Peer Review File

Manuscript Title: Few-fs resolution of a photoactive protein traversing a conical intersection

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee expertise:

Referee #1: ultrafast dynamics

Referee #2: TR-SFX

Referee #3: machine learning

Referees' comments:

Referee #1 (Remarks to the Author):

This manuscript presents very interesting results on the analysis of femtosecond x-ray serial crystallography data using geometric machine learning. Previously, the data was analyzed using standard methods by Pande et al (Ref 7 in the manuscript). The machine learning analysis reproduces the results of the previous analysis to a large extend, but with the impressive addition of 1 fs resolution, which is justified by the observation of an oscillation with 10 fs period in the signal. For the experimental conditions in which the data was taken, a 1 fs resolution would be truly impressive and unexpected, and thus needs strong supporting evidence beyond what is presented in the manuscript. A second point I would like to raise is that is seems difficult to connect the output of the machine learning algorithm with any physical properties of the system. I also have a few questions about the simulation, also explained in more detail below.

1. The evidence for the claim of 1 fs resolution is the observation of an oscillatory signal with 10 fs period. It is difficult to picture how this signal could be produced and measured if the system was pumped by a 140 fs laser pulse. The first question is, what is producing this signal? This should be discussed in the manuscript. The only thing that comes to mind that matches the period would be a C-H vibration. Then one needs to ask the question, does the excitation laser pulse have sufficient bandwidth to excite a coherent vibration with 10 fs period? No details are given on the exciting laser pulse in the manuscript. The authors should address this question. This would at least establish that such a fast signal could exist in the system.
The second question is whether the signal could be detected and not blurred out by the duration of the excitation laser pulse. Let us assume that a 10 fs vibration is excited by the laser pulse. This excitation could start at any point within the laser excitation, so it will essentially will be blurred (convoluted) in time with the 140 fs duration (probably of Gaussian or similar shape) of the pump pulse. The authors have verified the performance of their algorithm using synthetic data created using as a model the structures retrieved from their analysis. This is of course important, but what would be needed here is a model where the duration of the excitation pules is taken into account. For example, the authors could take the synthetic data already generate (which includes the structural dynamics with 10 fs time oscillation) and make N copies of it, then displace the copies in time according to the intensity profile of the excitation laser pulse and calculate a new synthetic data set composed of the superposition of the N (time displaced) copies (with say N = 100 or so). This would approximate the blurring in time produce by the excitation. Then run the algorithm over this new blurred synthetic data and see I the 10 fs signal is recovered. This would be strong evidence that the algorithm can capture dynamics much faster than the excitation laser

pulse, and truly impressive.
While Figure S5 shows that multiple high frequencies appear above the noise floor, this sort of analysis can be affected by artifacts, and alone is not sufficient to support the claim of 1 fs resolution.

2. It seems that a weakness of the machine learning approach is that there is no clear connection between the output of the analysis and the physical evolution of the system, i.e. how to combine the modes to build the actual trajectory of the system. The output of the algorithm needs to be fitted to simulated data in order to retrieve the trajectory through the conical intersection. Why is that? Isn't that information also encoded in the diffraction patterns?
A related question, what is the physical significance of the very fast turning point in the chronos, as seen in Fig. 2a?

3. The simulation assumes a trajectory in a two-dimensional PES. As far as I understood, the machine learning algorithm output indicates that the dynamics takes place on a multidimensional space. What is the justification to use a two-dimensional PES to fit the data to?
The fitting procedure was not entirely clear to me. The manuscript states "we fit each of the simulated trajectories to the six experimental trajectories." But what exactly is meant here by fitting the trajectories? Is it the structures vs time, or is the machine learning algorithm also run on the simulation to extract the modes?
The manuscript also states "In this, we consider both temporal and spatial translations of the simulated trajectories, and allow the simulated trajectories to be linearly scaled." What is meant by linear scaling of the trajectories, in time, in space? Why is this scaling of trajectories needed, and what is the physical significance of this scaling?
A more technical question: How is it possible to simulate 500k trajectories, when most quantum chemical simulation can only produce a few hundred trajectories in a reasonable time?

Referee #2 (Remarks to the Author):

Review Hosseneinizadeh et al.,

In their manuscript the authors describe a new methodology to reevaluate previously published time-resolved serial crystallographic (TR-SFX) data collected from the photoactive yellow protein (PYP) to increase temporal resolution. Due to the jitter between the optical pump and the X-ray probe, the time resolution in TR-SFX experiments is typically limited to 100-200 fs. This temporal resolution is not high enough to overserve quantum chemical phenomena like crossing of the conical intersection between two potential energy surfaces. The authors describe computational methods to increase the time resolution to a few femtoseconds, while preserving the near atomic resolution of the crystallographic approach.
Overcoming the timing uncertainty in ultrafast time-resolved crystallography is a spectacular feat many would have thought impossible. Presenting this breakthrough in an approachable and convincing way would certainly warrant publication in Nature. However, currently the manuscript is not written with a wider audience in mind and does not put the work in context of previous literature on the system.
Before I go on, I want to make it clear that I cannot evaluate the computational and mathematical basis of the work and will have to leave this to the other reviewers. But, as I suspect is true for many readers in the wider audience of Nature, I need to be convinced by how well the results relate to previous crystallographic, spectroscopic and computational work on the PYP system. Without this comparison it is unclear what specific new mechanistic insights have been obtained with the new method. Much of the manuscript reads like the introduction to a problem without a clear conclusion leaving me somewhat unsatisfied with the analysis. Before publication can be considered I recommend a major revision taking the specific recommendations and general guidelines below into account. I really want to believe such temporal resolution is possible but more work needs to be put into the analysis and the manuscript.

