

Photoinduced isomerization sampling of retinal in bacteriorhodopsin

Zhong Ren (✉ zren@uic.edu)

University of Illinois at Chicago <https://orcid.org/0000-0001-7098-3127>

Article

Keywords: charge separation, proton pump, serial crystallography, singular value decomposition, X-ray free electron laser

Posted Date: September 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-916600/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

1 Photoinduced Isomerization Sampling of Retinal in 2 Bacteriorhodopsin

3
4 Zhong Ren

5
6 Department of Chemistry, University of Illinois at Chicago, Chicago, IL 60607, USA
7 Renz Research, Inc., Westmont, IL 60559, USA

8
9 zren@uic.edu

10 ORCID 0000-0001-7098-3127

11 12 **Keywords**

13 charge separation, proton pump, serial crystallography, singular value decomposition,
14 X-ray free electron laser

15 16 **Abstract**

17 Photoisomerization of retinoids inside a confined protein pocket represents a critical
18 chemical event in many important biological processes from animal vision, non-visual
19 light effects, to bacterial light sensing and harvesting. Light driven proton pumping in
20 bacteriorhodopsin entails exquisite electronic and conformational reconfigurations
21 during its photocycle. However, it has been a major challenge to delineate transient
22 molecular events preceding and following the photoisomerization of the retinal from
23 noisy electron density maps when varying populations of intermediates coexist and
24 evolve as a function of time. Here I report several distinct early photoproducts
25 deconvoluted from the recently observed mixtures in time-resolved serial
26 crystallography. This deconvolution substantially improves the quality of the electron
27 density maps hence demonstrates that the all-*trans* retinal undergoes extensive
28 isomerization sampling before it proceeds to the productive 13-*cis* configuration. Upon
29 light absorption, the chromophore attempts to perform *trans*-to-*cis* isomerization at
30 every double bond coupled with the stalled *anti*-to-*syn* rotations at multiple single
31 bonds along its polyene chain. Such isomerization sampling pushes all seven
32 transmembrane helices to bend outward, resulting in a transient expansion of the
33 retinal binding pocket, and later, a contraction due to recoiling. These ultrafast

34 responses observed at the atomic resolution support that the productive photoreaction
35 in bacteriorhodopsin is initiated by light-induced charge separation in the prosthetic
36 chromophore yet governed by stereoselectivity of its protein pocket. The method of a
37 numerical resolution of concurrent events from mixed observations is also generally
38 applicable.

39

40 **Introduction**

41 Bacteriorhodopsin (bR) pumps protons outward from the cytoplasm (CP) against the
42 concentration gradient via photoisomerization of its retinal chromophore. The trimeric
43 bR on the native purple membrane shares the seven transmembrane helical fold and the
44 same prosthetic group (Fig. S1) with large families of microbial and animal rhodopsins
45 (Ernst et al., 2014; Kandori, 2015). An all-*trans* retinal in the resting state is covalently
46 linked to Lys216 of helix G through a Schiff base (SB), of which the double bond $C_{15}=N_{\zeta}$
47 is also in *trans*. Upon absorption of a visible photon, the all-*trans* retinal in bR
48 isomerizes efficiently and selectively to adopt the 13-*cis* configuration (Govindjee et al.,
49 1990). In contrast, an all-*trans* free retinal in organic solvents could isomerize about
50 various double bonds, but with poor quantum yields (Freedman and Becker, 1986;
51 Koyama et al., 1991).

52

53 A broad consensus is that the isomerization event takes place around 450-500 fs
54 during the transition from a blue-shifted species I to form a red-shifted intermediate J
55 (Herbst, 2002; Mathies et al., 1988). Various molecular events prior to the isomerization
56 have also been detected. Vibrational spectroscopy showed a variety of possible
57 motions, such as torsions about $C_{13}=C_{14}$ and $C_{15}=N_{\zeta}$, H-out-of-plane wagging at C_{14} , and
58 even protein responses (Diller et al., 1995; Kobayashi et al., 2001). Nevertheless, the
59 species I or a collection of species detected before 30 fs remain in a good *trans*
60 configuration about $C_{13}=C_{14}$ instead of a near 90° configuration (Zhong et al., 1996).
61 Recently, deep-UV stimulated Raman spectroscopy revealed strong signals of Trp and
62 Tyr motions in the protein throughout the I and J intermediates (Tahara et al., 2019).
63 Despite extensive studies, fundamental questions on the photoisomerization of retinal
64 remain unanswered at the atomic resolution. What is the quantum mechanical force
65 that causes the all-*trans* retinal to isomerize specifically to 13-*cis* after absorbing a
66 photon? Why not isomerize elsewhere in bR? How is the quantum yield of this specific
67 isomerization enhanced by the protein compared to those of free retinal in solution?

68 Does any isomerization sampling occur? This work addresses these questions by
69 solving a series of structures of the early intermediates based on the electron density
70 maps unscrambled from the published serial crystallography datasets using singular
71 value decomposition (SVD). These structures of “pure” photoproducts at atomic
72 resolution reveal widespread conformational changes in all seven helices prior to the
73 all-*trans* to 13-*cis* isomerization and after its completion, suggesting that isomerization
74 sampling takes place in bR, where rapid photoisomerizations and single bond rotations
75 are attempted everywhere along the polyene chain of the retinal before the only
76 successful one flips the SB at ~500 fs.

77
78 Several international consortiums carried out large operations of serial
79 crystallography at free electron lasers (XFELs). It is now possible to capture transient
80 structural species at room temperature in the bR photocycle as short-lived as fs
81 (Brändén and Neutze, 2021). Compared to cryo-trapping, authentic structural signals
82 from these XFEL data are expected to be greater in both amplitude and scope.
83 However, the signals reported so far do not appear to surpass those obtained by cryo-
84 trapping methods, suggesting much needed improvements in experimental protocols
85 and data analysis methods. Two major sources of data are used in this study (Table S1).
86 Nogly et al. captured retinal isomerization to 13-*cis* by the time of 10 ps and attributed
87 the specificity to the H-bond breaking between the SB and a water (Nogly et al., 2018).
88 Kovacs et al. contributed datasets at many short time delays (Kovacs et al., 2019). Those
89 sub-ps datasets demonstrate oscillatory signals at frequencies around 100 cm⁻¹. The
90 essence of this work is a numerical resolution of structural heterogeneity, a common
91 difficulty often encountered in cryo trapping and time-resolved serial crystallography.
92 To what extent a specific structural species can be enriched in crystals depends on the
93 reaction kinetics governed by many experimental parameters including but not limited
94 to the fluence, wavelength, and temperature of the light illumination. While it is
95 possible to reach higher fractional concentrations at specific time points for more stable
96 species such as K or M due to the ratio between the rates going into and exiting from
97 that species, transient species such as I and J are often poorly populated. If such
98 structural heterogeneity is not resolved, it is very difficult, if not impossible, to interpret
99 the electron density maps and to refine the intermediate structures (Ren et al., 2013).
100 An assumption in nearly all previous studies has been that each dataset, at a cryo
101 temperature or at a time delay, is derived from a mixture of a single photoinduced

102 species and the ground state. Therefore, the difference map reveals a pure intermediate
103 structure. This assumption is far from the reality thus often leads to misinterpretation
104 of the observed electron density maps. This work is yet another case study to
105 demonstrate the application of our analytical protocol based on SVD (Methods) that
106 makes no assumption on how many excited intermediates that contribute to the
107 captured signals at each time point (Ren, 2019; Ren et al., 2013; Yang et al., 2011). More
108 importantly, this work showcases that our resolution of structural heterogeneity enables
109 new mechanistic insights into the highly dynamic chemical or biochemical processes.

110

111 **Results and Discussion**

112 A total of 24 datasets and 18 time points up to 10 ps are analyzed in this study (Table
113 S1). Difference Fourier maps at different time points and with respect to their
114 corresponding dark datasets are calculated according to the protocols previously
115 described (Methods). A collection of 126 difference maps at short delays ≤ 10 ps are
116 subjected to singular value decomposition (SVD; Methods) followed by a numerical
117 deconvolution using the previously established Ren rotation in a multi-dimensional
118 Euclidean space (Ren, 2016, 2019). Such resolution of electron density changes from
119 mixed photoexcited species in the time-resolved datasets results in four distinct
120 intermediate structures in the early photocycle, which are then refined against the
121 reconstituted structure factor amplitudes (Table S2; Methods).

122

123 *Low frequency oscillations observed upon photoexcitation*

124 Ten out of 17 major components derived from the sub-ps delays of Kovacs et al. (Fig.
125 S2) describe five two-dimensional oscillatory behaviors at frequencies ranging from 60
126 to 400 cm^{-1} (Fig. S3). Compared to a bond stretching frequency commonly observed in
127 vibrational spectroscopy, these oscillations are at much lower frequencies. The lowest
128 frequency is $61 \pm 2 \text{ cm}^{-1}$, that is, a period of 550 ± 20 fs (Fig. S3a), which matches exactly
129 the oscillation detected in transient absorption changes in visual rhodopsin (Wang et al.,
130 1994). Although these ten components follow the oscillatory time dependencies, they
131 do not show any association with the chromophore or the secondary structure of the
132 protein (Fig. S4). Similar oscillatory components were also present in the XFEL datasets
133 of MbCO (Ren, 2019). Therefore, the same conclusion stands that these low frequency
134 vibrations induced by short laser pulses often detected by ultrafast spectroscopy are the
135 intrinsic property of a solvated protein molecule, here specifically bacteriorhodopsin

136 (bR) (Johnson et al., 2014; Liebel et al., 2014). Interestingly, the isomerization sampling
137 and productive photoisomerization observed in this study occur within the first
138 oscillatory period at the lowest frequency. While such coincidence begs the question
139 whether the protein oscillation is required for isomerization (see below), direct evidence
140 is lacking in these XFEL data to support any functional relevance of these oscillatory
141 signals.

142

143 *Intermediates I', I, and expansion of retinal binding pocket*

144 In contrast to the oscillating signals, three components \mathbf{U}_{10} , \mathbf{U}_{14} , and \mathbf{U}_{17} reveal strong
145 light-induced structural signals in terms of both extensiveness and quality (Figs. 1ab
146 and S5). These signals originate exclusively from a few time points of Nogly et al., too
147 few to fit the time dependency with exponentials. Instead, a spline fitting through these
148 time points gives rise to the estimated coefficients c_{10} , c_{14} , and c_{17} in the linear
149 combination of $c_{10}\mathbf{U}_{10} + c_{14}\mathbf{U}_{14} + c_{17}\mathbf{U}_{17}$ for reconstructing the electron density maps of the
150 states I, J, and their respective precursors I', J' (Fig. 2a). A reconstituted difference map
151 of I' - bR (Fig. 1c) is located on the spline trajectory from the origin, that is, bR at the
152 time point of 0-, to the first time point of 49-406 fs (PDB entry 6g7i). This state is
153 denoted I' as a precursor leading to the I state judged by the time point at ~30 fs.
154 However, this is not to say that a single species I' exists around 30 fs. Quite the
155 opposite, the population of the time-independent conformational species I' rises and
156 falls and peaks around 30 fs, while many other species during isomerization sampling
157 coexist with I' at the same time (see below). The reconstituted difference map is used to
158 calculate a set of structure factor amplitudes that would produce this difference map of
159 I' - bR (Methods). And the structure of I' is refined against this reconstituted dataset
160 (beige; Figs. 1cd and S6). The same protocol is used to refine the structure of I state
161 (purple; Fig. S7) with a reconstituted difference map I - bR (Figs. 1a, 2ab, 3a, and S5).
162 This SVD-dependent refinement strategy extends the commonly used method based on
163 an extrapolated map to another level. This newly developed method is able to refine a
164 structure against any linear combination of signal components while eliminating noise
165 and systematic error components, and components identified as other intermediate
166 species mixed in the data. Therefore, this method enables the refinement of an
167 unscrambled, hence pure, structural species (Methods).

168

169 The all-*trans* retinal chromophore in the ground state of bR is largely flat except the
 170 last atom C₁₅ (Fig. 2c 2nd panel). In contrast, the side chain of Lys216 is highly twisted
 171 forming two near-90° single bonds (Fig. 2c 4th panel), which results in a corner at C_ε that
 172 deviates dramatically from the plane of the all-*trans* retinal (Fig. 2c 2nd panel). The
 173 refined geometry of the retinal in I' retains a near perfect all-*trans* configuration,
 174 including the Schiff base (SB) double bond C₁₅=N_ζ, while various single bonds along the
 175 polyene chain deviate from the standard *anti* conformation significantly (Fig. 2c 4th
 176 panel). The torsional deviations from *anti* are in a descending order from the β-ionone
 177 ring to the SB. These torsional changes result in an S-shaped retinal shortened by ~4%
 178 (Fig. 2c 3rd panel). The distal segment C₆-C₁₂ moves inboard up to 0.9 Å and the
 179 proximal segment C₁₃-C_ε, including the SB, moves outboard up to 1.6 Å (Fig. 2c 1st and
 180 2nd panels; see Fig. S1 for orientations in bR). This creased retinal observed here at
 181 around 30 fs (Fig. 1d) is attributed to the direct consequence of a compression under an
 182 attraction force between the β-ionone ring and the SB (see below).

183
 184 The refined structure of the I state (Fig. S7) shows that the retinal remains in near
 185 perfect all-*trans*, including the SB, and as creased as its precursor I' (Fig. 3c). The
 186 torsional deviations from *anti* single bonds become even more severe compared to the I'
 187 state and remain in a descending order from the β-ionone ring to the SB (Fig. 2c 4th
 188 panel). The major difference from its precursor is that the single bond N_ζ-C_ε now
 189 adopts a perfect *syn* conformation (Figs. 2c 4th panel and 3c), and the anchor Lys216 has
 190 largely returned to its resting conformation. Such a lack of substantial change between
 191 the ground state and the intermediate I was previously noted by a comparison of a
 192 chemically locked C₁₃=C₁₄ with the native retinal (Zhong et al., 1996).

193
 194 Remarkably, the major component **U**₁₀ reconstituted into the difference map of I – bR
 195 contains widespread signal associated with all seven helices (Fig. 2b). The reconstituted
 196 map clearly shows collective outward motions from the center (Fig. 3a) suggesting an
 197 expansion of the retinal binding pocket at hundreds of fs, which is confirmed by the
 198 refined structure of the I state (Fig. 3d top panel). For example, the distance between
 199 the C_α atoms increases by 0.8 Å between Arg82 and Phe208 and by 0.7 Å between Tyr83
 200 and Trp182. It is noteworthy that similar protein signals are present in the raw
 201 difference map calculated from the time point of 457-646 fs from Nogly et al. (6g7j) prior
 202 to an SVD analysis (Fig. S8).

203

204 Transient bleaching at near UV of 265-280 nm was observed before 200 fs and
205 attributed to structural changes in the retinal skeleton and the surrounding Trp residues
206 (Schenkl et al., 2005). Recent deep-UV stimulated Raman spectroscopy also
207 demonstrated that motions of Trp and Tyr residues start to emerge at 200 fs and remain
208 steady until the isomerization is over at 30 ps (Tahara et al., 2019). Here the refined
209 structure of the I state with displaced helices and an expanded retinal binding pocket
210 offers an explanation for the stimulated Raman gain change at hundreds of fs.
211 However, it is unclear why and how such extensive protein responses take place even
212 before the retinal isomerization. According to the broadly accepted concept of
213 proteinquake, initial motions are generated at the epicenter where the chromophore
214 absorbs a photon and then propagated throughout the protein matrix (Ansari et al.,
215 1985). It is plausible that these ultrafast protein responses are the direct consequence of
216 isomerization sampling in a confined protein pocket. It was observed in organic
217 solvents using high-pressure liquid chromatography (HPLC) that all-*trans* retinal could
218 isomerize at various double bonds along the polyene chain to adopt 9-, 11-, and 13-*cis*
219 configurations, but with rather poor quantum yields (Freedman and Becker, 1986;
220 Koyama et al., 1991). This intrinsic property of the all-*trans* retinal would behave the
221 same even when it is incorporated in the protein except that the protein matrix herds
222 the chromophores on the right track of the productive photocycle and keeps the
223 concentrations of the attempted byproducts low. These byproduct conformations of the
224 retinal during isomerization sampling are too numerous and too minor to be observed
225 experimentally. Nevertheless, they cause a common effect, an expansion of its binding
226 pocket, since the all-*trans* retinal in the resting state is tightly boxed by massive side
227 chains all around (Fig. 3e). Any attempt to isomerize would push against this box one
228 way or another. For instance, triple attempts to isomerize simultaneously at 11, 13, and
229 15 positions were suggested by a quantum mechanics/molecular mechanics simulation
230 (Altoè et al., 2010). When the retinal binding pocket is altered in mutants, the quantum
231 yield of each isomerization byproduct is expected to increase resulting in an impaired
232 productive pathway (see below).

