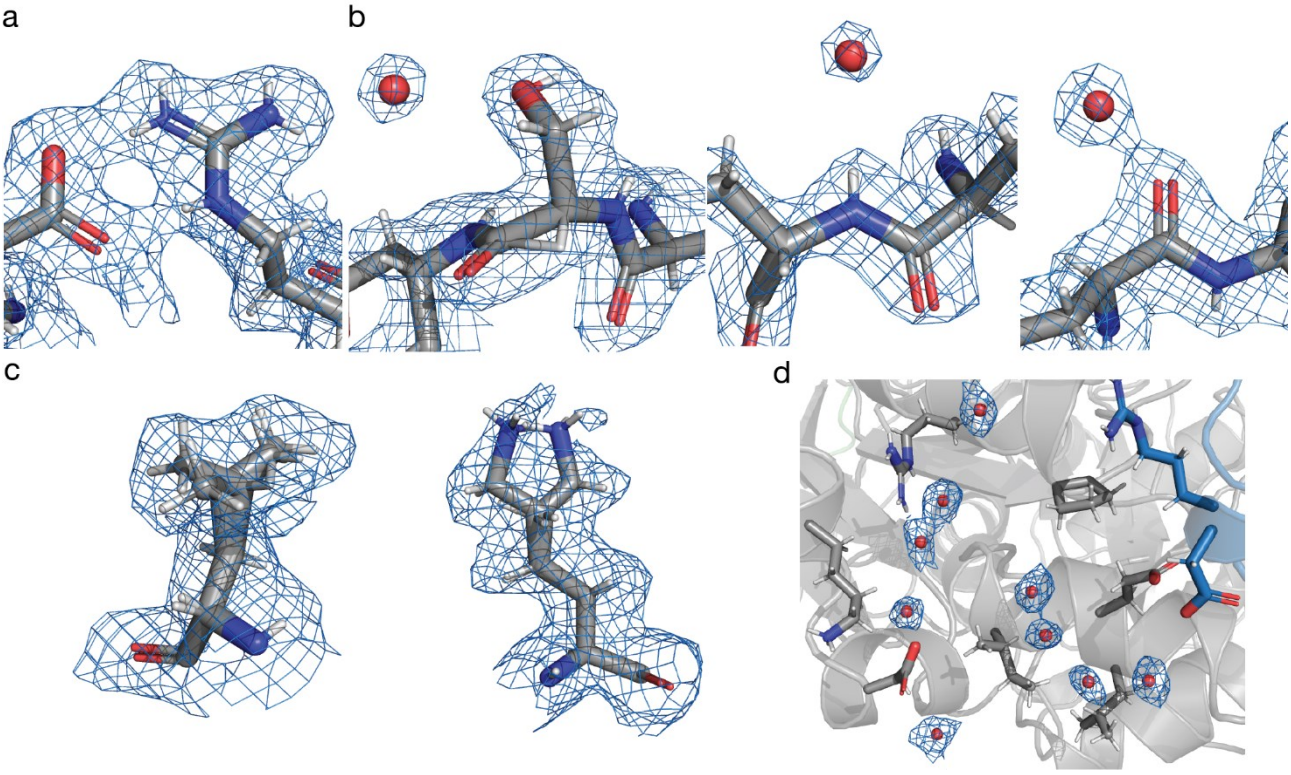
678 **Supplementary Figure 5** Alignment of 150 ureA, ureB and ureC sequences, chosen by
679 sampling from all homologs found for each protein. ClustalW was used for sequence
680 alignment and the ConSurf server for conservation analysis. Conservation is shown on a
681 gradient from dark purple to white to turquoise (arbitrary units). The light blue bar indicates
682 model completeness. Regions of interest are highlighted: The N-terminal extension in dark
683 blue, the oligomerization loop and its following helix are indicated in magenta, the mobile
684 flap is indicated in grey and the blue asterix indicates residues belonging to the active site.
685



Supplementary Figure 6 Distance comparison between residues of *Y. enterocolitica* urease against ureases with different modes of assembly. Tubes are colored by the pairwise distance of Cα atoms to the corresponding residue in *Y. enterocolitica* urease. Residues without equivalence after sequence alignment are shown in gray color. Segments with particularly high deviations are indicated.