• Too often the manuscript replaces clear information with vague general statements. This imprecision

already starts in the abstract where neither the protein under investigation nor the experimental method that generated the evaluated data is named. Instead the authors chose the vague "2'000-atom protein" and "extracted from experimental data" where photoactive yellow protein and time-resolved serial crystallography would have been much more precise. I understand that the authors want to highlight the general applicability of their method but still the abstract and text needs to contain the specifics and needs rewriting. The introduction could be shorter and better targeted.

• The title is somewhat misleading. Single-femtosecond resolution is likely overstated. Because of the physical limits of the experimental method it should be 5-10 fs at best (see discussion below on crystal size and homogenous excitation). Without refinements there is no atomic-resolution observation (see discussion on crystallographic evaluation below). And not the "2'000-atom protein" transverses a conical intersection but rather the 11 atoms of the p-coumaric acid. Of course, the chromophore is covalently bound to the protein but there is no description of the wider protein throughout the manuscript. I recommend to rephrase the title.

• Line 57 introduces p-cinnamic acid as the chromophore in PYP. This should be p-coumaric acid. The difference is only a singly hydroxy group, which however forms H-bonds with the protein and should not be neglected when describing the de-excitation and isomerization mechanism. Likely this is just a typo though as the chromophore shown in Figure 1 is correct.

• Further down in the abstract, the authors promise "single-femtosecond, atomic resolution movies". But in the end the 3D diffraction volumes obtained by the new method have not been used to define molecular structures to allow an atomistic interpretation. Figures S4 and S10 contain a comparison of electron density maps from conventional time-resolved crystallography and the new algorithm used in this paper. The similarity is convincing and provides validity to the new approach. But why were no structures refined and compared? Also, why only the 1.3 ps time delay which according to Figure S1 was not part of the data set obtained by subsampling of the experimental data. A series of electron density maps and a comparison of the refined structures at time delays that are discussed in the main manuscript and/or Pande et al would be much better. Both works have overlapping authors that could perform a crystallographic analysis to substantiate the progress in quality and temporal resolution as a highlight of the current manuscript.

• Another promise at the end of the abstract is the key parameters of the conical intersection and the de-excitation process. But again, I was not able to find these key parameters beside some poorly explained values in Table 1. For example, when does the trans-cis isomerization happen according to the new data? Figure 2 seems to imply that the features at 615 fs in Chrono 2-5 resemble the isomerization, which would fit well to the ~550 fs in Pande et al. Several times throughout the manuscript it is mentioned that the features around the C3-C4 bonds have "previously" or "normally" been associated with the trans-to-cis isomerization. If the authors want to interpret these features differently please be clear and less ambiguous in wording.

• I don't understand why Mode 1 is not shown since it represents the moving average of all modes. As such it should be the most important and the electron density maps obtained by it should be most similar to the ones obtained by conventional analysis. If electron density maps of mode 1 alone can't be shown because modes always have to be used in combination please think about a clearer nomenclature, for example relabel the maps to M1+M2, M1+M3 etc.

• On page 8-9 the modes 2-5 are compared in more detail. Please discuss how these modes relate to the reaction pathway. Photoactivation of PYP has a high quantum efficiency but still a significant proportion of the chromophore will relax back to the dark state while others will proceed towards the cis configuration within the investigated time frame. It might further be expected from QM/MM simulations that not all chromophores move fully synchronized which could also be due to experimental limitations (see below). Please interpret the modes in a common model of PYP photoactivation. Does the symmetry of Mode 5 for example reflect unproductive isomerizations? What is the relevance of the rapid charge oscillations in Mode 2 or the oscillatory charge redistribution in mode 3? How can it be that the same sequence of events is repeated after 615 fs in Mode 4? The manuscript needs some justification and discussion with respect to the literature why these observations are relevant.

• In line 182 it should be Tyr98 instead of Tye98

• Please include a Supplementary figure and discussion of the achievable time resolution taking the experimental parameters under which the data was collected into account. Beside the jitter mentioned in the main text, an important factor should be the size of the used crystals. Typical for such experiments is a crystal size of about 10 um which can be crossed by the activating pump laser in 33 fs or longer as the speed of light is reduced by ~25% in a watery environment. As the diffraction signal originates from an average of PYP molecules activated within this time frame there is a timing uncertainty inherent to the physical properties of the system. Even if PYP molecules in the first micrometer dominate the light signal because of the high optical density in a crystal, it is not clear to me how a time resolution of 1fs can be achieved as the described method is not sorting individual molecules. Similarly, please discuss how the physical limitations on timing from using 40 fs XFEL pulses and 140 fs optical laser pulses can be overcome. Pushing the temporal resolution to extreme limits is the most important outcome from this work. But I need more evidence as given by the Frequency analysis in Fig S5. Is the system just so deterministic that faster times can be extracted? If yes would that still be an observation or is it already a simulation?