233

234 *Intermediates J', J and productive isomerization of retinal*

235 The time point of 10 ps of Nogly et al. (6g7k) differs from the previous time point of
236 457-646 fs (6g7j) by negating the component of \mathbf{U}_{10} (Fig. 2ab), which leads to a

237 restoration of the normal retinal binding pocket in J' from an expanded one in the I state
238 followed by a contraction in J (Fig. 3d bottom panel). Two time-independent structures
239 of J' (green; Fig. S9) and J (gray; Fig. S10) are refined based on the respective
240 reconstituted difference maps with the same protocol (Methods). Their populations
241 peak at the approximate time of ~700 fs and ~20 ps, respectively. The observed
242 contraction of the retinal binding pocket is likely due to an elastic recoiling of the seven
243 helical bundle following its transient expansion caused by the isomerization sampling.
244

245 The creased retinal persists in both the J' and J structures (Fig. 2c 2nd panel and Fig.
246 3c). The difference map of J' – bR clearly shows the 13-*cis* configuration (Fig. 3b).
247 Indeed, near perfect 13-*cis* is successfully refined in both structures (Fig. 2c 4th panel).
248 While the SB double bond C₁₅=N_ζ is momentarily distorted from the *trans* configuration
249 in J' with a torsion angle of 133°, a perfect *trans* configuration at C₁₅=N_ζ is promptly
250 restored in J (Fig. 2c 4th panel). The refined structures of this series of early
251 intermediates show that the SB N_ζ is rotating clockwise in the entire process of the
252 isomerization of I' → I → J' → J, if the retinal is viewed from the proximal to distal
253 direction (Fig. 2c). It seems that the isomerization starts in an expanded retinal binding
254 pocket and finishes in a tighter one. Whether the pocket expansion and contraction are
255 required for the productive isomerization and what role the low frequency oscillations
256 play in isomerization will need more time points at short delays to further isolate the
257 molecular events temporally.
258

259 *Coulomb attraction as driving force of isomerization sampling*

260 The fundamental questions remain: What is the driving force that causes the all-*trans*
261 retinal to isomerize after a photon absorption, at several double bonds if not restrained
262 but exclusively at C₁₃=C₁₄ in bR? How does the protein environment enhance the
263 quantum yield of the isomerization to 13-*cis*? Here I hypothesize that a Coulomb
264 attraction between the β-ionone ring and the SB at the Frank-Condon point, 0+ time
265 point, provides the initial driving force upon a photon absorption. The electric field
266 spectral measurements (Mathies and Stryer, 1976) and the quantum mechanics
267 simulation (Nogly et al., 2018) suggested that a charge separation occurs along the
268 polyene chain at the excited state of bR. Such a dipole moment was also detected
269 through a transient bleaching signal at near UV region (Schenkl et al., 2005). It can be
270 shown that a plausible charge separation of ±0.1e between the β-ionone ring and the SB

271 would cause an attraction force > 1 pN. If calibrated with the measured range of dipole
272 moment of 10-16 D (Mathies and Stryer, 1976), the charge separation could reach the
273 level of $\pm 0.16e$ to $\pm 0.26e$, giving rise to an attraction force of 3.5-9 pN between the β -
274 ionone ring and the SB. This attraction force is evidently sufficient to crease the flat all-
275 *trans* retinal into an S-shape and to compress it slightly within tens of fs as observed in
276 I' and I states (Figs. 1d, 2c 2nd and 3rd panels, and 3c). In the meanwhile, this very
277 attraction force also triggers simultaneous attempts of double bond isomerizations and
278 single bond rotations along the polyene chain that cause the expansion of the retinal
279 binding pocket as observed at hundreds of fs. Following the only successful
280 isomerization at $C_{13}=C_{14}$, the chromophore segment from C_{15} to C_{δ} is attracted to the β -
281 ionone ring; and these two parts become significantly closer (Fig. 2c 3rd panel). None of
282 the single bond rotations can complete under the restraints of the protein. Especially,
283 the segment closer to the midpoint of the retinal is more confined due to the steric
284 hinderance of Thr90 and Tyr185 from the inboard and outboard sides, respectively (Fig.
285 3e). Therefore, the single bonds deviate from *anti* less and less towards the midpoint
286 (Fig. 2c 4th panel). The effect of charge separation seems eased gradually as the reaction
287 proceeds beyond the J state as indicated by the slow restoration of the *anti* conformation
288 (Fig. 2c 4th panel).

289
290 Apparently, the same charge separation and the attraction force upon photon
291 absorption also take place in a solution sample of free retinal. Compared to the retinal
292 embedded in protein, photoisomerization in solution is nonspecific, resulting in a range
293 of byproducts, since an isomerization at any position would bring the SB significantly
294 closer to the β -ionone ring. It is understandable that each of the byproducts could only
295 achieve a poor quantum yield (Freedman and Becker, 1986; Koyama et al., 1991) as
296 rotations at multiple single bonds driven by the same attraction force and achieving a
297 similar folding of the polyene chain would further sidetrack the double bond
298 isomerizations thus diminishing their quantum yields. However, these byproducts due
299 to single bond rotations are short-lived beyond detection by HPLC as they
300 spontaneously revert back in solution. The protein environment in bR plays a major
301 role in enhancing the quantum yield of the isomerization to 13-*cis* by shutting down all
302 other reaction pathways triggered by the charge separation. This is further elucidated
303 by the mutant functions below.

304

305 *Isomerization byproducts permitted by mutant protein environments*

306 The structure of a double mutant T90A/D115A (3cod) showed little difference from the
307 wildtype (Joh et al., 2008) while the single mutants T90V and T90A retain < 70% and <
308 20% of the proton pumping activity, respectively (Marti et al., 1991; Perálvarez et al.,
309 2001). These observations illustrate that some nonproductive pathways of the
310 isomerization sampling succeed more in the altered retinal binding pocket. In the
311 wildtype structure, Thr90 in helix C points towards the C₁₁=C₁₂-C₁₃-C₂₀ segment of the
312 retinal from the inboard with its C_γ atom 3.7 Å from the retinal plane. Given the van
313 der Waals radius r_c of 1.7 Å, only 0.3 Å is spared for the H atoms of the C_γ methyl group
314 thereby effectively shutting down the nonproductive pathways of the isomerization
315 sampling. Any motion of the retinal would have to push helix C toward inboard
316 causing an expansion of its binding pocket. Missing this close contact in T90A increases
317 the room to 1.9 Å for isomerization byproducts, which would greatly reduce the
318 quantum yield of the 13-*cis* productive isomerization thus retain < 20% of the activity.

319

320 In addition to 13-*cis*, the retinal in the light adapted T90V mutant showed 9- and 11-
321 *cis* configurations at the occupancies of 3% and 18%, respectively, while these
322 configurations were not detected in light adapted wildtype (Marti et al., 1991). Then
323 why would a Val residue at this position with an equivalent C_γ atom permit the
324 formation of some isomerization byproducts? In wildtype bR, the side chain of Thr90
325 engages two strong H-bonds Trp86O-Thr90O_γ-D115O_δ so that its C_γ methyl group is
326 aligned toward the retinal. Without these H-bonds in T90V, the isopropyl group of
327 Val90 is free to adopt other rotameric positions so that neither of the C_γ methyl groups
328 has to point directly to the retinal, which increases the available room for the formation
329 of some isomerization byproducts. Compared to the light adapted state, these
330 isomerization byproducts could reach even higher percentages during active
331 photocycles thus reduce the proton pumping activity below 70%.

332

333 From the outboard, the side chain of Tyr185 in helix F is nearly parallel to the retinal
334 plane with a distance of 3.5 Å. This close contact of a flat area from C₈ to C₁₄ of the
335 retinal prevents any significant motion of the retinal toward the outboard. Even slight
336 motions would push helix F away as observed here in the expansion of the retinal
337 binding pocket. The mutant Y185F largely retains the flat contact so that its proton
338 pumping activity does not reduce much (Hackett et al., 1987; Mogi et al., 1987).

339 However, it is predictable that various single mutants at this position with smaller and
340 smaller side chains would promote more and more isomerization byproducts and
341 eventually shut down proton pumping.

342
343 Two massive side chains of Trp86 and 182 from the EC and CP sides respectively do
344 not seem to play a significant role in suppressing byproduct formation as shown by the
345 mutant W182F that retains the most of the wildtype activity (Hackett et al., 1987), since
346 the motions involved in isomerization sampling are oriented more laterally. The
347 transient expansion and contraction of the retinal binding pocket (Fig. 3d) indicate that
348 the tight box surrounds the mid-segment of the retinal (Fig. 3e) is not completely rigid.
349 Rather, its plasticity must carry sufficient strength to prevent isomerization byproducts.
350 Presumably, this strength originates from the mechanical property of the helical bundle.

351
352 In summary, this work reveals the transient structural responses to many
353 unsuccessful attempts of double bond isomerization and single bond rotation by a
354 numerical resolution from the concurrent pathways, which are otherwise difficult to
355 observe. These findings underscore an important implication, that is, a nonspecific
356 Coulomb attraction provides the same driving force for the isomerization sampling
357 with and without a protein matrix. A productive isomerization at a specific double
358 bond is guided by the incorporation of the chromophore in a specific protein
359 environment. The productive pathway is selected from numerous possibilities via
360 stereochemical hinderance. Nevertheless, this nonspecific Coulomb attraction force
361 may not be directly applicable to the photoisomerization of retinal from 11-*cis* to all-
362 *trans* in the activation of visual rhodopsins. The key difference is bR as an energy
363 convertor versus a visual rhodopsin as a quantum detector (Lewis, 1978).

364 365 **Acknowledgements**

366 This work is supported in part by the grant R01EY024363 from National Institutes of
367 Health. The following database and software are used in this work: CCP4 (ccp4.ac.uk),
368 Coot (www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot), dynamix™ (Renz Research,
369 Inc.), gnuplot (gnuplot.info), PDB (rcsb.org), PHENIX (phenix-online.org), PyMOL
370 (pymol.org), Python (python.org), and SciPy (scipy.org).

371

372 **Competing interests**

373 ZR is the founder of Renz Research, Inc. that currently holds the copyright of the
374 computer software dynamix™.

375

376 **References**

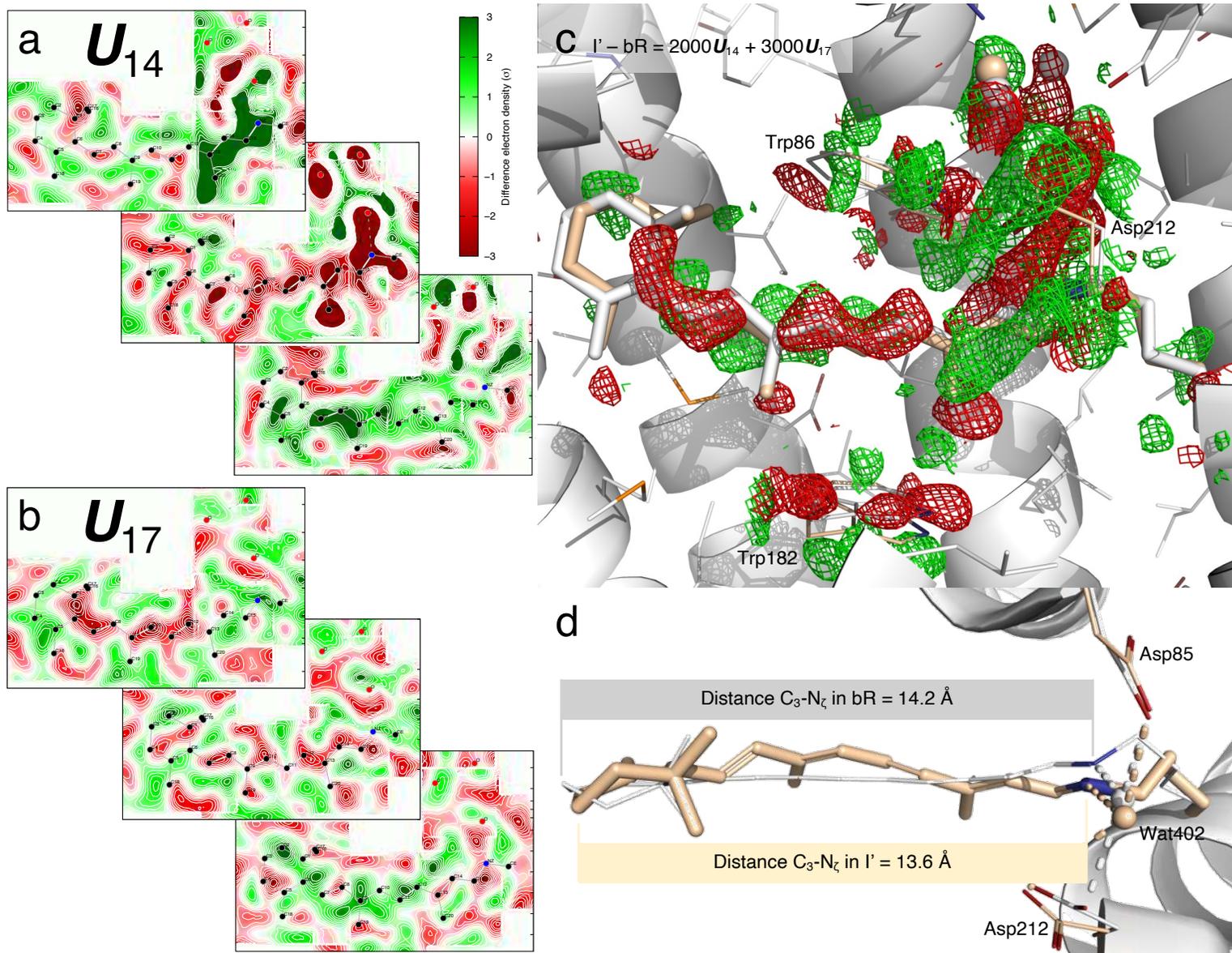
- 377 Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W.,
378 Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for
379 macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* *D66*, 213–221.
- 380 Altoè, P., Cembran, A., Olivucci, M., and Garavelli, M. (2010). Aborted double bicycle-pedal
381 isomerization with hydrogen bond breaking is the primary event of bacteriorhodopsin proton pumping.
382 *Proc. Natl. Acad. Sci.* *107*, 20172–20177.
- 383 Ansari, A., Berendzen, J., Bowne, S.F., Frauenfelder, H., Iben, I.E., Sauke, T.B., Shyamsunder, E., and
384 Young, R.D. (1985). Protein states and proteinquakes. *Proc. Natl. Acad. Sci.* *82*, 5000–5004.
- 385 Brändén, G., and Neutze, R. (2021). Advances and challenges in time-resolved macromolecular
386 crystallography. *Science* *373*, eaba0954.
- 387 Chandonia, J.-M., and Brenner, S.E. (2006). The impact of structural genomics: expectations and
388 outcomes. *Science* *311*, 347–351.
- 389 Diller, R., Maiti, S., Walker, G.C., Cowen, B.R., Pippenger, R., Bogomolni, R.A., and Hochstrasser, R.M.
390 (1995). Femtosecond time-resolved infrared laser study of the J–K transition of bacteriorhodopsin. *Chem.*
391 *Phys. Lett.* *241*, 109–115.
- 392 Ernst, O.P., Lodowski, D.T., Elstner, M., Hegemann, P., Brown, L.S., and Kandori, H. (2014). Microbial
393 and animal rhodopsins: Structures, functions, and molecular mechanisms. *Chem. Rev.* *114*, 126–163.
- 394 Freedman, K.A., and Becker, R.S. (1986). Comparative investigation of the photoisomerization of the
395 protonated and unprotonated n-butylamine Schiff bases of 9-cis-, 11-cis-, 13-cis-, and all-trans-retinals. *J.*
396 *Am. Chem. Soc.* *108*, 1245–1251.
- 397 Glynn, C., and Rodriguez, J.A. (2019). Data-driven challenges and opportunities in crystallography.
398 *Emerg. Top. Life Sci.* ETL520180177.
- 399 Govindjee, R., Balashov, S.P., and Ebrey, T.G. (1990). Quantum efficiency of the photochemical cycle of
400 bacteriorhodopsin. *Biophys. J.* *58*, 597–608.
- 401 Hackett, N.R., Stern, L.J., Chao, B.H., Kronis, K.A., and Khorana, H.G. (1987). Structure-function studies
402 on bacteriorhodopsin. V. Effects of amino acid substitutions in the putative helix F. *J. Biol. Chem.* *262*,
403 9277–9284.
- 404 Henry, E.R., and Hofrichter, J. (1992). Singular value decomposition: Application to analysis of
405 experimental data. In *Numerical Computer Methods*, (Academic Press), pp. 129–192.
- 406 Herbst, J. (2002). Femtosecond infrared spectroscopy of bacteriorhodopsin chromophore isomerization.
407 *Science* *297*, 822–825.
- 408 Joh, N.H., Min, A., Faham, S., Whitelegge, J.P., Yang, D., Woods, V.L., and Bowie, J.U. (2008). Modest
409 stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* *453*, 1266–
410 1270.