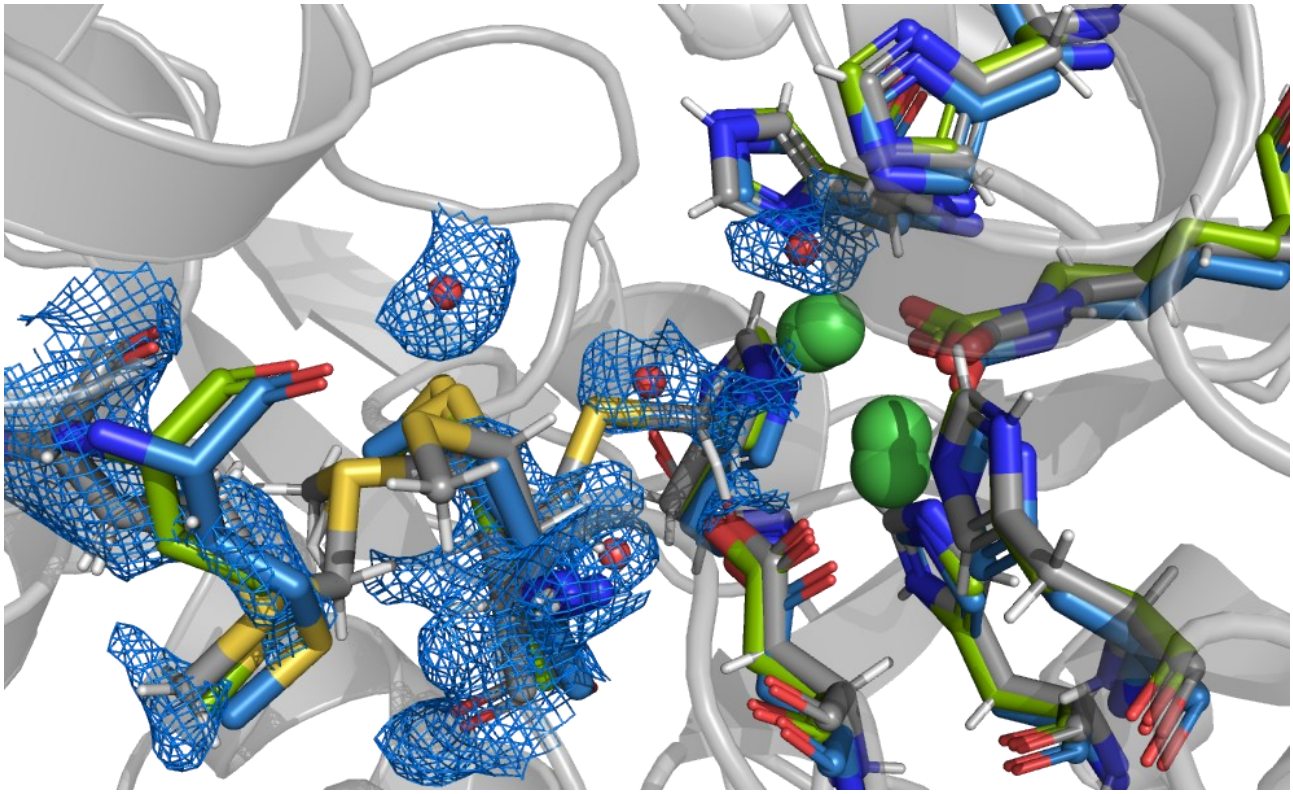
Supplementary Figure 7 Area (Å^2) of interfaces 1-7 (numbering as in Figure 4) plotted by organism. Left: intra-trimer interfaces; right: inter-trimer interfaces. The intra-trimer areas of interfaces 2 and 3 in *H. pylori* are higher because they are part of the same chain.



697

698 **Supplementary Figure 8** High-resolution structural features can be readily seen in the cryo-
699 EM map. a) Salt bridge between D69 and R63. b) Side chain hydration of S17 and back
700 bone hydration of the oxygen and the nitrogen, respectively. c) alternative side chain
701 conformations of L406 and K2. d) network of waters between two ureC proteins (grey, blue).

702



703

704 **Supplementary Figure 9** MET369 can adopt different conformations in the absence of the
705 substrate or an inhibitor. *Y. enterocolitica* urease (grey), *S. pasteurii* urease (blue), *K.*
706 *aerogenes* urease (green). One conformation could potentially reach the active site. There
707 is no described function for this amino acid in the catalysis. The conformation close to the
708 active site is only possible because the pocket is empty, which can be seen with an overlay
709 of the active site from *K. aerogenes* urease (PDB: 1EJW) at 1.9 Å and *S. pasteurii* urease
710 (PDB: 5OL4) at 1.28 Å resolution, respectively **(Figure 5d,e).**

711 # Supplementary Movies

712 **Supplementary Movie 1** Overview and feature highlights of the *Yersinia enterocolitica*
713 urease cryo-EM structure.