• Please include a description of how the crystallographic snapshots have been sorted for Fig S1. Is this the same sorting as in Pande et al., or have the timing tool signals also be reevaluated to increase timing accuracy in the input data to the current analysis?

• In Line 367 it should be "In order to confirm…"

• Figure 1: Please include the cis model of the chromophore. Also, the supplementary movies would benefit from containing the refined structures to guide interpretation of the density fluctuations.

• Figure 2b: Please include the contour level at which the maps are shown. Arrows indicating the specific ED features discussed in the main text would increase accessibility of the manuscript.

• Figure 3a: Please add a unit to the color bar.

• Figure 4: I don't see how this Figure is relevant enough for a main figure of the manuscript. With such arbitrary axes it does poorly represent the new data that could be obtained with the method. Combining the figure with Table 1 or other relevant parameters that describe the conical intersection to a wider audience would provide more specific information.

• Table 1: Please add units to the different parameters. Please explain or highlight the relevance of these parameters, if possible, in Figure 4 to make it more meaningful. Can some of these values be cross-checked with literature values obtained with orthogonal techniques?

• Figure S1: Please include reference to the data source. Please include 1.3 ps data used for Figure S4.

• Figure S3: For simplicity please mention in the figure legend to which step synthetic and output 3D diffraction volumes belong in Figure S2.

• Figure S4: Please include panels presenting more than the chromophore region, so that noise levels in the maps can be compared. Please include the contour levels at which the maps are shown. Could be combined with Fig S10 to provide a convincing case for the validity of the methodology, especially if earlier time delays can be included as mentioned above.

• Please provide resolution limits in Å instead of 1/Å to increase readability.

• The material and methods section mentioned on line 354 that also the dark data yielded a single mode from the NLSA analysis. A supplementary figure comparing this dark mode the light modes on the same scale may act as convincing control for the technique, especially if no high frequency components are

present.

• From line 529 onwards the authors explain how they fitted experimental trajectories to their experimental counterparts. This leaves me puzzled because (line 65-67) in the introduction the authors highlight that electronic structure calculations of PYP are unfeasible for the foreseeable future due to its size. However only the chromophore excitation has to be modeled and is relevant in the ultrafast time domain unless the authors found other significant motions further out in the protein as already mentioned above.

Referee #3 (Remarks to the Author):

In the work performed by Hosseinizadeh and colleagues the authors used time-resolved crystallographic data to characterize the dynamics of the structural states of the PYP system. This was achieved by a combination of machine learning with experimental and theoretical techniques.
From the methodological standpoint on the machine learning side the work is sound. The methods used are mostly based on the framework described by Coifman et al where the nonlinear spectral analysis is described. On the way the technique is described it is perhaps an exaggeration to refer to it as geometric machine learning (although not wrong), as mostly what it seems that is being done is non linear dimension reduction, and ultimately then vectors are used as features to perform the learning, so the learning is not really being done in the manifold. Generally, in geometric deep learning you would learn directly from the manifold with other types of techniques that could include graph neural networks. In any case these comments are more of style and terminology but the bottom-line is that the methodological approach seem sound.

Author Rebuttals to Initial Comments:

Response to Reviewer Comments

Reviewer #1

We are grateful to the distinguished Reviewer for carefully reading our submission, and the insightful comments. Below, we discuss each item in turn, and list the changes made in thelight of the comment.

Comment 1a:

The evidence for the claim of 1 fs resolution is the observation of an oscillatory signal with 10 fs period. It is difficult to picture how this signal could be produced and measured if the system was pumped by a 140 fs laser pulse. The first question is, what is producing this signal? This should be discussed in the manuscript. The only thing that comes to mind that matches the period would be a C-H vibration.

Response:

The evidence for the femtosecond temporal resolution stems primarily from two observations. First, the presence of sharp (~2 fs wide) turning points at 615 fs in all chronos(Fig. 2a). Second, a high signal-to-noise-ratio, oscillatory signal with a 10.5 fs period, supporting single-femtosecond timing acuity via (non-uniform) Shannon-Nyquist sampling.(As shown by [Jerri, A. J. Proceedings of the IEEE 65, 1565-1596 (1977)], the clear observation of a frequency *f* in a noisy, non-uniformly sampled signal requires a timing acuity of about *f*/10.) These observations are now buttressed by close agreement between the peaks we observe in the Fourier spectra of the chronos and the peaks observed by ultrafast Raman spectroscopy in the

experimentally investigated (3 - 30) THz range [Kuramochi et al., Nature Chemistry 9, 660 (2017)]. A new section has been added to Methods regarding this comparison.

We agree with the Reviewer that the 10.5 fs oscillatory signal may be associated with a C-H bond stretch, typically observed in the (10 – 12) fs range at moderate to high intensities [G. Socrates, "Infrared and Raman Characteristic Group Frequencies: Tables and Charts", 3rd Edition, John Wiley & Sons, 2007]. It is also possible that the oscillatory signal stems from an N-H vibrational bond stretch, which typically results in an absorption band in the (9.5 - 10.4 fs) range. A note to this effect has been inserted.