- 411 Johnson, P.J.M., Halpin, A., Morizumi, T., S. Brown, L., I. Prokhorenko, V., P. Ernst, O., and Miller, R.J.D.
412 (2014). The photocycle and ultrafast vibrational dynamics of bacteriorhodopsin in lipid nanodiscs. *Phys.*
413 *Chem. Chem. Phys.* *16*, 21310–21320.
- 414 Jung, Y.O., Lee, J.H., Kim, J., Schmidt, M., Moffat, K., Šrajcar, V., and Ihee, H. (2013). Volume-conserving
415 trans–cis isomerization pathways in photoactive yellow protein visualized by picosecond X-ray
416 crystallography. *Nat. Chem.* *5*, 212–220.
- 417 Kandori, H. (2015). Ion-pumping microbial rhodopsins. *Front. Mol. Biosci.* *2*.
- 418 Kobayashi, T., Saito, T., and Ohtani, H. (2001). Real-time spectroscopy of transition states in
419 bacteriorhodopsin during retinal isomerization. *Nature* *414*, 531–534.
- 420 Kovacs, G.N., Colletier, J.-P., Grünbein, M.L., Yang, Y., Stensitzki, T., Batyuk, A., Carbajo, S., Doak, R.B.,
421 Ehrenberg, D., Foucar, L., et al. (2019). Three-dimensional view of ultrafast dynamics in photoexcited
422 bacteriorhodopsin. *Nat. Commun.* *10*, 3177.
- 423 Koyama, Y., Kubo, K., Komori, M., Yasuda, H., and Mukai, Y. (1991). Effect of protonation on the
424 isomerization properties of n-butylamine Schiff base of isomeric retinal as revealed by direct HPLC
425 analyses: Selection of isomerization pathways by retinal proteins. *Photochem. Photobiol.* *54*, 433–443.
- 426 Lanyi, J.K., and Schobert, B. (2007). Structural changes in the L photointermediate of bacteriorhodopsin. *J.*
427 *Mol. Biol.* *365*, 1379–1392.
- 428 Lewis, A. (1978). The molecular mechanism of excitation in visual transduction and bacteriorhodopsin.
429 *Proc. Natl. Acad. Sci.* *75*, 549–553.
- 430 Liebel, M., Schnedermann, C., Bassolino, G., Taylor, G., Watts, A., and Kukura, P. (2014). Direct
431 observation of the coherent nuclear response after the absorption of a photon. *Phys. Rev. Lett.* *112*,
432 238301.
- 433 Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W.,
434 Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and
435 electrons: recent developments in Phenix. *Acta Crystallogr. Sect. Struct. Biol.* *75*, 861–877.
- 436 Marti, T., Otto, H., Mogi, T., Rösselet, S.J., Heyn, M.P., and Khorana, H.G. (1991). Bacteriorhodopsin
437 mutants containing single substitutions of serine or threonine residues are all active in proton
438 translocation. *J. Biol. Chem.* *266*, 6919–6927.
- 439 Mathies, R., and Stryer, L. (1976). Retinal has a highly dipolar vertically excited singlet state: implications
440 for vision. *Proc. Natl. Acad. Sci.* *73*, 2169–2173.
- 441 Mathies, R., Brito Cruz, C., Pollard, W., and Shank, C. (1988). Direct observation of the femtosecond
442 excited-state cis-trans isomerization in bacteriorhodopsin. *Science* *240*, 777–779.
- 443 Mogi, T., Stern, L.J., Hackett, N.R., and Khorana, H.G. (1987). Bacteriorhodopsin mutants containing
444 single tyrosine to phenylalanine substitutions are all active in proton translocation. *Proc. Natl. Acad. Sci.*
445 *84*, 5595–5599.
- 446 Nogly, P., Weinert, T., James, D., Carbajo, S., Ozerov, D., Furrer, A., Gashi, D., Borin, V., Skopintsev, P.,
447 Jaeger, K., et al. (2018). Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser.
448 *Science* *361*, eaat0094.
- 449 Perálvarez, A., Barnadas, R., Sabés, M., Querol, E., and Padrós, E. (2001). Thr90 is a key residue of the
450 bacteriorhodopsin proton pumping mechanism. *FEBS Lett.* *508*, 399–402.
- 451 Ren, Z. (2013a). Reaction trajectory revealed by a joint analysis of Protein Data Bank. *PLoS ONE* *8*, e77141.

- 452 Ren, Z. (2013b). Reverse engineering the cooperative machinery of human hemoglobin. *PLoS ONE* 8,
453 e77363.
- 454 Ren, Z. (2016). Molecular events during translocation and proofreading extracted from 200 static
455 structures of DNA polymerase. *Nucleic Acids Res.* 6, 1–13.
- 456 Ren, Z. (2019). Ultrafast structural changes decomposed from serial crystallographic data. *J. Phys. Chem.*
457 *Lett.* 10, 7148–7163.
- 458 Ren, Z., Perman, B., Srajer, V., Teng, T.-Y., Praderyand, C., Bourgeois, D., Schotte, F., Ursby, T., Kort, R.,
459 Wulff, M., et al. (2001). A molecular movie at 1.8 Å resolution displays the photocycle of photoactive
460 yellow protein, a eubacterial blue-light receptor, from nanoseconds to seconds. *Biochemistry* 40, 13788–
461 13801.
- 462 Ren, Z., Chan, P.W.Y., Moffat, K., Pai, E.F., Royer, W.E., Šrajer, V., and Yang, X. (2013). Resolution of
463 structural heterogeneity in dynamic crystallography. *Acta Cryst D* 69, 946–959.
- 464 Schaffer, J.E., Kukshal, V., Miller, J.J., Kitainda, V., and Jez, J.M. (2021). Beyond X-rays: an overview of
465 emerging structural biology methods. *Emerg. Top. Life Sci.* ETL520200272.
- 466 Schenkl, S., Mourik, F. van, Zwan, G. van der, Haacke, S., and Chergui, M. (2005). Probing the ultrafast
467 charge translocation of photoexcited retinal in bacteriorhodopsin. *Science* 309, 917–920.
- 468 Schmidt, M., Rajagopal, S., Ren, Z., and Moffat, K. (2003). Application of singular value decomposition to
469 the analysis of time-resolved macromolecular X-ray data. *Biophys. J.* 84, 2112–2129.
- 470 Schmidt, M., Graber, T., Henning, R., and Srajer, V. (2010). Five-dimensional crystallography. *Acta*
471 *Crystallogr. A* 66, 198–206.
- 472 Šrajer, V., Ren, Z., Teng, T.-Y., Schmidt, M., Ursby, T., Bourgeois, D., Praderyand, C., Schildkamp, W.,
473 Wulff, M., and Moffat, K. (2001). Protein conformational relaxation and ligand migration in myoglobin: A
474 nanosecond to millisecond molecular movie from time-resolved Laue X-ray diffraction. *Biochemistry* 40,
475 13802–13815.
- 476 Tahara, S., Kuramochi, H., Takeuchi, S., and Tahara, T. (2019). Protein dynamics preceding
477 photoisomerization of the retinal chromophore in bacteriorhodopsin revealed by deep-UV femtosecond
478 stimulated Raman spectroscopy. *J. Phys. Chem. Lett.* 10, 5422–5427.
- 479 Ursby, T., and Bourgeois, D. (1997). Improved estimation of structure-factor difference amplitudes from
480 poorly accurate data. *Acta Crystallogr. A* 53, 564–575.
- 481 Wang, Q., Schoenlein, R.W., Peteanu, L.A., Mathies, R.A., and Shank, C.V. (1994). Vibrationally coherent
482 photochemistry in the femtosecond primary event of vision. *Science* 266, 422–424.
- 483 Yang, X., Ren, Z., Kuk, J., and Moffat, K. (2011). Temperature-scan cryocrystallography reveals reaction
484 intermediates in bacteriophytochrome. *Nature* 479, 428–432.
- 485 Zhong, Q., Ruhman, S., Ottolenghi, M., Sheves, M., Friedman, N., Atkinson, G.H., and Delaney, J.K.
486 (1996). Reexamining the primary light-induced events in bacteriorhodopsin using a synthetic C13=C14-
487 locked chromophore. *J Am Chem Soc* 118, 12828–12829.
- 488 (2012). The Protein Data Bank at 40: Reflecting on the Past to Prepare for the Future. *Structure* 20, 391–
489 396.
- 490
- 491

492 Figures and Legends

493

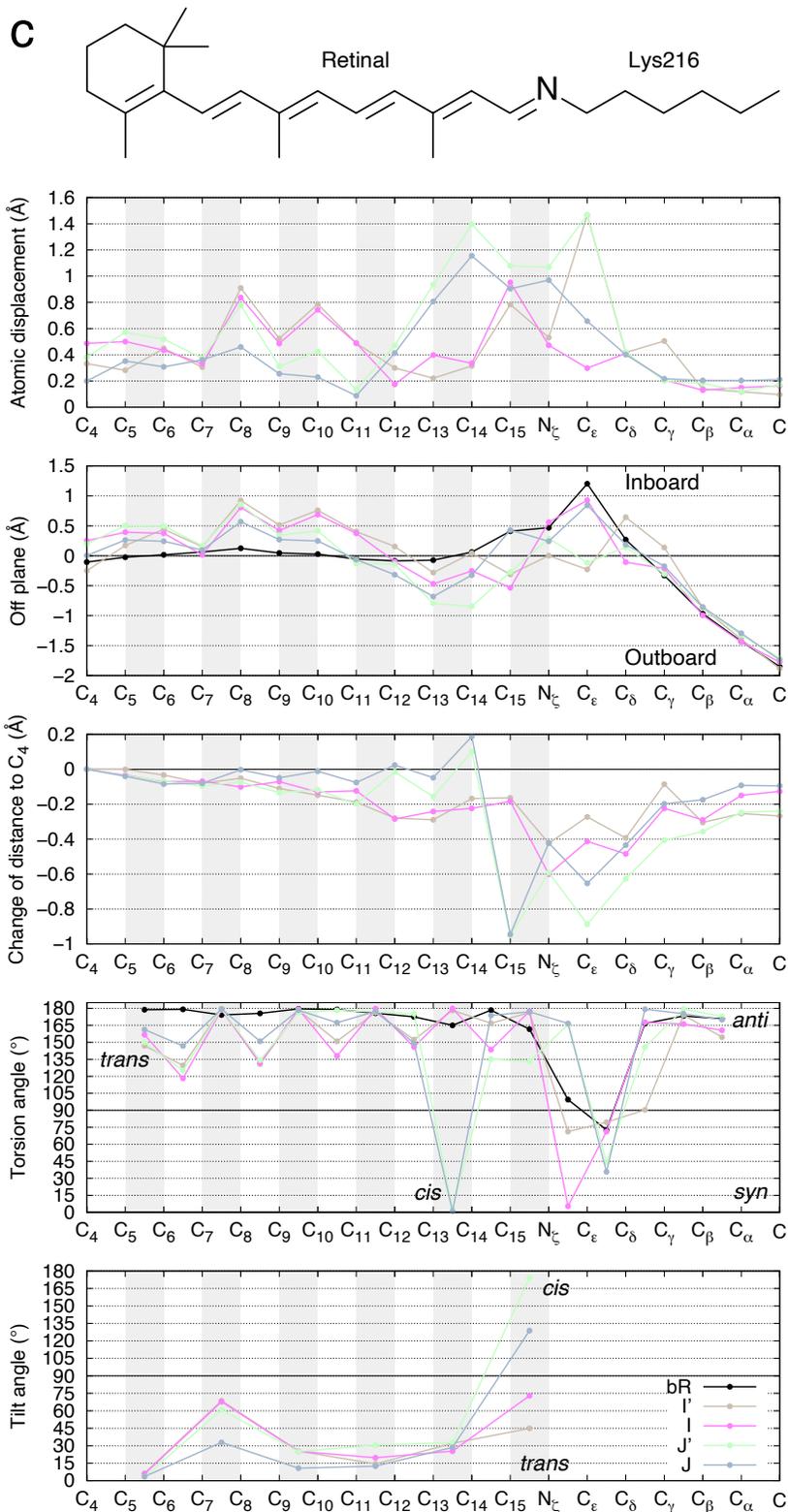
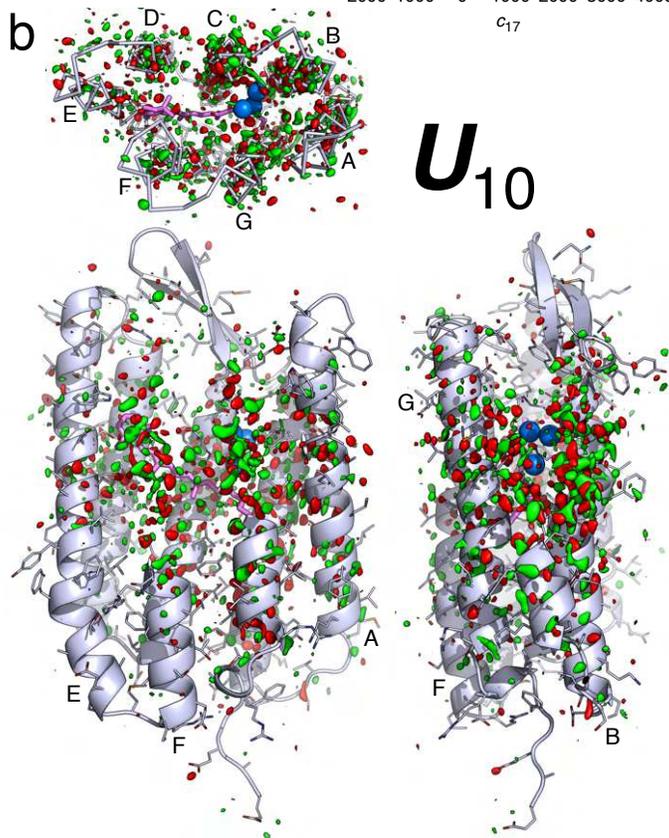
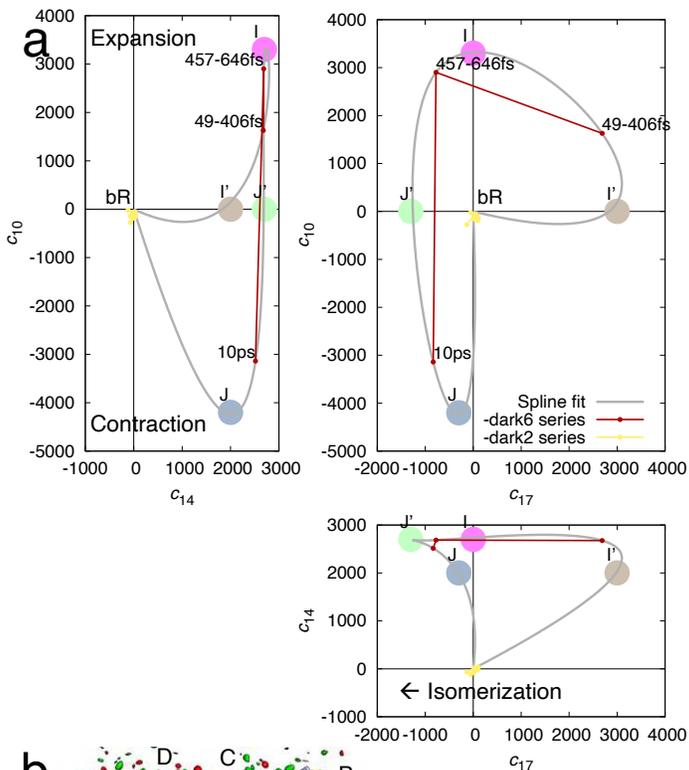


495

496 Figure 1. Shortened retinal in S-shape since earliest intermediate I' . (a) Cross sections
 497 of component map U_{14} . The middle cross section is an integration ± 0.2 Å around the
 498 surface through the retinal. The top cross section is an integration 1.2-1.8 Å outboard
 499 from the retinal surface and the bottom one is an integration 0.8-1.2 Å inboard. See Fig.
 500 S1 for definitions of inboard, outboard, and other orientations in bR molecule. Green
 501 and red indicate electron density gain and loss, respectively. Nearly the entire retinal is
 502 in negative densities. The proximal segment and three waters are in intense negative

503 densities. On the other hand, strong positive densities flank the proximal and distal
504 segments from the outboard and inboard, respectively. Such signal distribution results
505 in the S-shaped retinal by the refinement shown in (d). (b) Cross sections of component
506 map \mathbf{U}_{17} . The middle cross section is an integration $\pm 0.2 \text{ \AA}$ around the surface through
507 the retinal. The top panel is an integration $0.5\text{-}0.9 \text{ \AA}$ outboard and the bottom is an
508 integration $0.8\text{-}1.2 \text{ \AA}$ inboard. Negative and positive densities flank the retinal from the
509 outboard and inboard, respectively. (c) Difference map of $I' - \text{bR}$ reconstituted from \mathbf{U}_{14}
510 and \mathbf{U}_{17} (a and b). The map is contoured at $\pm 3\sigma$ in green and red mesh, respectively.
511 The opposite displacements of the distal and proximal segments of the retinal are
512 obvious. Extensive signals indicate changes in the water network and Asp85 and 212.
513 (d) Refined retinal conformation in beige overlaid on the resting state in white. This
514 view is orthographical to (c). The marked distances from C_3 to N_ζ show a shortened
515 retinal creased into an S-shape. C_{20} methyl group is tilted 33° toward outboard from its
516 resting state bR. Wat402 remains in H-bonds with both Asp85 and 212.
517