714 **Supplementary Movie 2** Morphing between maps and models obtained from different sets
715 of frames along the exposure.

716 # Supplementary Tables

717 **Supplementary Table 1** Cryo-EM data collection and image processing summary.

	Single shot	Multi shot
Data collection		
Microscope	Titan Krios	Titan Krios
Voltage [kV]	300	300
Direct electron detector	Gatan K2	Gatan K2
Zero-loss energy filter	GIF (20 eV slit width)	GIF (20 eV slit width)
Physical pixel size [Å]	0.639	0.639
Super-resolution mode	No	No
Total exposure [e^-/Å²]	42	42
Exposure time [s]	8	8
Number of frames [per movie]	40	40
Movies acquired	2,252	2,243
Beam-image shift?	No	3 shots per hole
Image processing (before merging)		
Movies processed	2,115	2,197
Pixel size [Å]	0.639	0.639
Box size [pixels²]	512	512
Particles picked (Gautomatch w/ templates)	87,204	107,399
Particles after 2D classification	87,032	80,239
Particles after 3D classification	51,173	62,884
Unmasked resolution [Å] (FSC 0.143)	2.52	2.39
Masked resolution [Å] (FSC 0.143)	2.20	2.10
Image processing (after merging)		
Particles in final reconstruction		119,020
Unmasked resolution [Å] (FSC 0.143)		2.20
Masked resolution [Å] (FSC 0.143)		1.98

718

719 **Supplementary Table 2** Beam tilt estimation after merging the two datasets.

	Multi-shot #1	**Multi-shot #2**	**Multi-shot #3**	**Single-shot**
Beamtilt x *(mrad)*	0.008	-0.218	-0.028	-0.078
Beamtilt y *(mrad)*	-0.091	-0.127	-0.000	-0.065

720

721 **Supplementary Table 3** Bayesian polishing training results.

	Multi Shot ***(before merging)***	**Single Shot** ***(before merging)***	**Merged**	**Multi Shot #1**	**Multi Shot #2**	**Multi Shot #3**	**Single shot**
Number of particles in training ***(approximate)***	5,000	5,000	10,000	10,000	10,000	10,000	10,000
Sigma for Velocity (Å/dose) ***(smaller = shorter tracks)***	0.531	0.819	0.779	0.810	0.693	0.636	0.681
Sigma for Divergence (Å) ***(higher = more homogeneous tracks across micrograph, i.e. "rigid block")***	3,600	10,455	10,620	10,260	7,350	5,220	8,355
Sigma for Acceleration (Å/dose) ***(smaller = straighter tracks)***	2.175	3.735	1.620	1.485	1.425	1.500	0.795

722

723 **Supplementary Table 4** Sequence identity between different ureases and *Y. enterocolitica*
724 urease.

Organism	PDB code	Nr. Of Genes	Stoichiometry α-(β)-(γ)	ureA (%)	ureB (%)	ureC (%)
S. pasteurii	5O4L	3	3-3-3	60.6	46.7	57.5
K. aerogenes	1EJW	3	3-3-3	60.6	52.9	58.7
H. pylori	1E9Z	2	12-12	52.0	50.4	57.6