Changed:
Lines 143 – 163 now read: Fourier analysis of the chronos by multi-taper methods [30,31] reveals the clear presence of frequencies of up to 95 THz (10.5 fs) at signal-to-noise ratios of ~ 5 or more (Supplementary Fig. S5). As the observation of a frequency component in an initially non-uniform set of timepoints requires a time resolution ~ 5 – 10 times shorter than the period of the component [32], the observation of a 10.5 fs signal validates the high time resolution of our approach. This is further supported by the clear presence of sharp (~2-fs-

wide) turning points in all chronos at 615 fs (Fig. 2a), which we regard as a fiduciary markerfor the conical intersection.

Fourier analysis of the chronos before 615 fs, i.e., before the encounter with the conical intersection brings multiple species into play, reveals three prominent peaks at 4, 21, and 33.5 THz in the spectroscopically well explored region spanning the (3 – 30) THz range and its vicinity. These peaks have been previously observed by ultrafast time-domain Raman spectroscopy of PYP before the conical intersection [33]. The peak frequencies match those we observe to within 7% or better. (See Methods section entitled "Comparison with spectroscopically determined PYP frequency spectrum".) Such close agreement with independently known "ground truths" is the ultimate test of any machine learning approach.

The signal at 95 THz may be associated with a C-H bond stretch, typically observed in the 100 THz (10 – 12) fs band at moderate to high intensities. It is also possible that this oscillatory signal stems from an N-H vibrational bond stretch, which typically results in an absorption band in the (9.5 – 10.4) fs range [34].

See also new Table S1 and associated caption in lines 1167 – 1181, which read:

Source	Peak Position (± 2 THz)							
Chrono 2						70		
Chrono 3			37					
Chrono 4	3	21	30	58	64	72		98
Chrono 5	5				64	74	89	96
Average peak frequencies	4	21	33.5	58	64	72	89	97
Time-resolved Raman freq.	4	22.5	34.7	Not available				
Difference	0%	7%	3%					

Table S1. Peak positions obtained from multi-taper Fourier analysis of the chronos before the encounter with the conical intersection vs. peaks positions obtained from time-resolved Raman spectra of PYP, all in THz. Each chrono displays a characteristic frequency spectrum, which sometimes includes a subset of the peaks observed in another chrono. The exact peak position can change by ~ 2 THz as multi-taper parameters are varied. Closely separated peaks are identified as one peak at the average frequency position. Using time- resolved Raman spectroscopy, Kuramochi et al.[33] have investigated the approximately 3 – 30THz frequency range in PYP. As shown in the Table, all frequency peaks revealed by multi- taper analysis of the chronos before the encounter with the conical intersection can be identified with a peak observed by Kuramochi et al. with a frequency accuracy of 7% or better. However, the time-resolved Raman spectra contain additional peaks not observed inour analysis of the chronos. This suggests that not all spectroscopically observed frequencies pertain to the structure dynamical collective variables we have extracted from time-resolved scattering data.

Comment 1b:
The second question is whether the signal could be detected and not blurred out by the duration of the excitation laser pulse. Let us assume that a 10 fs vibration is excited by the laser pulse. This excitation could start at any point within the laser excitation, so it will essentially be blurred (convoluted) in time with the 140 fs duration (probably of Gaussian or similar shape) of the pump pulse. The authors have verified the performance of their algorithm using synthetic data created using as a model the structures retrieved from their analysis. This is of course important, but what would be needed here is a model where the duration of the excitation pules is taken into account.

Response:
We appreciate the Reviewer's comment, and regret that the issue was not treated adequately in the previous manuscript.

As the Reviewer notes, pump and probe pulses used in serial femtosecond crystallography are typically tens of femtoseconds long. In principle, simulation can be used to investigate the effect of the pulse characteristics on the achievable time resolution. However, extensive studies of the spatial and temporal characteristics of incident pulses [43] has shown that it is extremely difficult, if not impossible, to determine the actual pulse characteristics in time- resolved serial femtosecond crystallography. We therefore follow a data-driven machine- learning approach, whereby the veracity of our approach is judged by the extent to which the approach reproduces independently known "ground-truths". Using experimental XFEL pump-probe data obtained with optical pulses as long as 75 fs, we have previously shown that the vibrational frequencies revealed by our approach accurately match those of well- known systems such as N_2, even when the incident pulse lengths are long compared with the vibrational frequencies [19]. The relevant table from this reference is reproduced here for the Reviewer's convenience. Similarly, each of the PYP vibrational frequencies revealed by our present work has been independently observed (see Table S1). This clearly demonstratesthat accurate dynamical information can be extracted by our technique.

d **Vibrational Periods (fs)**

Measured	Theory	State	Remarks
15.2	15.1	$X^2\Sigma^+_g$	Previously observed in time domain
16.7	16.7	$X^1\Sigma^+_g$	
19.6	17.7	$A^2\Pi_u$	Overlap between peaks degrades accuracy
20.2	22.4	$a^3\Pi_u$	
23.6	23.7	$A^1\Pi_u$	

Changed:
Lines 400 – 415 now read: The pump and probe pulses used in serial femtosecond crystallography are typically tens of femtoseconds long. In principle, simulation can be usedto investigate the effect of the pulse characteristics on the achievable time resolution. However, extensive studies of the spatial and temporal characteristics of incident pulses [43] has shown that it is extremely difficult, if not impossible, to determine the actual pulse characteristics in time-resolved serial femtosecond crystallography. We therefore follow a data-driven machine-learning approach, whereby the veracity of our results is determinedby the extent to which our algorithm reproduces independently known ground-truths.
Using experimental XFEL pump-probe data obtained with optical pulses as long as 75 fs, we have previously shown that the vibrational frequencies revealed by our approach accurately match those of well-known systems such as N_2, even when the incident pulse lengths are long compared with the vibrational frequencies [19]. Similarly, in the spectroscopically examined frequency range, each of the PYP vibrational frequencies revealed by our present work has been independently observed by time-resolved Raman techniques [33]. (See also Table S1). This clearly demonstrates that reliable dynamical information can be extracted with few-femtosecond time resolution.