Ren: Isomerization sampling

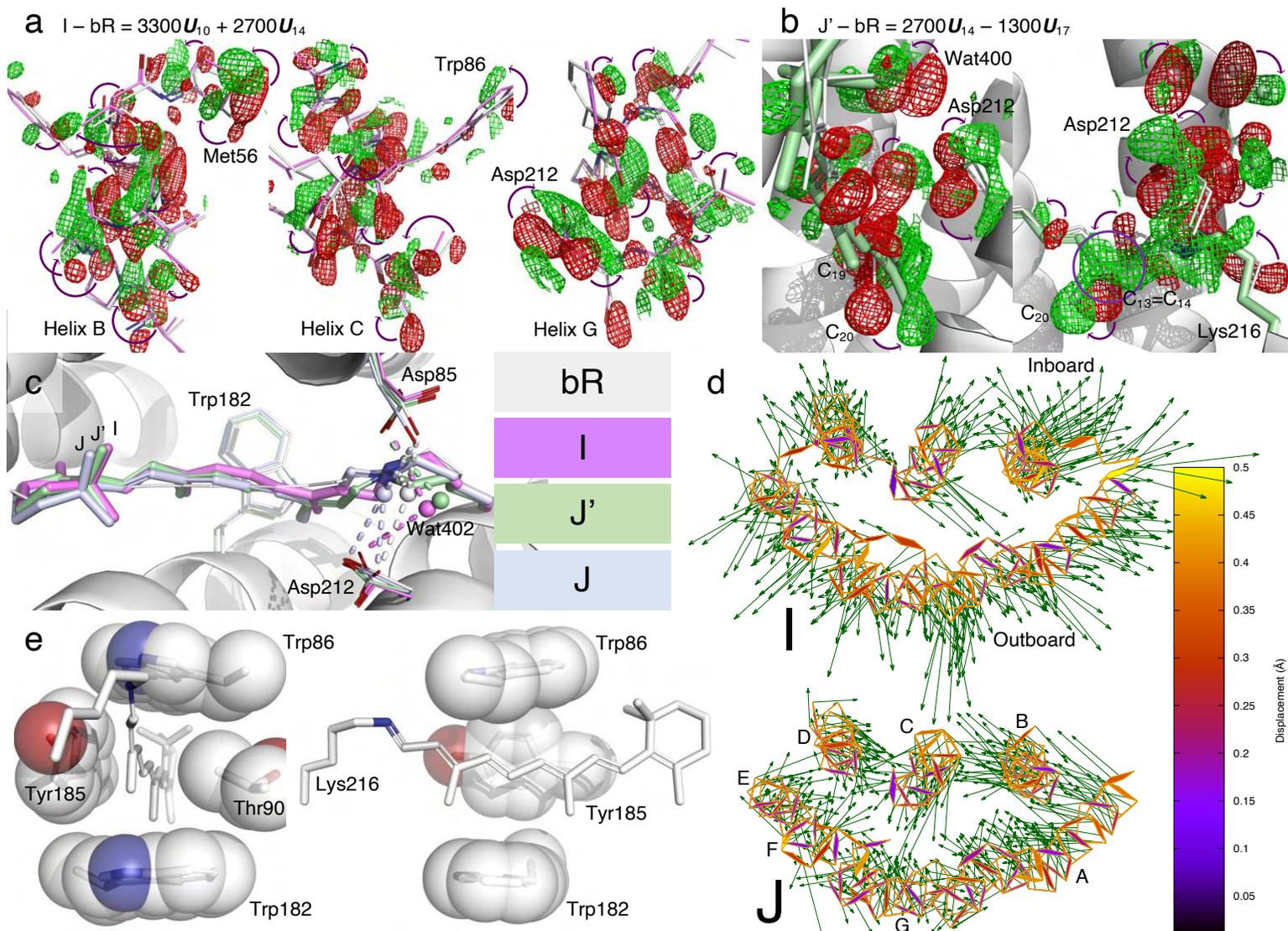


520 Figure 2. Early intermediates identified in SVD space. (a) Multi-dimensional spaces of
 521 SVD. The SVD analysis of difference Fourier maps at short delays ≤ 10 ps results in
 522 time-dependent coefficients $c_k(t)$, where $k = 1, 2, \dots$, each corresponding to a time-
 523 independent components \mathbf{U}_k . Each raw difference map at a time delay t can be closely
 524 represented by a linear combination of these components, $c_1(t)\mathbf{U}_1 + c_2(t)\mathbf{U}_2 + \dots$, that is
 525 called a reconstituted difference map. Each of these components \mathbf{U}_k and the
 526 reconstituted difference maps can be rendered in the same way as an observed
 527 difference map. The coefficient set $c_k(t)$ is therefore a trace of the photocycle trajectory,
 528 when these time-dependent functions are plotted in a multi-dimensional space or
 529 plotted together against the common variable t . Coefficients corresponding to
 530 components \mathbf{U}_{10} , \mathbf{U}_{14} , and \mathbf{U}_{17} are plotted in three orthographical views. Three time
 531 points from Nogly et al. in red contain \mathbf{U}_{14} equally. These time points vary in \mathbf{U}_{10} and
 532 \mathbf{U}_{17} . Datasets from Kovacs et al. in yellow do not carry any of these signals, therefore
 533 cluster near the origin. The component map of \mathbf{U}_{10} is displayed in (b) and \mathbf{U}_{14} is
 534 displayed in Figs. 1a and S5. \mathbf{U}_{17} is displayed in Fig. 1b. Several apices of the spline
 535 fitting are chosen as the potential pure states of I' , I , J' , and J marked by large dots. This
 536 choice is only an approximate due to the insufficient number of time points observed.
 537 (b) Component map \mathbf{U}_{10} . The main chain and side chains of the protein are rendered
 538 with ribbons and sticks, respectively. The retinal and Lys216 are in purple sticks.
 539 Several key waters are in blue spheres. Parts of the structure are omitted to reveal more
 540 of the interior. The map is contoured at $\pm 3\sigma$ in green and red, respectively. Three
 541 orthographical views of \mathbf{U}_{10} clearly show that the signals are distributed around the
 542 middle segment of the molecule and taper off to both CP and EC surfaces. The signals
 543 also concentrate at all seven helices. (c) Conformational parameters calculated from the
 544 refined chromophore. The chemical structure of the chromophore on top is aligned to
 545 the horizontal axis. Double bonds are shaded in gray. Atomic displacements of each
 546 intermediate from the resting state show greater changes in the proximal segment (top
 547 panel). A plane is least-squares fitted to C_4 through C_{14} of the resting state. The
 548 distances of all atoms to this plane in the inboard and outboard directions show the
 549 curvature of the chromophore. The creased retinal in early intermediates and the
 550 inboard protruding corner at C_6 in the resting state are clearly shown (2nd panel).
 551 Distances to atom C_4 are calculated for all refined chromophores. Changes in these
 552 distances with respect to the resting state show the shortened chromophore in I' and I .
 553 Once isomerization to 13-*cis* occurs, the segment from C_{15} to C_8 around the SB becomes

554 significantly closer to the β -ionone ring due to the Coulomb attraction force, while the
555 distal segment of the retinal from C₁₄ and beyond stretches (3rd panel). The torsion
556 angles of single and double bonds quantify *anti/syn* or *trans/cis* for the ground state and
557 all intermediates (4th panel). Only a single bond can be twisted with its torsion angle
558 near 90°. A twisted double bond would be energetically costly. Each double bond is
559 least-squares fitted with a plane. The interplanar angle between a double bond and the
560 corresponding one in the ground state measures the local tilting of the retinal (bottom
561 panel).

562

563



565
 566 Figure 3. Intermediates I, J', J, and retinal binding pocket expansion and contraction.
 567 (a) Reconstituted difference map I – bR from U_{10} and U_{14} (Figs. 1a, 2ab, and S5). The
 568 map is contoured at $\pm 2.5\sigma$ in green and red mesh, respectively. The difference map at
 569 three middle segments of helices B, C, and G show main chain displacements toward
 570 inboard or outboard as indicated by the arrows marking the negative and positive pairs
 571 of densities. These difference densities are the direct evidence of the expansion of the
 572 retinal binding pocket. The refined structure of I is in purple and the resting state is in
 573 white. (b) Reconstituted difference map J' – bR from U_{14} and U_{17} (Figs. 1ab, 2a, and S5).
 574 The map is contoured at $\pm 3.5\sigma$ in green and red mesh, respectively. These difference
 575 densities are the direct evidence of isomerization at hundreds of fs. The refined
 576 structure of J' in 13-*cis* is in green. (c) The refined 13-*cis* retinal conformation compared

577 with the resting state in white. I, J', and J are in purple, green, and bluish gray,
578 respectively. The creased S-shape is easing gradually (Fig. 2c 2nd panel). (d) The refined
579 structures of I and J compared with the resting state viewed along the trimer three-fold
580 axis from the EC side. Atomic displacements in the main chain from bR to I and J are
581 color coded and marked by arrows with lengths 20× of the actual displacements. All
582 seven helices in I move away from the center except a small segment in helix C showing
583 an expansion of the retinal binding pocket (top panel). However, all seven helices in J
584 move closer to one another showing a contraction with respect to the resting state bR
585 (bottom panel). This contraction is much more significant if compared directly with the
586 expanded I state. (e) Two orthographical views of the retinal tightly boxed at its middle
587 segment. The closest contact is Thr90 and Tyr185 on the inboard and outboard sides of
588 the retinal plane, respectively. The minimum distance between them is $7.0 \text{ \AA} = 4r_C + 0.2$
589 \AA , where $r_C = 1.7 \text{ \AA}$ is the van der Waals radius of C. See also (Kandori, 2015).
590
591

592 **Photoinduced Isomerization Sampling of Retinal in** 593 **Bacteriorhodopsin**

594

595 Zhong Ren

596

597 Department of Chemistry, University of Illinois at Chicago, Chicago, IL 60607, USA

598 Renz Research, Inc., Westmont, IL 60559, USA

599

600 zren@uic.edu

601 ORCID 0000-0001-7098-3127

602

603 **Methods**

604 From the outset, the key presumption is that every crystallographic dataset, at a given
605 temperature and a given time delay after the triggering of a photochemical reaction,
606 captures a mixture of unknown number of intermediate species at unknown fractions.
607 Needless to say, all structures of the intermediates are also unknown except the
608 structure at the ground state that has been determined and well refined by static
609 crystallography. A simultaneous solution of all these unknowns requires multiple
610 datasets that are collected at various temperatures or time delays so that a common set
611 of intermediate structures are present in these datasets with variable ratios. If the
612 number of available datasets is far greater than the number of unknowns, a linear
613 system can be established to overdetermine the unknowns with the necessary
614 stereochemical restraints (Ren et al., 2013). The analytical methods used in this work to
615 achieve such overdetermination have been incrementally developed in the past years
616 and recently applied to another joint analysis of the datasets of carbonmonoxy
617 myoglobin (Ren, 2019). Time-resolved datasets collected with ultrashort pulses from an
618 X-ray free electron laser were successfully analyzed by these methods to visualize
619 electron density components that reveal transient heating, *3d* electrons of the heme iron,
620 and global vibrational motions. This analytical strategy is recapped below.

621

622 The methodological advance in this work is the refinement of each pure
623 intermediate structure that has been deconvoluted from multiple mixtures. Structure
624 factor amplitudes of a single conformation free of heterogeneity are overdetermined.

625 Given the deconvoluted structure factor amplitude set of a pure state, the standard
626 structural refinement software with the built-in stereochemical constraints is taken full
627 advantage of, e.g. PHENIX (Adams et al., 2010; Liebschner et al., 2019). In case that the
628 computed deconvolution has not achieved a single pure structural species, the
629 structural refinement is expected to make such indication.

630

631 *Difference Fourier maps*

632 A difference Fourier map is synthesized from a Fourier coefficient set of $F_{\text{light}} - F_{\text{reference}}$
633 with the best available phase set, often from the ground state structure. Before Fourier
634 synthesis, F_{light} and $F_{\text{reference}}$ must be properly scaled to the same level so that the
635 distribution of difference values is centered at zero and not skewed either way. A
636 weighting scheme proven effective assumes that a greater amplitude of a difference
637 Fourier coefficient $F_{\text{light}} - F_{\text{reference}}$ is more likely caused by noise than by signal (Ren et al.,
638 2001, 2013; Šrajer et al., 2001; Ursby and Bourgeois, 1997). Both the dark and light
639 datasets can serve as a reference in difference maps. If a light dataset at a certain delay
640 is chosen as a reference, the difference map shows the changes since that delay time but
641 not the changes prior to that delay. However, both the dark and light datasets must be
642 collected in the same experiment. A cross reference from a different experimental
643 setting usually causes large systematic errors in the difference map that would swamp
644 the desired signals. Each difference map is masked 3.5 Å around the entire molecule of
645 bacteriorhodopsin (bR). No lipid density is analyzed.

646

647 *Meta-analysis of protein structures*

648 Structural meta-analysis based on singular value decomposition (SVD) has been
649 conducted in two forms. In one of them, an interatomic distance matrix is calculated
650 from each protein structure in a related collection. SVD of a data matrix consists of
651 these distance matrices enables a large-scale joint structural comparison but requires no
652 structural alignment (Ren, 2013a, 2013b, 2016). In the second form, SVD is performed
653 on a data matrix of electron densities of related protein structures (Ren, 2019; Ren et al.,
654 2013; Schmidt et al., 2003, 2010). Both difference electron density maps that require a
655 reference dataset from an isomorphous crystal form and simulated annealing omit
656 maps that do not require the same unit cell and space group of the crystals are possible
657 choices in a structural meta-analysis (Ren, 2019; Ren et al., 2013). The interatomic
658 distances or the electron densities that SVD is performed on are called core data. Each

659 distance matrix or electron density map is associated with some metadata that describe
660 the experimental conditions under which the core data are obtained, such as
661 temperature, pH, light illumination, time delay, mutation, etc. These metadata do not
662 enter the SVD procedure. However, they play important role in the subsequent
663 interpretation of the SVD result. This computational method of structural analysis takes
664 advantage of a mathematical, yet practical, definition of conformational space with
665 limited dimensionality (Ren, 2013a). Each experimentally determined structure is a
666 snapshot of the protein structure. A large number of such snapshots taken under a
667 variety of experimental conditions, the metadata, would collectively provide a survey of
668 the accessible conformational space of the protein structure and reveal its reaction
669 trajectory. Such joint analytical strategy would not be effective in early years when far
670 fewer protein structures were determined to atomic resolution. Recent rapid growth in
671 protein crystallography, such as in structural genomics (Chandonia and Brenner, 2006,
672 2012) and in serial crystallography (Glynn and Rodriguez, 2019; Schaffer et al., 2021),
673 has supplied the necessarily wide sampling of protein structures for a joint analytical
674 strategy to come of age. The vacancies or gaps in a conformational space between well-
675 populated conformational clusters often correspond to less stable transient states whose
676 conformations are difficult to capture, if not impossible. These conformations are often
677 key to mechanistic understanding and could be explored by a back calculation based on
678 molecular distance geometry (Ren, 2013a, 2016), the chief computational algorithm in
679 nucleic magnetic resonance spectroscopy (NMR), and by a structure refinement based
680 on reconstituted dataset, a major methodological advance in this work (see below).
681 These structures refined to atomic resolution against reconstituted datasets may reveal
682 short-lived intermediate conformation hard to be captured experimentally.
683 Unfortunately, a protein structure refined against a reconstituted dataset currently
684 cannot be recognized by the Protein Data Bank (PDB). Because crystallographic
685 refinement of a macromolecular structure is narrowly defined as a correspondence from
686 one dataset to one structure. A never-observed dataset reconstituted from a collection
687 of experimental datasets does not match the well-established crystallographic template
688 of PDB; let alone a refinement of crystal structure with the NMR algorithm.

689

690 A distance matrix contains M pairwise interatomic distances of a structure in the
691 form of Cartesian coordinates of all observed atoms. An everyday example of distance
692 matrix is an intercity mileage chart appended to the road atlas. Differences in the

693 molecular orientation, choice of origin, and crystal lattice among all experimentally
 694 determined structures have no contribution to the distance matrices. Due to its
 695 symmetry, only the lower triangle is necessary. A far more intimate examination of
 696 protein structures in PDB is a direct analysis of their electron density maps instead of
 697 the atomic coordinates. M such (difference) electron densities, often called voxels in
 698 computer graphics, are selected by a mask of interest. In the case of difference maps,
 699 only the best refined protein structure in the entire collection supplies a phase set for
 700 Fourier synthesis of electron density maps. This best structure is often the ground state
 701 structure determined by static crystallography. Other refined atomic coordinates from
 702 the PDB entries are not considered in the meta-analysis. That is to say, a meta-analysis
 703 of difference electron density maps starts from the X-ray diffraction data archived in
 704 PDB rather than the atomic coordinates interpreted from the diffraction data, which
 705 removes any potential model bias.

706

707 *Singular value decomposition of (difference) electron density maps*

708 An electron density map, particularly a difference map as emphasized here, consists of
 709 density values on an array of grid points within a mask of interest. All M grid points in
 710 a three-dimensional map can be serialized into a one-dimensional sequence of density
 711 values according to a specific protocol. It is not important what the protocol is as long
 712 as a consistent protocol is used to serialize all maps of the same grid setting and size,
 713 and a reverse protocol is available to erect a three-dimensional map from a sequence of
 714 M densities. Therefore, a set of N serialized maps, also known as vectors in linear
 715 algebra, can fill the columns of a data matrix \mathbf{A} with no specific order, so that the width
 716 of \mathbf{A} is N columns, and the length is M rows. Often, $M \gg N$, thus \mathbf{A} is an elongated
 717 matrix. If a consistent protocol of serialization is used, the corresponding voxel in all N
 718 maps occupies a single row of matrix \mathbf{A} . This strict correspondence in a row of matrix
 719 \mathbf{A} is important. Changes of the density values in a row from one structure to another
 720 are due to either signals, systematic errors, or noises. Although the order of columns in
 721 matrix \mathbf{A} is unimportant, needless to say, the metadata associated with each column
 722 must remain in good bookkeeping.