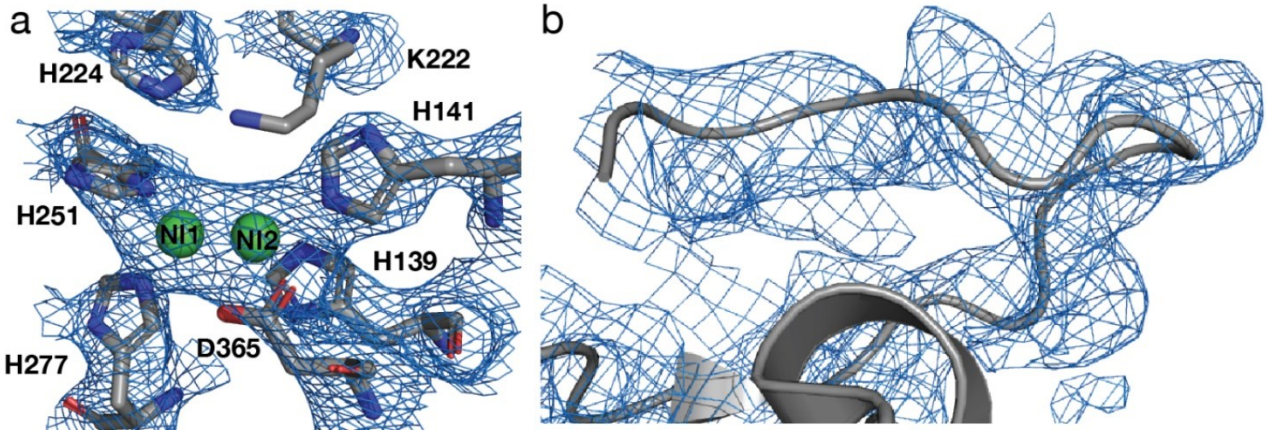
725

726 **Supplementary Table 5** RMSD values between each chain of our *Y. enterocolitica* urease
727 model versus structures deposited at the PDB. Values are given in Ångstroms. Values in
728 parenthesis indicate the number of Cα atom pairs matched with and without pruning. The
729 program matchmaker from UCSF Chimera was used to calculate sequence alignments and
730 RMSD values between each chain in respective models.

	Chain A	Chain B	Chain C	Active site
1E9Z	0.860/0.942 (98/100)	0.702/4.424 (107/124)	0.642/1.248 (537/561)	0.634
Alignment Score	339.8	376.8	1874.2	-
5OL4	0.706/0.706 (100/100)	0.756/2.999 (108/120)	0.673/1.171 (542/563)	0.293
Alignment Score	369.3	327.5	2002.7	-
1EJW	0.607/0.607 (100/100)	0.622/1.054 (97/101)	0.611/0.916 (538/559)	0.270
Alignment Score	637.2	269.4	1957.0	-

731

Supplementary Note 1



Supplementary Figure 10 Low resolution crystal structure of *Y. enterocolitica* urease. Structure determination by X-ray crystallography yielded a 3.01 Å resolution density, which was sufficient to see the higher oligomeric state of *Y. enterocolitica* urease, but not for detailed analysis of the assembly mechanism and active site.

Comparison of the model based on the high-resolution cryo-EM data *vs.* the previous crystal structure shows a few differences. The previous X-ray model did not include residue 100 of ureA and residues 31-33, 148-162 of ureB. Residues 148-162 form a helix, which was previously not observable. This helix sits right at the interface between three ureB proteins and constitutes the interaction between the trimers.

743 # Supplementary Methods

744 For X-ray crystallographic analysis of the *Y. enterocolitica* urease, *Y. enterocolitica* was
745 differently purified than described in the **Methods** for cryo-EM. The protein was precipitated
746 using 40-60% w/v ammonium sulfate, resuspended and dialyzed in 0.15M NaCl, 50mM Tris
747 pH 8.0. It was further purified using a 45 ml self-packed DEAE FF XK26/20 (Sigma, DFF100)
748 and a Superdex 200 10/300 GL column. The SEC was also used for buffer exchange to 20
749 mM HEPES, 100 mM NaCl pH 7. Urease crystals grew at 20°C at 10 mg/ml in 0.1M CHES
750 pH 9.5; 50% (v/v) PEG 200. Urease crystals belonged to space group H32 with unit cell
751 parameters of a= 157.2 Å, b = 157.2 Å and c= 774.6 Å, with four molecules per asymmetric
752 unit. The structure was determined by molecular replacement with PHASER (McCoy *et al.*,
753 2007) using the urease crystal structure from *Klebsiella aerogenes* (PDB: 1FWB) (Pearson
754 *et al.*, 1997). Model building and structure refinement were performed with Coot (Emsley &
755 Cowtan, 2004) and Buster-TNT (Blanc *et al.*, 2004). The atomic coordinates for this model
756 have been deposited in the Protein Data Bank under the accession code 4Z42.

Supplementary References

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M. & Bricogne, G. (2004). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2210–2221.

Emsley, P. & Cowtan, K. (2004). *Acta Crystallogr. Sect. D*. **60**, 2126–2132.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Crystallogr.* **40**, 658–674.

Pearson, M. A., Michel, L. O., Hausinger, R. P. & Karplus, P. A. (1997). *Biochemistry*. **36**, 8164–8172.