Comment 2a:
It seems that a weakness of the machine learning approach is that there is no clear connection between the output of the analysis and the physical evolution of the system, i.e. how to combine the modes to build the actual trajectory of the system. The output of the algorithm needs to be fitted to simulated data in order to retrieve the trajectory through theconical intersection. Why is that? Isn't that information also encoded in the diffraction patterns?

Response:
As described in [Henry & Hofrichter, Methods in Enzymology, 210, 129 (1992)] and [Schmidtet al. Biophys J 84, 2112 (2003)], additional information is needed to reconstruct the physical evolution of a system from the modes obtained by standard (linear) singular value decomposition (SVD). This stems from noise attenuating the singular value (relative weight)of each mode by an unknown amount. As such, additional information (for example a kinetic model) is needed to reconstruct the physical behavior of the system. Fung et al. [Fung et al., Nature 532, 471 (2016)] demonstrate how this issue can be resolved for dynamical systems with minimal additional information.

Changes:
Lines 215 - 218 now read: As in standard singular value decomposition, additional information is needed to accomplish this task [Henry & Hofrichter (1992), Schmidt et al. (2003), Fung et al. (2016)]. Away from the conical intersection, we determine the appropriate mode combinations as previously described in [37], and extract structures fromthe resulting DED's as outlined in [35].

Comment 2b:
A related question, what is the physical significance of the very fast turning point in the chronos, as seen in Fig. 2a?

Response:
The sharp turning points reflect the rapid changes in each of the dynamical collective modesas the conical intersection is traversed. This assignment is now explicitly stated in the manuscript.

Changed:
Lines 151 - 152 read: Fourier analysis of the chronos before 615 fs, i.e., before the encounter with the conical intersection ff.

Comment 3a:
The simulation assumes a trajectory in a two-dimensional PES. As far as I understood, the machine learning algorithm output indicates that the dynamics takes place on a multidimensional space. What is the justification to use a two-dimensional PES to fit the data to?

Response:
We agree that the dynamics could indeed unfold on a multi-dimensional PES manifold. In principle, this space could be described by an effective Hamiltonian involving all experimentally observed modes. Such an approach would, however, substantially increase the number of free parameters. To circumvent this problem, we adopt a two-dimensional PES as the minimal model needed to describe the vicinity of the conical intersection. This approximation significantly reduces the danger of overfitting, and enables us to solve the time-dependent Schrödinger equation for a sufficiently dense set of points in parameter space. The appropriateness of this approach is supported by the observation that all five pairwise combinations of experimentally determined structure-dynamical modes yield two-dimensional PES parameter-sets, which are the same to within uncertainty. This indicates that, in the vicinity of the conical intersection, all dynamical trajectories unfold on the sametwo-dimensional PES manifold. Further away from the conical intersection, our analysis reveals the presence of at least five distinct trajectories corresponding to different conduitsto and from the vicinity of conical intersection.

Changed:
Lines 236 - 243 now read: The PES's are approximated by a second-order Taylor expansion, known as the vibronic coupling model [38,39,] and are assumed to be symmetric with respect to both normal modes, possibly differing in their respective vibrational frequencies. Inter-state coupling is mediated via one mode only. To determine the dynamics in the space spanned

by two modes, we numerically solve the time-dependent Schrödinger equation for a wave packet initially occupying the excited electronic state.

Comment 3b:
The fitting procedure was not entirely clear to me. The manuscript states "we fit each of the simulated trajectories to the six experimental trajectories." But what exactly is meant here by fitting the trajectories? Is it the structures vs time, or is the machine learning algorithm also run on the simulation to extract the modes? The manuscript also states "In this, we considerboth temporal and spatial translations of the simulated trajectories, and allow the simulated trajectories to be linearly scaled." What is meant by linear scaling of the trajectories, in time, in space? Why is this scaling of trajectories needed, and what is the physical significance of this scaling?

Response:
We apologize for the lack of clarity on this point. The purpose of identifying simulated trajectories with experimental ones is to determine the set of physical parameters best able to describe each of the experimental trajectories. This is performed by comparing simulated dynamical trajectories with experimental ones. In principle, one would simply select the simulated trajectories most closely resembling (have the smallest chi-squared value with respect to) each of the experimental trajectories. In practice, the axes describing the experimental and simulated trajectories may be rotated and/or rigidly shifted with respect to each other. Attenuation due to noise and timing jitter may also change the "length" of each experimental axis by an unknown amount. The best-fit search must therefore allow rigid shifts in the origin, axis rotations, and scaling of the simulated trajectories.