723

724 SVD of the data matrix \mathbf{A} results in $\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^T$, also known as matrix factorization.
 725 Matrix \mathbf{U} has the same shape as \mathbf{A} , that is, N columns and M rows. The N columns
 726 contain decomposed basis components \mathbf{U}_k , known as left singular vectors of M items,

727 where $k = 1, 2, \dots, N$. Therefore, each component \mathbf{U}_k can be erected using the reverse
 728 protocol to form a three-dimensional map. This decomposed elemental map can be
 729 presented in the same way as the original maps, for example, rendered in molecular
 730 graphics software such as Coot and PyMol. It is worth noting that these decomposed
 731 elemental maps or map components \mathbf{U}_k are independent of any metadata. That is to
 732 say, these components remain constant when the metadata vary. Since each left
 733 singular vector \mathbf{U}_k has a unit length due to the orthonormal property of SVD (see
 734 below), that is, $|\mathbf{U}_k| = 1$, the root mean squares (rms) of the items in a left singular
 735 vector is $1/\sqrt{M}$ that measures the quadratic mean of the items.

736
 737 The second matrix \mathbf{W} is a square matrix that contains all zeros except for N positive
 738 values on its major diagonal, known as singular values w_k . The magnitude of w_k is
 739 considered as a weight or significance of its corresponding component \mathbf{U}_k . The third
 740 matrix \mathbf{V} is also a square matrix of $N \times N$. Each column of \mathbf{V} or row of its transpose \mathbf{V}^T ,
 741 known as a right singular vector \mathbf{V}_k , contains the relative compositions of \mathbf{U}_k in each of
 742 the N original maps. Therefore, each right singular vector \mathbf{V}_k can be considered as a
 743 function of the metadata. Right singular vectors also have the same unit length, that is,
 744 $|\mathbf{V}_k| = 1$. Effectively, SVD separates the constant components independent of the
 745 metadata from the compositions that depend on the metadata.

746
 747 A singular triplet denotes 1) a decomposed component \mathbf{U}_k , 2) its singular value w_k ,
 748 and 3) the composition function \mathbf{V}_k . Singular triplets are often sorted in a descending
 749 order of their singular values w_k . Only a small number of n significant singular triplets
 750 identified by the greatest singular values w_1 through w_n can be used in a linear
 751 combination to reconstitute a set of composite maps that closely resemble the original
 752 ones in matrix \mathbf{A} , where $n < N$. For example, the original map in the i th column of
 753 matrix \mathbf{A} under a certain experimental condition can be closely represented by the i th
 754 composite map $w_1 v_{1i} \mathbf{U}_1 + w_2 v_{2i} \mathbf{U}_2 + \dots + w_n v_{ni} \mathbf{U}_n$, where (v_{1i}, v_{2i}, \dots) is from the i th row of
 755 matrix \mathbf{V} . The coefficient set for the linear combination is redefined here as $c_{ki} =$
 756 $w_k v_{ki} / \sqrt{M}$. The rms of the density values in a map component, or the average magnitude
 757 measured by the quadratic mean, acts as a constant scale factor that resets the modified
 758 coefficients c_{ki} back to the original scale of the core data, such as Å for distance matrices
 759 and $e/\text{Å}^3$ for electron density maps if these units are used in the original matrix \mathbf{A} .
 760 Practically, an electron density value usually carries an arbitrary unit without a

761 calibration, which makes this scale factor unnecessary. In the linear combination $c_{1i}\mathbf{U}_1 +$
762 $c_{2i}\mathbf{U}_2 + \dots + c_{ni}\mathbf{U}_n$, each component \mathbf{U}_k is independent of the metadata while how much of
763 each component is required for the approximation, that is, c_{ki} , depends on the metadata.
764

765 Excluding the components after \mathbf{U}_n in this approximation is based on an assumption
766 that the singular values after w_n are very small relative to those from w_1 through w_n . As
767 a result, the structural information evenly distributed in all N original maps is
768 effectively concentrated into a far fewer number of n significant components, known as
769 information concentration or dimension reduction. On the other hand, the trailing
770 components in matrix \mathbf{U} contain inconsistent fluctuations and random noises.
771 Excluding these components effectively rejects noises (Schmidt et al., 2003). The least-
772 squares property of SVD guarantees that the rejected trailing components sums up to
773 the least squares of the discrepancies between the original core data and the
774 approximation using the accepted components.

775
776 However, no clear boundary is guaranteed between signals, systematic errors, and
777 noises. Systematic errors could be more significant than the desired signals. Therefore,
778 excluding some components from 1 through n is also possible. If systematic errors are
779 correctly identified, the reconstituted map without these significant components would
780 no longer carry the systematic errors.

781
782 *The orthonormal property of SVD*

783 The solution set of SVD must guarantee that the columns in \mathbf{U} and \mathbf{V} , the left and right
784 singular vectors \mathbf{U}_k and \mathbf{V}_k , are orthonormal, that is, $\mathbf{U}_h \bullet \mathbf{U}_k = \mathbf{V}_h \bullet \mathbf{V}_k = 0$ (ortho) and $\mathbf{U}_k \bullet \mathbf{U}_k$
785 $= \mathbf{V}_k \bullet \mathbf{V}_k = 1$ (normal), where $h \neq k$ but both are from 1 to N . The orthonormal property
786 also holds for the row vectors. As a result, each component \mathbf{U}_k is independent of the
787 other components. In other words, a component cannot be represented by a linear
788 combination of any other components. However, two physical or chemical parameters
789 in the metadata, such as temperature and pH, may cause different changes to a
790 structure. These changes are not necessarily orthogonal. They could exhibit some
791 correlation. Therefore, the decomposed components \mathbf{U}_k not necessarily represent any
792 physically or chemically meaningful changes (see below).

793

794 Due to the orthonormal property of SVD, an N -dimensional Euclidean space is
795 established, and the first n dimensions define its most significant subspace. Each
796 coefficient set $c_i = (c_{1i}, c_{2i}, \dots, c_{ni})$ of the i th composite map is located in this n -dimensional
797 subspace. All coefficient sets for $i = 1, 2, \dots, N$ in different linear combinations to
798 approximate the N original maps in a least-squares sense can be represented by N
799 points or vectors c_1, c_2, \dots, c_N in the Euclidean subspace. This n -dimensional subspace is
800 essentially the conformational space as surveyed by the jointly analyzed core data. The
801 conformational space is presented as scatter plots with each captured structure
802 represented as a dot located at a position determined by the coefficient set c_i of the i th
803 observed map. When the subspace has greater dimensionality than two, multiple two-
804 dimensional orthographical projections of the subspace are presented, such as Fig. 2a.
805 These scatter plots are highly informative to reveal the relationship between the
806 (difference) electron density maps and their metadata.

807
808 If two coefficient sets $c_i \approx c_j$, they are located close to each other in the
809 conformational space. Therefore, these two structures i and j share two similar
810 conformations. Two structures located far apart from each other in the conformational
811 space are dissimilar in their conformations, and distinct in the compositions of the map
812 components. A reaction trajectory emerges in this conformational space if the temporal
813 order of the core data is experimentally determined (Fig. 2a). Otherwise, an order could
814 be assigned to these structures based on an assumed smoothness of conformational
815 changes along a reaction trajectory (Ren, 2013a, 2013b, 2016). Causation and
816 consequence of structural motions could be revealed from the order of the structures in
817 a series, which may further lead to structural mechanism. In addition, an off-trajectory
818 location in the conformational space or a location between two clusters of observed
819 structures represents a structure in a unique conformation that has never been
820 experimentally captured. Such a hypothetical structure can be refined against a
821 reconstituted distance matrix using molecular distance geometry (Ren, 2013a, 2013b,
822 2016) or a reconstituted electron density map with the method proposed below.

823

824 *Rotation in SVD space*

825 Dimension reduction is indeed effective in meta-analysis of protein structures when
826 many datasets are evaluated at the same time. However, the default solution set of SVD
827 carries complicated physical and chemical meanings that are not immediately obvious.

828 The interpretation of a basis component \mathbf{U}_k , that is, “what-does-it-mean”, requires a
 829 clear demonstration of the relationship between the core data and their metadata. The
 830 outcome of SVD does not guarantee any physical meaning in a basis component.
 831 Therefore, SVD alone provides no direct answer to “what-does-it-mean”, thus its
 832 usefulness is very limited to merely a mathematical construction. However, the
 833 factorized set of matrices \mathbf{U} , \mathbf{W} , and \mathbf{V} from SVD is not a unique solution. That is to say,
 834 they are not the only solution to factorize matrix \mathbf{A} . Therefore, it is very important to
 835 find one or more alternative solution sets that are physically meaningful to elucidate a
 836 structural interpretation. The concept of a rotation after SVD was introduced by Henry
 837 & Hofrichter (Henry and Hofrichter, 1992). But they suggested a protocol that fails to
 838 preserve the orthonormal and least-squares properties of SVD. The rotation protocol
 839 suggested by Ren incorporates the metadata into the analysis and combines with SVD
 840 of the core data. This rotation achieves a numerical deconvolution of multiple physical
 841 and chemical factors after a pure mathematical decomposition, and therefore, provides
 842 a route to answer the question of “what-does-it-mean” (Ren, 2019). This rotation shall
 843 not be confused with a rotation in the three-dimensional real space, in which a
 844 molecular structure resides.

845
 846 A rotation in the n -dimensional Euclidean subspace is necessary to change the
 847 perspective before a clear relationship emerges to elucidate scientific findings. It is
 848 shown below that two linear combinations are identical before and after a rotation
 849 applied to both the basis components and their coefficients in a two-dimensional
 850 subspace of h and k . That is,

$$c_h \mathbf{U}_h + c_k \mathbf{U}_k = f_h \mathbf{R}_h + f_k \mathbf{R}_k, \quad (1)$$

852
 853 where c_h and c_k are the coefficients of the basis components \mathbf{U}_h and \mathbf{U}_k before the
 854 rotation; and f_h and f_k are the coefficients of the rotated basis components \mathbf{R}_h and \mathbf{R}_k ,
 855 respectively. The same Givens rotation of an angle θ is applied to both the components
 856 and their coefficients:

$$\begin{cases} \mathbf{R}_h = \mathbf{U}_h \cos\theta - \mathbf{U}_k \sin\theta; \\ \mathbf{R}_k = \mathbf{U}_h \sin\theta + \mathbf{U}_k \cos\theta. \end{cases} \quad (2)$$

858

859 Obviously, the rotated components \mathbf{R}_h and \mathbf{R}_k remain mutually orthonormal and
 860 orthonormal to other components. And

$$861 \quad \begin{cases} f_h = s_h t_h = c_h \cos \theta - c_k \sin \theta; \\ f_k = s_k t_k = c_h \sin \theta + c_k \cos \theta. \end{cases} \quad (3)$$

862
 863 Here $s_{h|k} = \sqrt{\sum f_{h|k}^2}$ are the singular values that replace w_h and w_k , respectively, after the
 864 rotation. They may increase or decrease compared to the original singular values so
 865 that the descending order of the singular values no longer holds. $\mathbf{T}_{h|k} = (t_{h|k1}, t_{h|k2}, \dots,$
 866 $t_{h|kN}) = (f_{h|k1}, f_{h|k2}, \dots, f_{h|kN})/s_{h|k}$ are the right singular vectors that replace \mathbf{V}_h and \mathbf{V}_k ,
 867 respectively. \mathbf{T}_h and \mathbf{T}_k remain mutually orthonormal after the rotation and
 868 orthonormal to other right singular vectors that are not involved in the rotation.

869
 870 To prove Eq. 1, Eqs. 2 and 3 are combined and expanded. All cross terms of sine and
 871 cosine are self-canceled:

$$872 \quad \begin{aligned} 873 \quad f_h \mathbf{R}_h + f_k \mathbf{R}_k &= (c_h \cos \theta - c_k \sin \theta)(\mathbf{U}_h \cos \theta - \mathbf{U}_k \sin \theta) + (c_h \sin \theta + c_k \cos \theta)(\mathbf{U}_h \sin \theta + \mathbf{U}_k \cos \theta) \\ 874 \quad &= c_h \mathbf{U}_h \cos^2 \theta + c_k \mathbf{U}_k \sin^2 \theta + c_h \mathbf{U}_h \sin^2 \theta + c_k \mathbf{U}_k \cos^2 \theta \pm c_h \mathbf{U}_k \sin \theta \cos \theta \pm c_k \mathbf{U}_h \sin \theta \cos \theta \\ 875 \quad &= c_h \mathbf{U}_h (\cos^2 \theta + \sin^2 \theta) + c_k \mathbf{U}_k (\sin^2 \theta + \cos^2 \theta) \\ 876 \quad &= c_h \mathbf{U}_h + c_k \mathbf{U}_k \end{aligned}$$

877
 878 A rotation in two-dimensional subspace of h and k has no effect in other dimensions,
 879 as the orthonormal property of SVD guarantees. Multiple steps of rotations can be
 880 carried out in many two-dimensional subspaces consecutively to achieve a multi-
 881 dimensional rotation. A new solution set derived from a rotation retains the
 882 orthonormal property of SVD. The rotation in the Euclidean subspace established by
 883 SVD does not change the comparison among the core data of protein structures. Rather
 884 it converts one solution set $\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^T$ to other alternative solutions $\mathbf{A} = \mathbf{R}\mathbf{S}\mathbf{T}^T$ so that
 885 an appropriate perspective can be found to elucidate the relationship between the core
 886 data and metadata clearly and concisely.

887
 888 For example, if one physical parameter could be reoriented along a single dimension
 889 k but not involving other dimensions by a rotation, it can be convincingly shown that
 890 the left singular vector \mathbf{U}_k of this dimension illustrates the structural impact by this

891 physical parameter. Before this rotation, the same physical parameter may appear to
892 cause structural variations along several dimensions, which leads to a difficult
893 interpretation. Would a proper rotation establish a one-on-one correspondence from all
894 physical or chemical parameters to all the dimensions? It depends on whether each
895 parameter induces an orthogonal structural change, that is, whether structural
896 responses to different parameters are independent or correlated among one another. If
897 structural changes are indeed orthogonal, it should be possible to find a proper rotation
898 to cleanly separate them in different dimensions. Otherwise, two different rotations are
899 necessary to isolate two correlated responses, but one at a time.

900
901 For another example, if the observed core datasets form two clusters in the
902 conformational space, a rotation would be desirable to separate these clusters along a
903 single dimension k but to align these clusters along other dimensions. Therefore, the
904 component \mathbf{U}_k is clearly due to the structural transition from one cluster to the other.
905 Without a proper rotation, the difference between these clusters could be complicated
906 with multiple dimensions involved. A deterministic solution depends on whether a
907 clear correlation exists between the core data and metadata. A proper rotation may
908 require a user decision. A wrong choice of rotation may select a viewpoint that hinders
909 a concise conclusion. However, it would not alter the shape of the reaction trajectory,
910 nor create or eliminate an intrinsic structural feature. A wrong choice of rotation cannot
911 eliminate the fact that a large gap exists between two clusters of observed core datasets
912 except that these clusters are not obvious from that viewpoint. A different rotation may
913 reorient the perspective along another direction. But the structural conclusion would be
914 equivalent. See example of before and after a rotation in (Ren, 2016).

915
916 This rotation procedure finally connects the core crystallographic datasets to the
917 metadata of experimental conditions and accomplishes the deconvolution of physical or
918 chemical factors that are not always orthogonal to one another after a mathematical
919 decomposition. SVD analysis presented in this paper employs rotations extensively
920 except that no distinction is made in the symbols of components and coefficients before
921 and after a rotation except in this section. This method is widely applicable in large-
922 scale structural comparisons. Furthermore, Ren rotation after SVD is not limited to
923 crystallography and may impact other fields wherever SVD is used. For example, SVD
924 is frequently applied to spectroscopic data, images, and genetic sequence data.

925

926 *Structural refinement against reconstituted dataset*

927 The linear combination $\Delta\rho(t) = f_1(t)\mathbf{R}_1 + f_2(t)\mathbf{R}_2 + \dots + f_n(t)\mathbf{R}_n$ after a rotation reconstitutes
928 one of the observed difference maps at a specific time point t . This time-dependent
929 difference map depicts an ever-evolving mixture of many excited species. A
930 reconstituted difference map $\Delta\rho(E)$ for a time-independent, pure, excited species $E =$
931 intermediate I' , I , J' , and J deconvoluted from many mixtures would take the same form
932 except that only one or very few coefficients remain nonzero if a proper rotation has
933 been found (Table S2). In order to take advantage of the mature refinement software for
934 macromolecular structures with extensive stereochemical restraints, a set of structure
935 factor amplitudes is needed. Therefore, it is necessary to reconstitute a set of structure
936 factor amplitudes that would produce the target difference map $\Delta\rho(E)$ based on a
937 known structure at the ground state. First, an electron density map of the structure at
938 the ground state is calculated. This calculated map is used as a base map. Second, this
939 base map of the ground state is combined with the positive and negative densities in the
940 target difference map $\Delta\rho(E)$ so that the electron densities at the ground state are skewed
941 toward the intermediate state. Third, structure factors are calculated from the
942 combined map. Finally, the phase set of the calculated structure factors is discarded,
943 and the amplitudes are used to refine a single conformation of the intermediate species
944 E that $\Delta\rho(E)$ represents.