Changed:
Lines 586 – 596: A new section entitled "Identifying each experimental trajectory with a simulated counterpart" has been inserted to elucidate these points.

Comment 3c:
A more technical question: How is it possible to simulate 500k trajectories, when most quantum chemical simulation can only produce a few hundred trajectories in a reasonable time?

Response:
Most quantum chemical simulations are computationally expensive because their complexity increases rapidly with the number of degrees of freedom. We circumvent this problem by employing an effective two-modes model, which meets the minimal requirements of a conical intersection, and for which the time-dependent Schrödinger equation can be solved efficiently. This enables us to simulate a large number of trajectories, each of which requires only tractably small computational resources.

Changed:
Lines 223 - 227 now read: The calculations employ an effective model Hamiltonian in a compressed collective-mode space to capture the properties of a conical intersection and its vicinity. This allows numerically exact calculation of the nuclear quantum-dynamics as a function of a small number of molecule-specific model parameters.

Response to Reviewer #2

We are grateful to the distinguished Reviewer for the careful reading of our submission, and the insightful comments and questions. Below, we address each comment, and list the changes made in the light of that comment.

Comment 1:
I need to be convinced by how well the results relate to previous crystallographic, spectroscopic and computational work on the PYP system. Without this comparison it is unclear what specific new mechanistic insights have been obtained with the new method. Much of the manuscript reads like the introduction to a problem without a clear conclusion leaving me somewhat unsatisfied with the analysis. Before publication can be considered I recommend a major revision taking the specific recommendations and general guidelines below into account. I really want to believe such temporal resolution is possible, but more work needs to be put into the analysis and the manuscript.

Response:
We are grateful to the distinguished Reviewer for the opportunity to revise our submissionin the light of her/his comments.

Comment 2:
Too often the manuscript replaces clear information with vague general statements. This imprecision already starts in the abstract where neither the protein under investigation nor the experimental method that generated the evaluated data is named. Instead the authors chose the vague “2’000-atom protein” and “extracted from experimental data” where photoactive yellow protein and time-resolved serial crystallography would have been much more precise. I understand that the authors want to highlight the general applicability of their method but still the abstract and text needs to contain the specifics and needs rewriting. The introduction could be shorter and better targeted.

Response:
We regret the lack of specifics. We have revised manuscript to provide more specific details without sacrificing accessibility.

Changes:
Throughout.

Comment 3:
The title is somewhat misleading. Single-femtosecond resolution is likely overstated. Because of the physical limits of the experimental method it should be 5-10 fs at best (see discussion below on crystal size and homogenous excitation). Without refinements there is no atomic-resolution observation (see discussion on crystallographic evaluation below). Andnot the "2'000-atom protein" transverses a conical intersection but rather the 11 atoms of the p-coumaric acid. Of course, the chromophore is covalently bound to the protein but there is no description of the wider protein throughout the manuscript. I recommend to rephrase the title.

Response:
The title, abstract, and text have been changed to reflect a time resolution in the few-femtosecond regime. We agree that only a specific part of the chromophore undergoes major changes in traversing the conical intersection. However, in general, it is not possible to determine this region objectively without performing the analysis we have outlined. As itturns out, only a small region of the chromophore participates in the isomerization reaction,but the nature and extent of this region are a priori unknown. Indeed, although PYP has been the subject of study for decades, our results now reveal that "peripheral" side-chains host previously unanticipated charge oscillations, whose importance remains to be investigated. In sum, we agree that only a part of the chromophore actively participates in the isomerization process, but exactly which part must be identified by the type of whole- chromophore analysis we have presented.

Changes:
Throughout, including the title and abstract.

Comment 4:
Line 59 introduces p-cinnamic acid as the chromophore in PYP. This should be p-coumaric acid. The difference is only a singly hydroxy group, which however forms H-bonds with the protein and should not be neglected when describing the de-excitation and isomerization mechanism. Likely this is just a typo though as the chromophore shown in Figure 1 is correct.
Response:
We thank the Reviewer for this correction, which has been implemented.

Changed:
Line 59: "p-cinnamic acid" has been changed to "p-coumaric acid".

Comment 5:
Further down in the abstract, the authors promise "single-femtosecond, atomic resolution movies". But in the end the 3D diffraction volumes obtained by the new method have not been used to define molecular structures to allow an atomistic interpretation. Figures S4 and S10 contain a comparison of electron density maps from conventional time-resolved crystallography and the new algorithm used in this paper. The similarity is convincing and provides validity to the new approach. But why were no structures refined and compared?

Response:
We apologize for this omission. Fig 2c now provides the refined cis structures at 800 fs, 900 fs and 3 ps. These were obtained as outlined in the Supplementary Information section 7 of [Fung et al., Nature 532, 471 (2016)] and in [Pandey et al. Nature Methods 17, 73 (2020)]. In principle, a movie of many refined structures can be compiled. That remains a future task.

Changed:
Fig. 2c now includes refined cis structures at 285 fs, 800 fs, 900 fs, and 3 ps. Additional references ([36] and [37]) have been provided. Finally, Table 2d provides the changes in the torsional angle with time. This angle is often used as a structurally intuitive isomerization reaction coordinate.