945

946 This protocol following the SVD and Ren rotation of components achieves a
947 refinement of a pure structural species without the need of alternative conformations.
948 Several points are noteworthy. First, the minimization protocol in this refinement is
949 performed against a numerically reconstituted amplitude set that has never been
950 directly measured from a crystal. This reconstituted dataset could be considered as an
951 extrapolated dataset "on steroids" if compared to the traditional extrapolation of small
952 differences, such as, the Fourier coefficient set to calculate a 3Fo-2Fc map, a technique
953 often used to overcome a partial occupancy of an intermediate structure. An
954 extrapolation of small differences is not directly observed either but computed by an
955 exaggeration of the observed difference based on an assumption that the intermediate
956 state is partially occupied, such as the doubling of the observed difference in 3Fo-2Fc =
957 $F_o + 2(F_o - F_c)$. In contrast to the conventional technique of extrapolation, the
958 deconvolution method applied here is an interpolation among many experimental

959 datasets rather than an extrapolation. Secondly, the deconvolution is a simultaneous
960 solution of multiple intermediate states mixed together instead of solving a single
961 excited state.

962
963 Second, a map calculated from the ground state structure is chosen as the base map
964 instead of an experimental map such as F_o or $2F_o-F_c$ map. If the second step of the
965 protocol is skipped, that is, no difference map is combined with the ground state map,
966 the refinement would result in an R factor of nearly zero, since the refinement is
967 essentially against the calculated structure factors (bR in Table S2). This is to say, the
968 residuals of the refinement are solely due to the difference component instead of the
969 base map. This is desirable since errors in the static structure of the ground state are
970 gauged during its own refinement. On the other hand, if an experimental map is
971 chosen as a base map, the refinement R factors would reflect errors in both the base map
972 and the difference map, which leads to a difficulty in an objective evaluation of this
973 refinement protocol.

974
975 Third, the combination of the base map and a difference map is intended to
976 represent a pure intermediate species. Therefore, alternative conformations in
977 structural refinement that model a mixture of species would defeat this purpose.
978 However, this combined map could be very noisy and may not represent a single
979 species without a proper rotation. This is particular the case, if the target difference
980 map $\Delta\rho$ is not derived from an SVD analysis and Ren rotation. The SVD analysis
981 identifies many density components that are inconsistent among all observed difference
982 maps and excludes them, which greatly reduces the noise content. Therefore, this
983 refinement protocol may not be very successful without an SVD analysis. Another
984 source of noise originates from the phase set of the structure factors. Prior to the
985 refinement of the intermediate structure, the phase set remains identical to that of the
986 ground state. This is far from the reality when an intermediate structure involves
987 widespread changes, such as those refined in this study. If the rotation after SVD is not
988 properly selected, the target difference map would remain as a mixture minus the
989 ground state. Therefore, the refinement of a single conformation would encounter
990 difficulty or significant residuals, as judged by the R factors, the residual map, and the
991 refined structure. A proper solution to this problem is a better SVD solution by Ren
992 rotation rather than alternative conformations. A successful refinement of near perfect

993 *trans* or *cis* double bonds is a good sign to indicate that the reconstituted amplitude set
994 after a rotation reflects a relatively homogeneous structure. If a double bond could not
995 be refined well to near perfect *trans* or *cis* configuration, the dataset of structure factor
996 amplitudes is likely from a mixture of heterogeneous configurations, which occurred
997 frequently in previous studies of bR and photoactive yellow protein (Jung et al., 2013;
998 Lanyi and Schobert, 2007; Nogly et al., 2018). It has been a great difficulty in
999 crystallographic refinement in general that a heterogeneous mixture of conformations
1000 cannot be unambiguously refined even with alternative conformations. This difficulty
1001 becomes more severe when a mixture involves more than two conformations or when
1002 some conformations are very minor.

1003
1004 Lastly, the refinement protocol proposed here could be carried out in the original
1005 unit cell and space group of the crystal at the ground state. However, this is not always
1006 applicable as the original goal of the meta-analysis is a joint examination of all available
1007 structures from a variety of crystal forms. It would be highly desirable to evaluate
1008 difference maps of the same or similar proteins from non-isomorphous crystals together
1009 by SVD. Alternatively, the refinement protocol could also be performed in the space
1010 group of P1 with a virtual unit cell large enough to hold the structure, which is the
1011 option in this study (Table S2). This is to say, the entire analysis of SVD-rotation-
1012 refinement presented here could be extracted and isolated from the original crystal
1013 lattices, which paves the way to future applications to structural data acquired by
1014 experimental techniques beyond crystallography, most attractively, to single particle
1015 reconstruction in cryo electron microscopy.

1016

1017

1018 **Supplementary Tables**

Table S1. Datasets analyzed in this work					
Publication	PDB	Label	Resolution	Main conclusions	New findings in this work
Nogly et al. Science 361, eaat0094, 2018	6g7h	dark6	1.5 Å	Retinal fully isomerizes by 10 ps. But the SB water dissociates earlier.	The short-delay datasets contribute to the structures of I' → I → J' → J. Photoisomerization in J'; retinal binding pocket expansion before 1 ps in I and contraction at 10 ps in J
	6g7i	49-406fs	1.9 Å		
	6g7j	457-646fs	1.9 Å		
	6g7k	10ps	1.9 Å		
Kovacs et al. Nat. Commun. 10, 3177, 2019	6g7l	8.33ms	1.9 Å	The exceedingly high power density of the pump laser causes two-photon absorption. Vibrational motions were observed.	The sub-ps datasets exhibit extensive vibrations at various frequencies. The vibrational signals are widespread over the entire bR molecule and not associated with any structural elements. Therefore, it is concluded that these global vibrations are intrinsic properties of bR induced by short laser pulses. The vibrational signals are more prominent under higher power density of the laser pulses. However, these vibrations are irrelevant to the light-driven proton pumping function of bR.
	6ga1	dark1	1.7 Å		
	6ga2	dark2	1.8 Å		
	6rmk	dark3	1.8 Å		
	6ga7	240fs	1.8 Å		
	6ga8	330fs	1.8 Å		
	6ga9	390fs	1.8 Å		
	6gaa	430fs	1.8 Å		
	6gab	460fs	1.8 Å		
	6gac	490fs	1.8 Å		
	6gad	530fs	1.8 Å		
	6gae	560fs	1.8 Å		
	6gaf	590fs	1.8 Å		
	6gag	630fs	1.8 Å		
6gah	680fs	1.8 Å			
6gai	740fs	1.8 Å			
6ga4	1ps	1.8 Å			
6ga5	3ps	1.9 Å			
6ga6	10ps	1.8 Å			
6ga3	33ms	2.1 Å			

1019

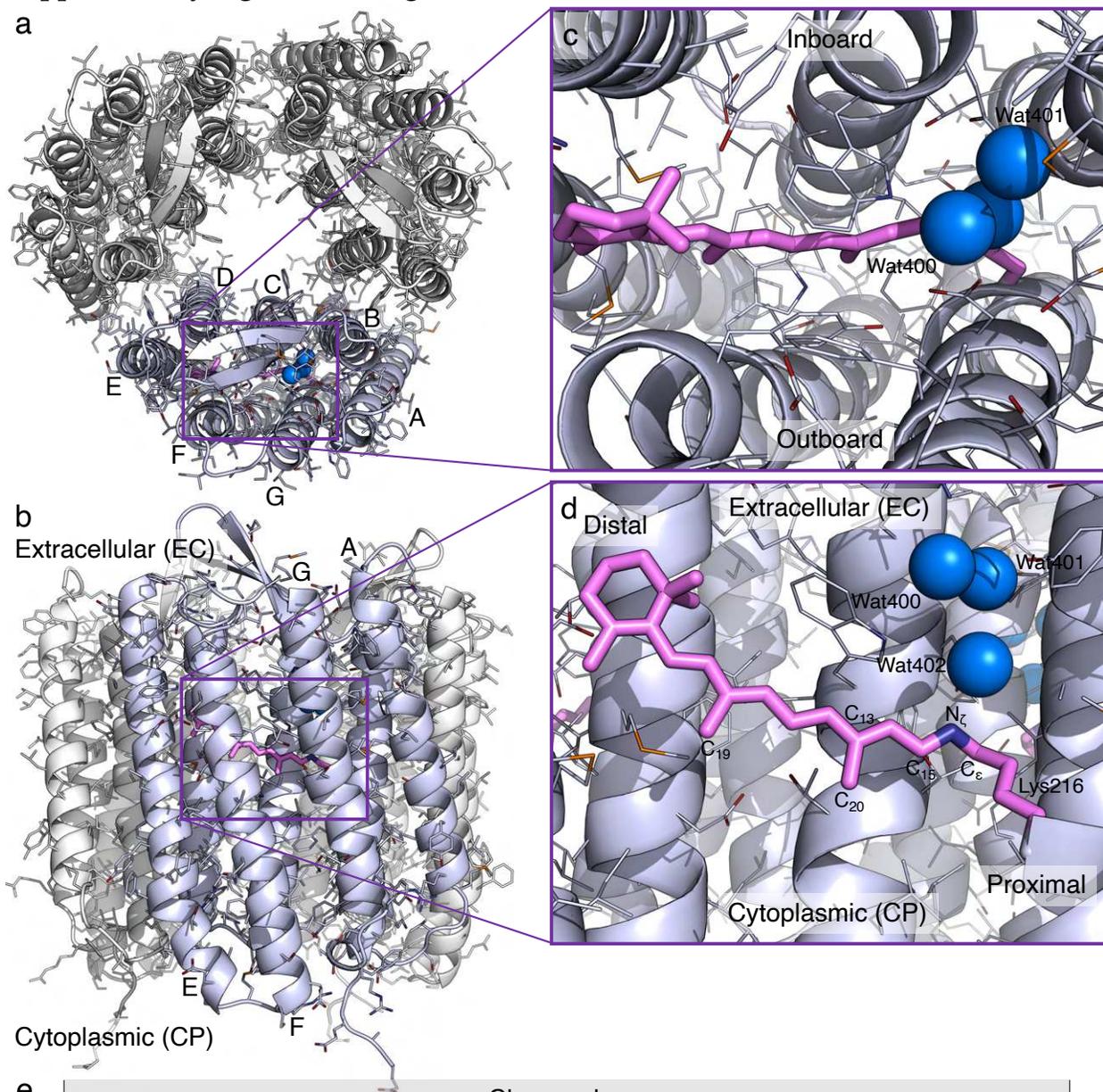
1020

Intermediate	bR	I'	I	J'	J
Time period	0-	< 50 fs	40-700 fs	0.5-2 ps	1-30 ps
<i>c</i> ₁₀	0	0	3,300	0	-4,200
Coefficient <i>c</i> ₁₄	0	2,000	2,700	2,700	2,000
<i>c</i> ₁₇	0	3,000	0	-1,300	-300
Starting model	PDB 6g7h				
Resolution range	50-2.1 Å				
Space group	P1				
Unit cell	$a = b = 62.32 \text{ Å}; c = 111.10 \text{ Å}; \alpha = \beta = 90^\circ; \text{ and } \gamma = 120^\circ$				
Unique reflections	80,354 in working set + 4,236 in test set = 84,590 total				
Completeness	95% in working set + 5% in test set = 100% reconstituted				
<i>R</i> (%)	1.8	29.4	31.0	29.1	30.0
<i>R</i> _{free} (%)	1.9	31.1	32.4	30.4	30.7
Refined content	230 protein residues + retinal + water molecules				
Number of atoms	1,798	1,795	1,798	1,796	1,795
Water molecules	8	5	8	6	5
RMSD bonds (Å)	0.005	0.009	0.009	0.009	0.009
RMSD angles (°)	0.793	1.206	1.105	1.085	1.068
Rama. favored (%)	98.7	96.5	95.6	96.1	96.5
Rama. outliers (%)	0.0	0.0	0.4	0.4	0.4
Clash score	4	9	5	4	6

1021

1022

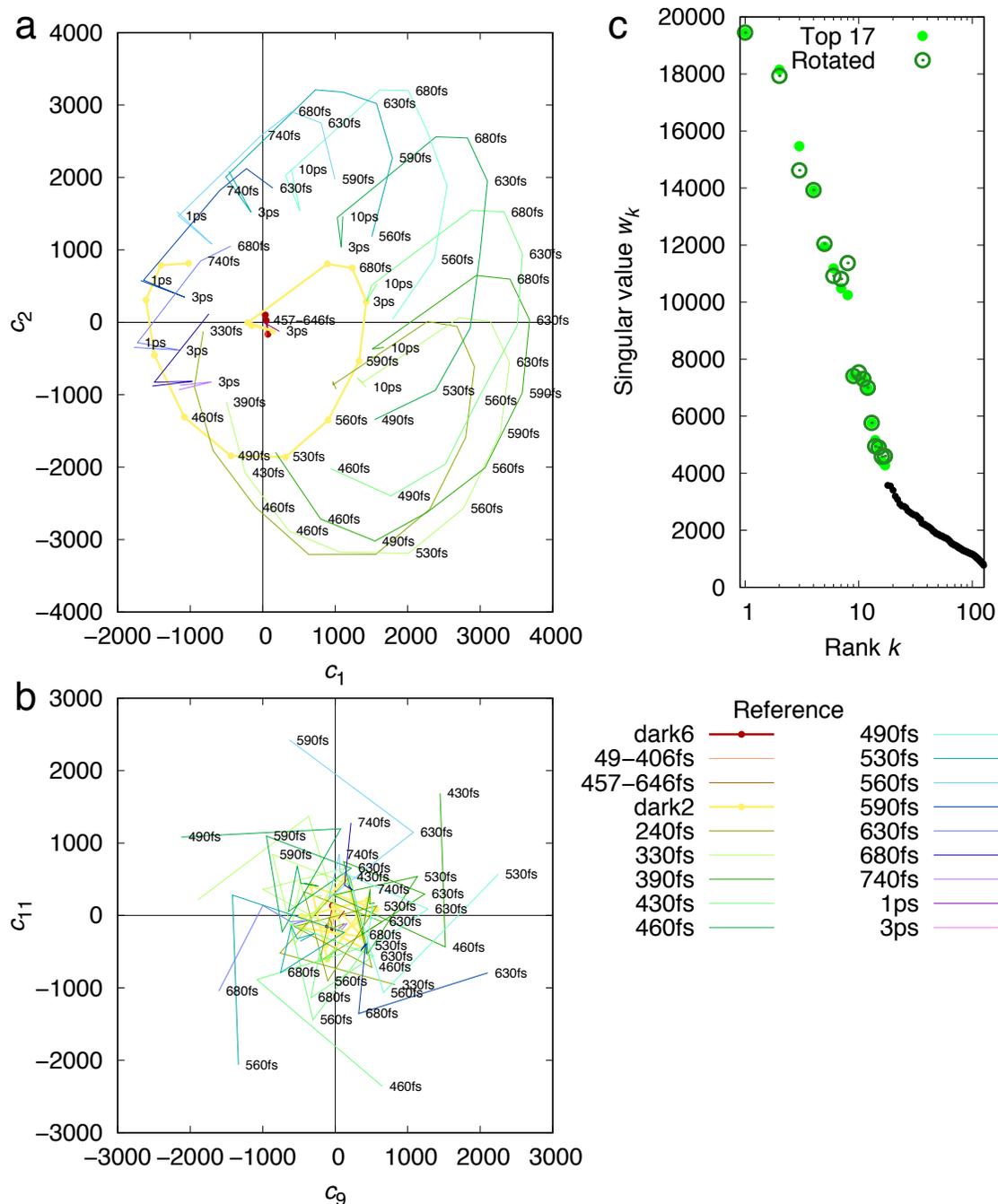
1023 Supplementary Figures and Legends



1024

1025 Figure S1. Orientations in bacteriorhodopsin. (a) Bacteriorhodopsin (bR) trimer viewed
1026 from the extracellular (EC) side along the three-fold axis. (b) An orthographical view to
1027 (a) looking from the outside of the trimer. (c and d) Two orthographical views of the
1028 retinal chromophore looking along the three-fold and normal to the three-fold axis. The
1029 plane of retinal is largely parallel to the three-fold axis. Therefore, two sides of the
1030 plane are called inboard and outboard with respect to the three-fold axis. The direction
1031 toward the anchor Lys216 is called proximal. The β -ionone ring direction is therefore
1032 distal. (e) Chemical structure of retinal incorporated to its anchor Lys216. The atom
1033 numbers and various segment names are marked.
1034

Ren: Isomerization sampling



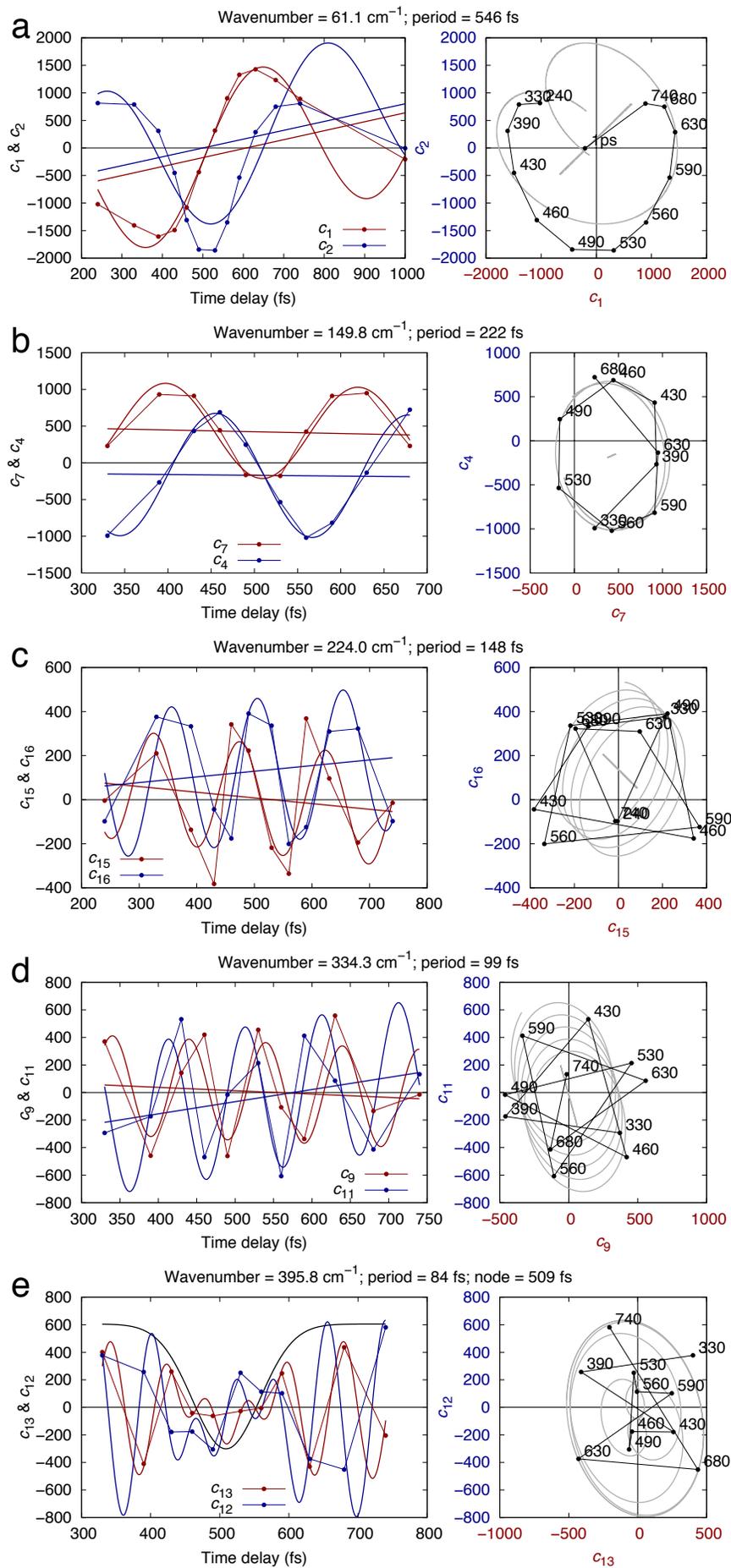
1035

1036

1037 Figure S2. SVD applied to difference Fourier maps. Difference Fourier maps at the
 1038 short delays $t \leq 10$ ps are decomposed into component maps. Each difference map at a
 1039 time delay t can be represented by a linear combination of these components, $c_1(t)\mathbf{U}_1 +$
 1040 $c_2(t)\mathbf{U}_2 + \dots$, where \mathbf{U}_k are the time-independent components and $c_k(t)$ are their
 1041 corresponding time-dependent coefficients (Methods). (a and b) Two example plots
 1042 show circular correlations between c_1 and c_2 , c_9 and c_{11} . These circular correlations

1043 indicate two-dimensional oscillations. Each colored trace represents difference maps in
1044 a time series calculated with a common reference. Those time series with a dark
1045 reference are plotted with thick lines. Other series are in thin lines. (c) Singular values
1046 before and after Ren rotation (Ren, 2016, 2019) (Methods). Singular values derived from
1047 SVD indicate the significance of the components. 17 of them stand out.
1048

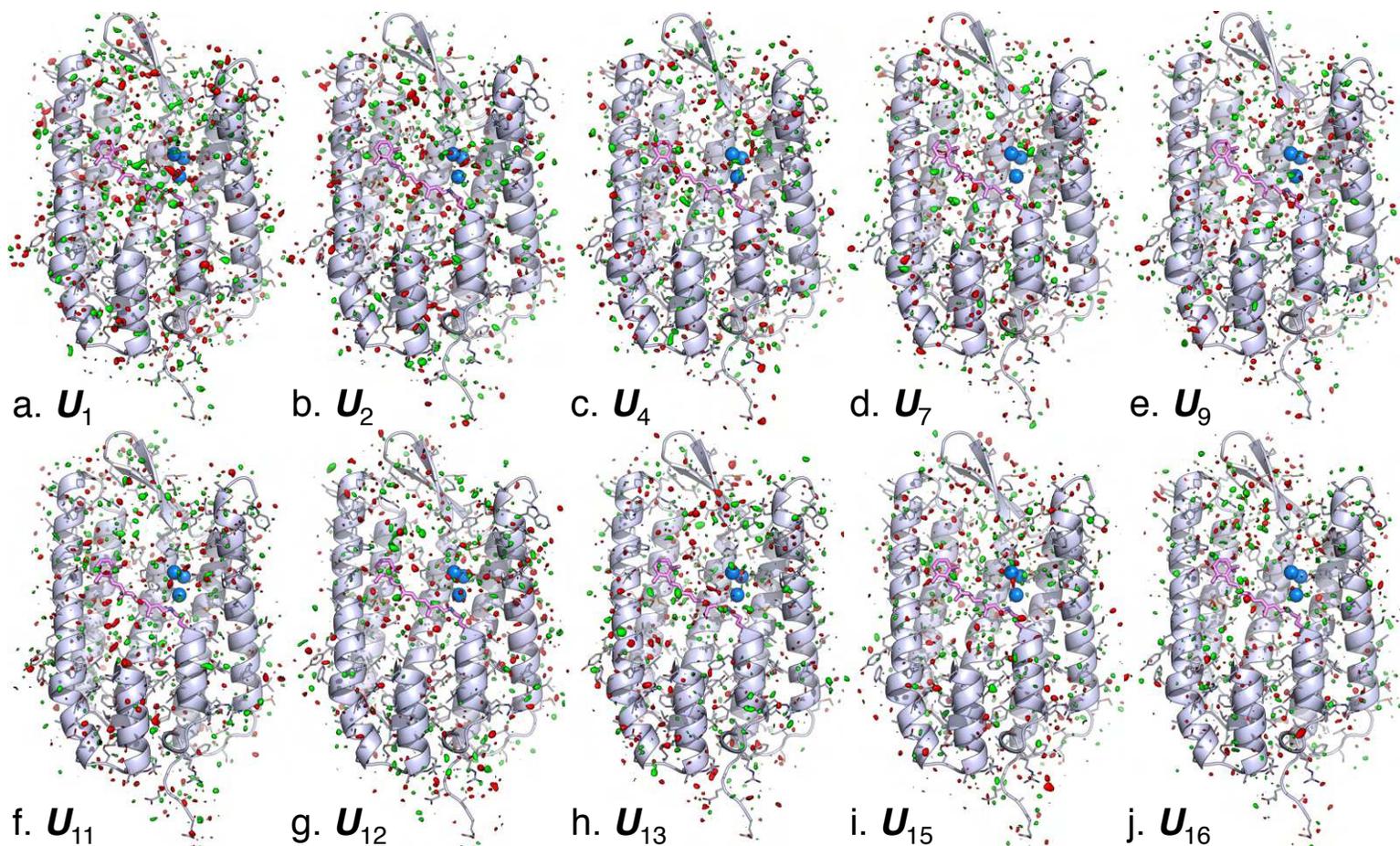
Ren: Isomerization sampling



1050 Figure S3. Oscillations of SVD components. The coefficients of ten components $c_1, c_2; c_4,$
1051 $c_7; c_{15}, c_{16}; c_9, c_{11};$ and c_{12}, c_{13} are found oscillating at frequencies ranging from 60 to 400
1052 cm^{-1} . Each pair of the coefficients oscillate at a common frequency. These frequencies
1053 are $61\pm 2, 150\pm 3, 224\pm 7, 334\pm 8,$ and $396\pm 3 \text{ cm}^{-1}$, respectively. These coefficients are
1054 plotted against the time delay t (left) and against each other in a pair (right). Each
1055 coefficient is fitted with a sine function around a straight baseline $c_k = a \sin\left(\frac{2\pi t}{T} + \varphi\right) +$
1056 $b + ct$. Both the fitted function and the baseline are plotted. The amplitude a for the
1057 last pair of coefficients c_{12} and c_{13} are replaced with a Gaussian function $a -$
1058 $\exp\left(-\frac{(t-t_0)^2}{\tau^2}\right)$ to implement a node at $t_0 = 509\pm 5 \text{ fs}$ (e).

1059

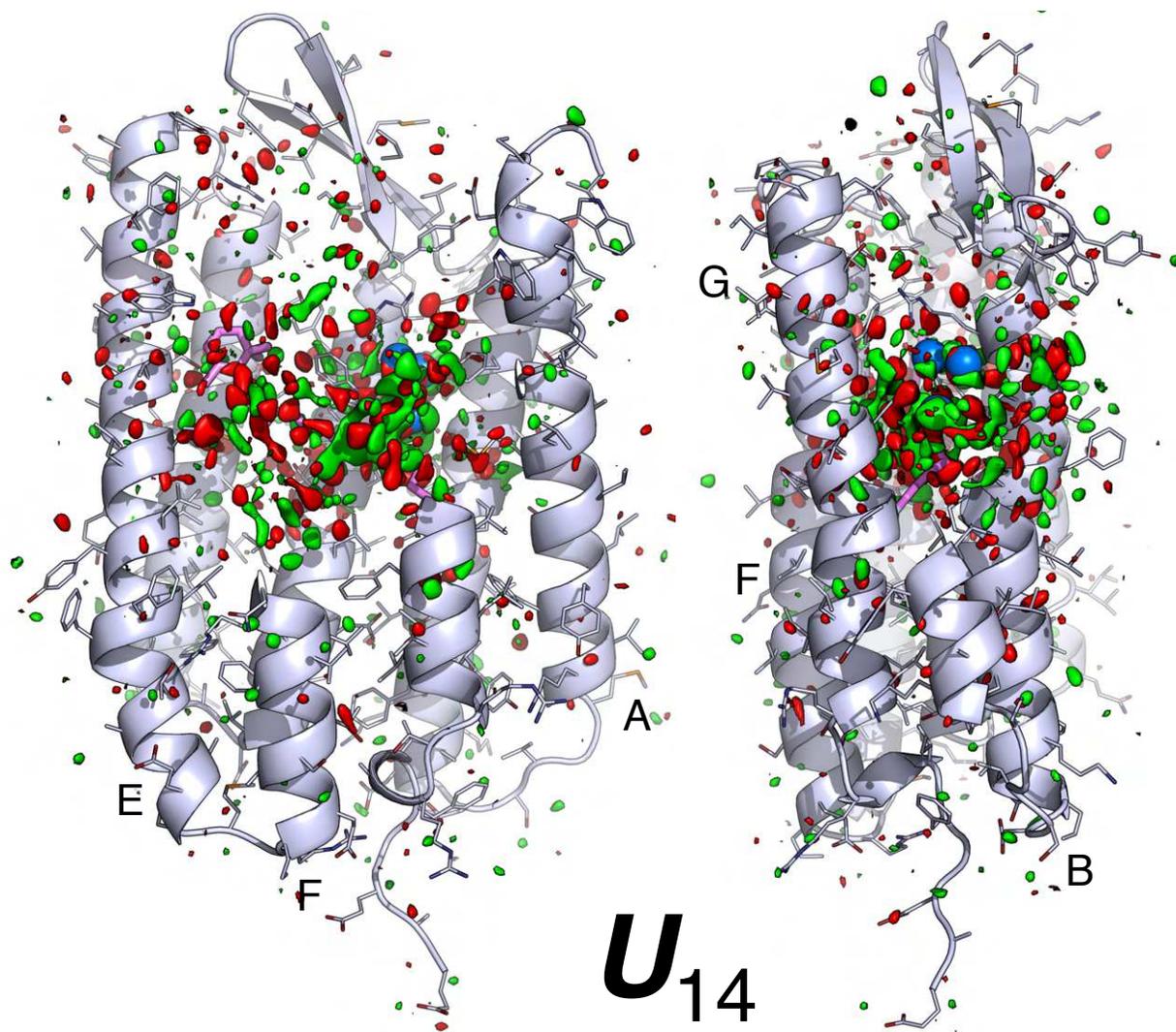
1060



1062

1063 Figure S4. No structural signal in oscillating components. Ten oscillating components
 1064 are contoured at $\pm 3\sigma$ in green and red, respectively. The main chain and side chains of
 1065 the protein are rendered with ribbon and sticks, respectively. The retinal and Lys216
 1066 are in purple sticks. Several key waters are in blue spheres. Parts of the structure are
 1067 omitted to reveal more of the interior. Despite that the time-dependent coefficients to
 1068 these components contain strong oscillatory signals (Figs. S2 and S3), these components
 1069 themselves display no obvious association with any structural features such as the
 1070 retinal or secondary structures. They are in stark contrast to the signal distributions of
 1071 the non-oscillating components (Figs. 1ab, 2b, and S5).

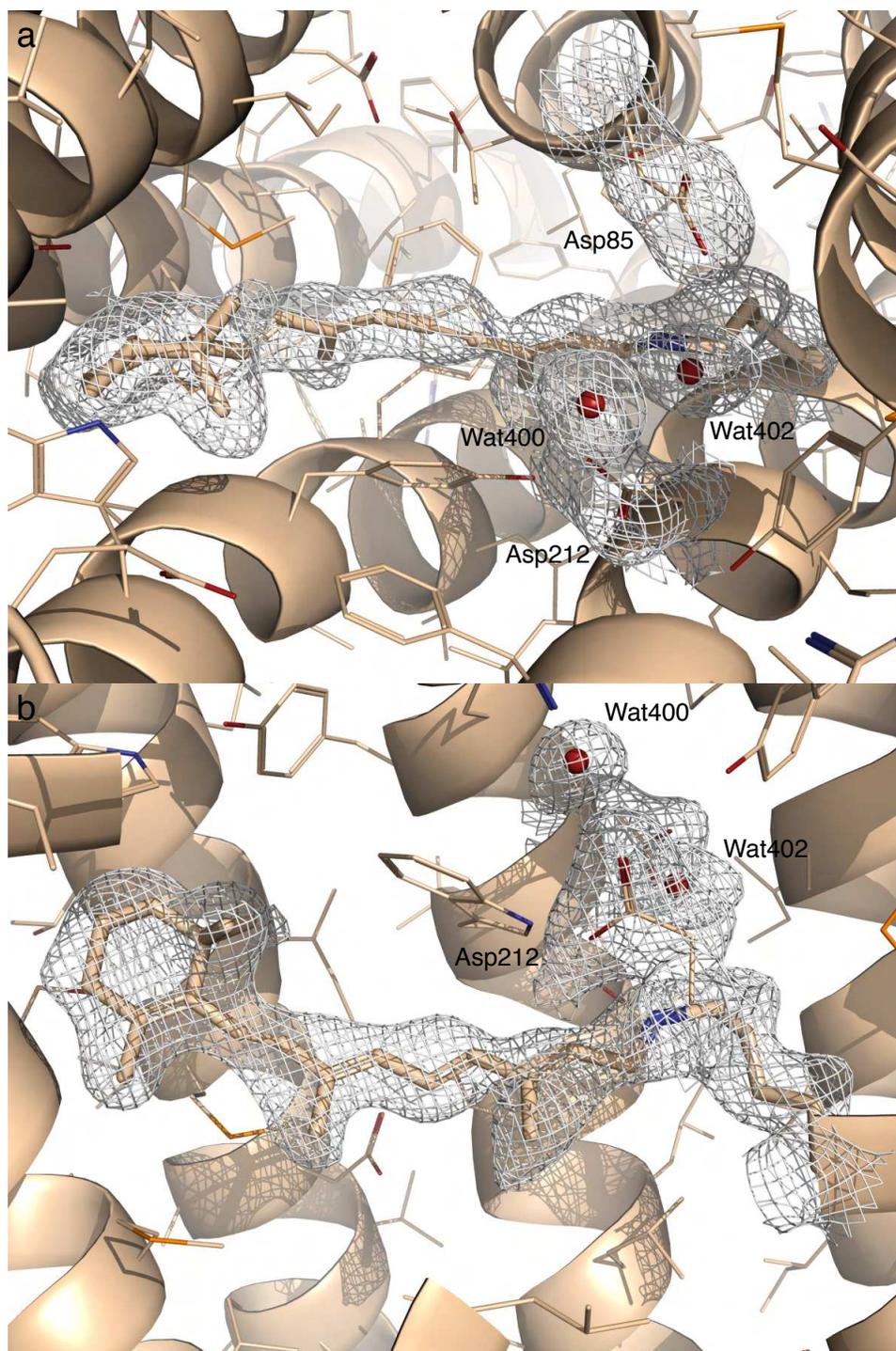
1072



1073
1074

1075 Figure S5. Two orthographical views of component map U_{14} . The main chain and side
1076 chains of the protein are rendered with ribbons and sticks, respectively. The retinal and
1077 Lys216 are in purple sticks. Several key waters are in blue spheres. Parts of the
1078 structure are omitted to reveal more of the interior. The map is contoured at $\pm 3\sigma$ in
1079 green and red, respectively. The signals are largely associated with the chromophore
1080 and its immediate vicinity.