Comment 6:
Also, why only the 1.3 ps time delay which according to Figure S1 was not part of the data set obtained by subsampling of the experimental data. A series of electron density maps and a comparison of the refined structures at time delays that are discussed in the main manuscript and/or Pande et al would be much better. Both works have overlapping authorstemporal resolution as a highlight of the current manuscript.

Response:
Unfortunately, "1.3 ps" was a typographical error, and should have read "3 ps". The "continuous" experimental data extend only to ~ 900 fs, with the 3-ps data forming a separate cluster in time. We have, however, applied our approach to the 3-ps cluster of jittered data, and the results are shown in Figs. 2c and S4.

Changed:
1.3 ps has been changed to 3 ps throughout.

Comment 7:
Another promise at the end of the abstract is the key parameters of the conical intersectionand the de-excitation process. But again, I was not able to find these key parameters besidesome poorly explained values in Table 1.

Response:
We regret that the role and importance of the key parameters were insufficiently clear. The physically important parameters are the vibrational frequencies of the potential energy surfaces, and the coupling strength, i.e., the transition probability between the participating electronic states. In combination, the vibrational frequencies and the coupling define the topography of the CI. A new paragraph elucidates the meaning and significance of these parameters.

Changed:
Lines 269 - 274 now read: Specifically, the parameter ω characterizes the kinetic energy, and (r, θ) the initial position of the wave packet, respectively. The shape of the potential energy surfaces is determined mainly by the vibrational frequencies γ_1, γ_2. The coupling parameter λ defines the probability of a transition between the two electronic states as a result of the wave packet moving through, or close to the conical intersection. Vibrationalfrequencies and coupling strength together define the topography of the conical intersection.

Comment 8:
When does the trans-cis isomerization happen according to the new data? Figure 2 seems to imply that the features at 615 fs in Chrono 2-5 resemble the isomerization, which would fit well to the ~550 fs in Pande et al. Several times throughout the manuscript it is mentioned that the features around the C3-C4 bonds have "previously" or "normally" been associated with the trans-to-cis isomerization. If the authors want to interpret these featuresdifferently please be clear and less ambiguous in wording.

Response:
The answer to the question regarding the "time of isomerization" depends somewhat on definitions. Based on the sharp changes in the topos and chronos of modes 1 - 5, isomerization occurs at 615 fs. The point of closest approach to the conical intersection (indicated by a red circle in Fig. 3c) is reached (25±3) fs later. As noted in the text, inter- state transition is a statistical process. Our estimates therefore represent weighted averages. The qualifiers "previously" and "normally" were used to clarify the consensus understanding of the isomerization process in PYP, not to call this consensus into question.In order not to confuse the issue, we have reduced the use of these qualifiers as far as possible.

Changes: Throughout, as appropriate.

Comment 8:
I don't understand why Mode 1 is not shown since it represents the moving average of all modes. As such it should be the most important and the electron density maps obtained by

it should be most similar to the ones obtained by conventional analysis. If electron density maps of mode 1 alone can't be shown because modes always have to be used in combination please think about a clearer nomenclature, for example relabel the maps to M1+M2, M1+M3 etc.

Response:
As the Reviewer notes, Mode 1 represents a moving average over the data. As the excitation-induced changes are small compared with the mean, this mode is not particularlyinformative. However, Mode 1 has now been included in Figs. 2a, and S11. The Reviewer iscorrect that Mode 1 is added to all difference electron density maps. This point has been clarified.

Changed:
Line 844: Caption to Fig. 2 now makes it clear that Mode 1 has been added.
Lines 171 - 173 now read: As Mode 1 represents the moving average of the signal, its topo is added to each of the subsequent modes to facilitate comparison with DED's obtained by conventional means.

Comment 9:
On page 8-9 the modes 2-5 are compared in more detail. Please discuss how these modes relate to the reaction pathway.

Response:
The modes constitute the collective variables involved in isomerization via the conical intersection. These variables define the axes of the space in which the reaction pathways (aka dynamical trajectories) are traversed. Every point on the reaction pathway can thus be described in terms of the contribution of two collective variables.

Changed:
Lines 136 - 141 read: More specifically, each topo represents a characteristic difference electron density map, evolving in time as prescribed by its corresponding chrono (Fig. 2a). Each topo-chrono pair thus represents a characteristic dynamical mode of the charge distribution. In essence, these modes constitute the basis functions, which may be used in combination to describe the dynamical trajectories ("the reaction paths") of the system (see Fig. 3c, and below).

Comment 10a:
Photoactivation of PYP has a high quantum efficiency but still a significant proportion of the chromophore will relax back to the dark state while others will proceed towards the cis configuration within the investigated time frame. It might further be expected from QM/MM simulations that not all chromophores move fully synchronized which could also be due to

experimental limitations (see below). Please interpret the modes in a common model of PYP photoactivation.

Response:
As noted in response to the previous question, each mode represents an axis of the space in which a reaction pathway (dynamical trajectory) unfolds. Just as a dynamic trajectory in two-dimensional space can be defined as {x(t), y(t)}, a reaction pathway can be described asa succession of {x, y} values, thus combining the contribution of two collective variables.
Our work identifies the collective variables relevant to the isomerization process. While the modes are mutually orthogonal, the choice of modes is not unique, and a single mode need not correspond to an experimental observable. The observables correspond to dynamical trajectories, rather than a specific mode. Our work is able to capture these trajectories on the femtosecond scale, and, in the vicinity of the conical intersection, relate them to population transfer between states (Fig. 3).