1081



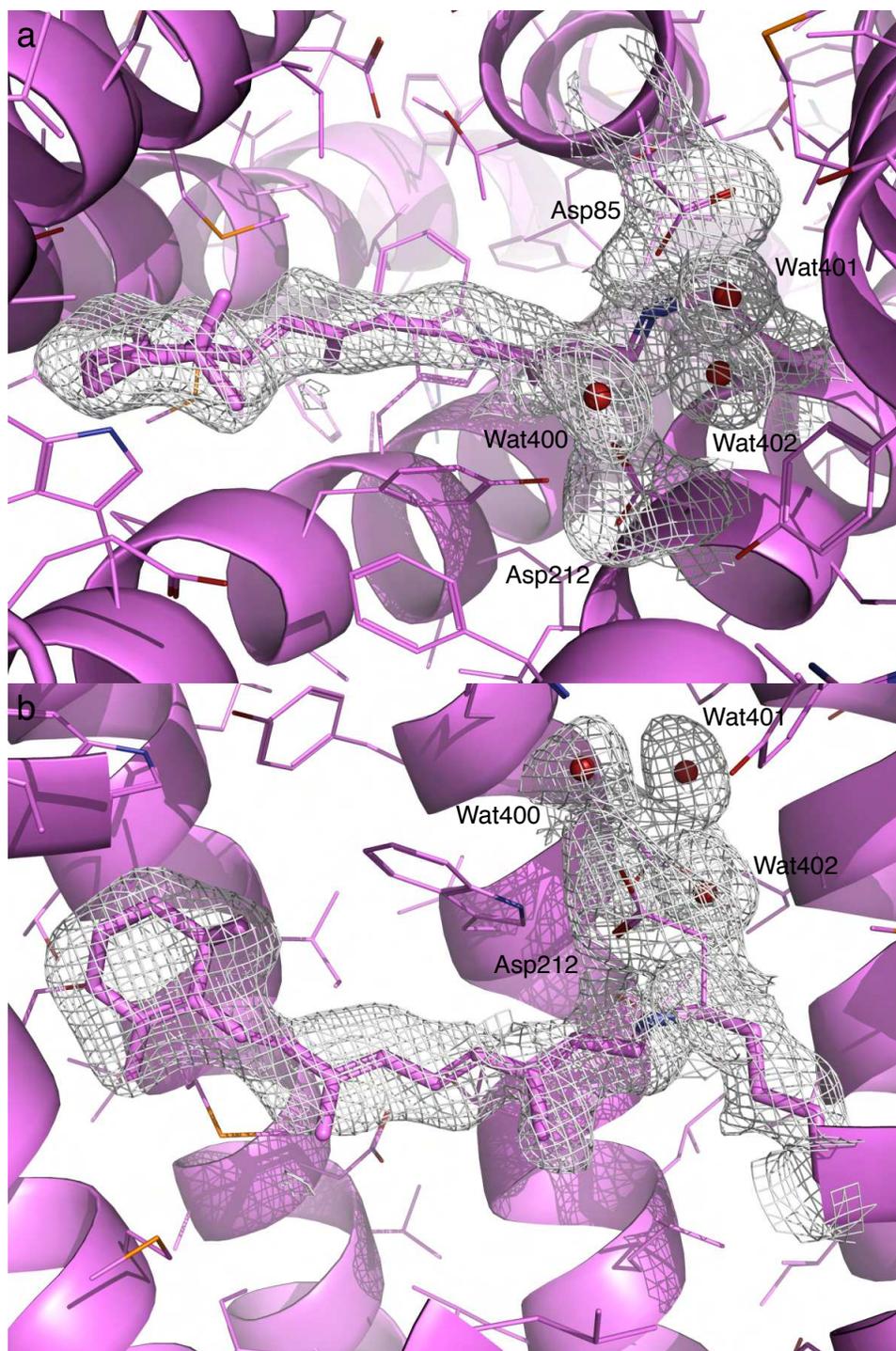
1082
1083

1084 Figure S6. Two orthographical views of the 2Fo-Fc map of I' contoured at 3.5 σ . Here Fo
1085 is the reconstituted structure factor amplitudes rather than observed amplitudes (Table
1086 S2). Fc is the structure factor amplitudes calculated from the refine structure (Methods).

1087 The same protocol applies to the Fourier synthesis of 2Fo-Fc maps of other
1088 intermediates (Figs. S7, S9, and S10).

1089

1090

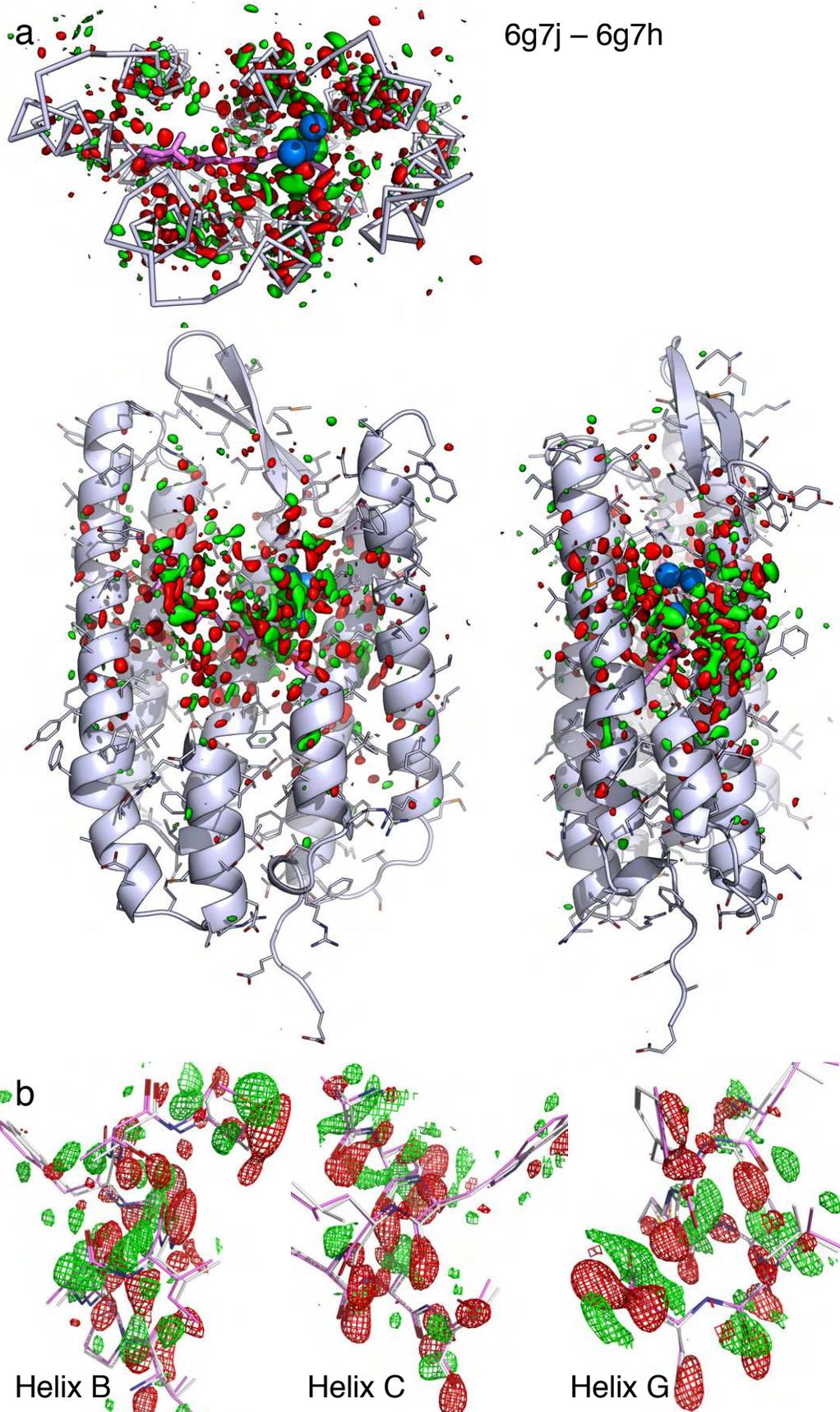


1091

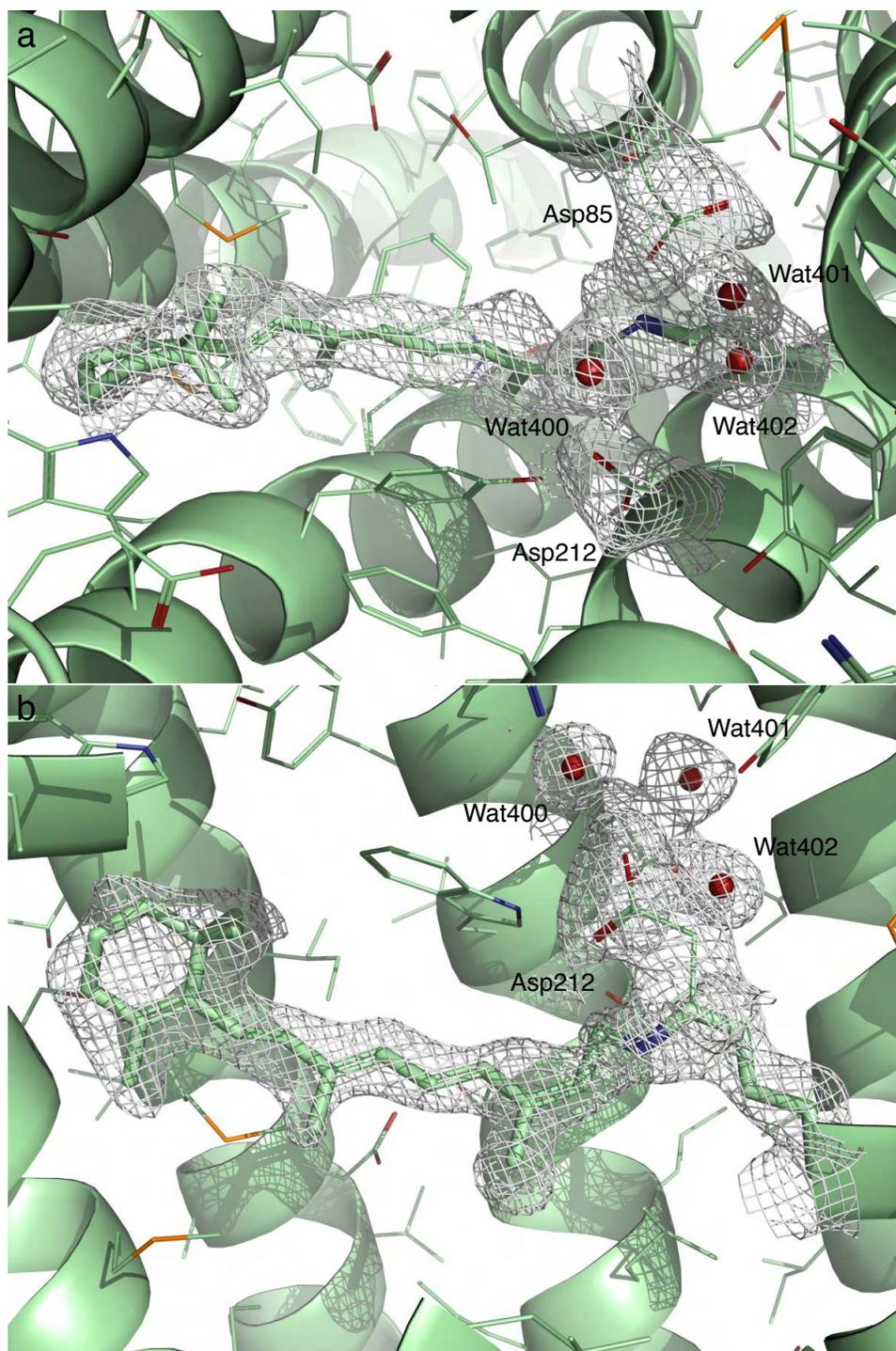
1092

1093 Figure S7. Two orthographical views of the 2Fo-Fc map of I contoured at 3σ. Here Fo is
1094 the reconstituted structure factor amplitudes rather than observed amplitudes (Table
1095 S2). Fc is the structure factor amplitudes calculated from the refine structure (Methods).

1096



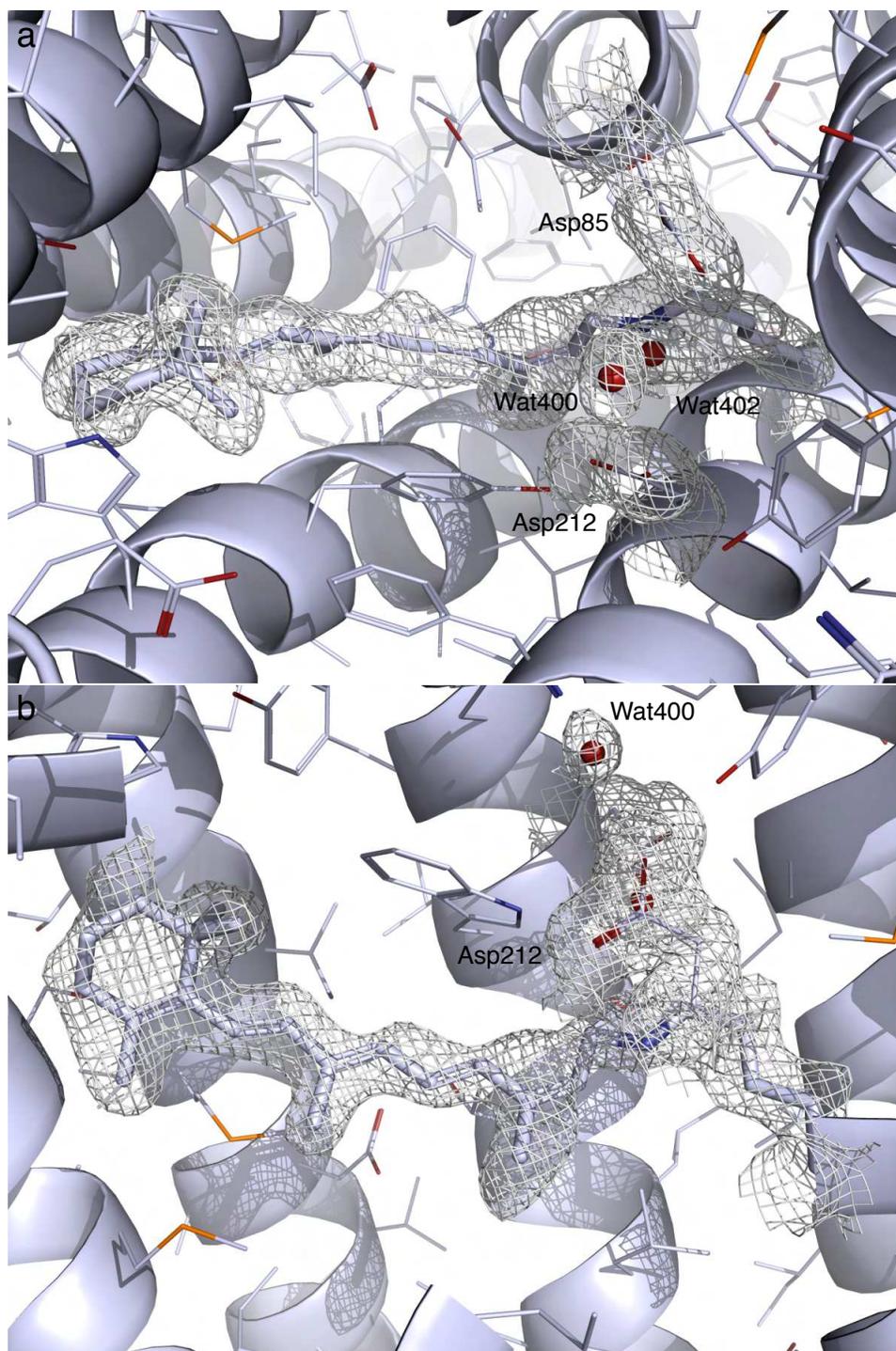
1098 Figure S8. Raw difference Fourier map at 457-646 fs. This difference Fourier map is
1099 calculated from the dataset 6g7j at the time point of 457-646 fs by subtracting the dark
1100 dataset 6g7h. The map is contoured at $\pm 3\sigma$ in green and red, respectively. This map is
1101 prior to SVD analysis. Compared with \mathbf{U}_{10} (Fig. 2b) and the reconstituted map (Fig. 3a),
1102 it is clear that this is the original source of the widespread signals except that the σ
1103 value of this map is higher than those after SVD. (a) The raw difference map contoured
1104 in the entire molecule shows the association of the signals with the structural elements
1105 at an excellent signal-to-noise ratio. (b) Details of the raw difference map show
1106 displacements of helices. The raw difference map is largely the same as the
1107 reconstituted map (Fig. 3a).
1108
1109



1110
1111

1112 Figure S9. Two orthographical views of the 2Fo-Fc map of J' contoured at 4σ . Here Fo
1113 is the reconstituted structure factor amplitudes rather than observed amplitudes (Table
1114 S2). Fc is the structure factor amplitudes calculated from the refine structure (Methods).

1115



1116
1117

1118 Figure S10. Two orthographical views of the 2Fo-Fc map of J contoured at 5σ. Here Fo
1119 is the reconstituted structure factor amplitudes rather than observed amplitudes (Table
1120 S2). Fc is the structure factor amplitudes calculated from the refine structure (Methods).

1121