As shown in Fig. 2c, our results can also be used to determine refined structures, and thus make contact with traditional interpretations of PYP isomerization. Our work extends this possibility to the deep femtosecond range. The primary focus of the present paper, however, is the crossing of conical intersections.

Changed:
These points are clarified in lines 134ff.

Comment 10b:
Does the symmetry of Mode 5 for example reflect unproductive isomerizations?

Response:
Yes. The fact that chrono-5 starts and ends at about the same value indicates a collective variable corresponding to unproductive isomerization.

Changed:
Lines 200-202 now read: Taken together, the modes described above reveal the structure-dynamical motifs involved in PYP de-excitation, including collective motions leading to successful and unsuccessful isomerization attempts.

Comment 10c:
What is the relevance of the rapid charge oscillations in Mode 2 or the oscillatory charge redistribution in mode 3?

Response:
We have highlighted these rapid charge oscillations in order to move away from the notion that the isomerization consists of a smooth transition from one well-ordered state to

another. Our results show that the reality is more complex, involving, for example, rapidcharge oscillations.

Comment 10d:
How can it be that the same sequence of events is repeated after 615 fs in Mode 4? The manuscript needs some justification and discussion with respect to the literature why these observations are relevant.

Response:
The chronos display two types of approximate symmetry about 615 fs: a) mirror symmetry; and, b) inversion symmetry. These symmetries necessarily entail repeating the same or reverse sequence of events after 615 fs as before this timepoint. As noted earlier, the dynamical trajectories, rather than the modes themselves, correspond to experimental observables. Our observations are potentially significant, because they elucidate, in unprecedented detail, the structure-dynamical trajectories of de-excitation via a conical intersection, and yield the properties of the relevant PES manifold.

Comment 10d:
In line 183 it should be Tyr98 instead of Tye98.

Response:
We are grateful for the correction.

Changed:
Appropriately corrected.

Comment 11a:
Please include a Supplementary figure and discussion of the achievable time resolution taking the experimental parameters under which the data was collected into account.

Response:
To assess the veracity of our approach, we rely on a data-driven machine-learning approach, whereby the veracity of our results is determined by the extent to which our algorithm reproduces independently known ground-truths. [...], we have previously shown that the vibrational frequencies as high as 50 THz revealed by our approach accurately match those of well-known systems such as N_2, even when the incident pulse lengths are long compared with the vibrational frequencies [19]. Similarly, in the spectroscopically examined frequency range, each of the PYP vibrational frequencies revealed by our present work has been independently observed by time-resolved Raman techniques [33]. (See also table inserted below and Table S1). The table below clearly demonstrates that reliable dynamical information can be extracted with few-femtosecond time resolution. Theoretically, the achievable time resolution is determined by the total number of snapshots, the timing jitter

and pixel noise, and the average time interval between snapshots. Application of the expression (8) of the Supplementary section 3 of Fung et al. to the present PYP dataset leadsto a time resolution estimate in the sub-femtosecond regime. Such estimates, of course, represent the rarely achievable best case.

d **Vibrational Periods (fs)**

Measured	Theory	State	Remarks
15.2	15.1	$X^2\Sigma^+_g$	Previously observed in time domain
16.7	16.7	$X^1\Sigma^+_g$	
19.6	17.7	$A^2\Pi_u$	Overlap between peaks degrades accuracy
20.2	22.4	$a^3\Pi_u$	
23.6	23.7	$A^1\Pi_u$	

Changed:

Lines 406 – 415 now read: We follow a data-driven machine-learning approach, whereby the veracity of our results is determined by the extent to which our algorithm reproduces independently known ground-truths. […] we have previously shown that the vibrational frequencies revealed by our approach accurately match those of well-known systems such as N_2 [19]. Similarly, in the spectroscopically well examined frequency range of about (3 – 30) THz, each of the PYP vibrational frequencies revealed by our present work has been independently observed by time-resolved Raman techniques [33] (see Table S1). This clearly demonstrates that reliable dynamical information can be extracted with few-femtosecond time resolution.

Comment 11b:

Beside the jitter mentioned in the main text, an important factor should be the size of the used crystals. Typical for such experiments is a crystal size of about 10 um which can be crossed by the activating pump laser in 33 fs or longer as the speed of light is reduced by ~25% in a watery environment. As the diffraction signal originates from an average of PYP molecules activated within this time frame there is a timing uncertainty inherent to the physical properties of the system. Even if PYP molecules in the first micrometer dominate the light signal because of the high optical density in a crystal, it is not clear to me how a time resolution of 1fs can be achieved as the described method is not sorting individual molecules.

Response:

We agree that the crystal size and illumination conditions likely play an important role. Some of these effects have been examined in detail in [Grünbein et al. Nat. Methods 17, 681(2020)]. The primary message of such analyses is that it is extremely difficult, if not impossible, to characterize, let alone simulate the details of the experimentally achieved

illumination conditions. As noted above, we therefore adopt a data-driven machine-learning approach, in which the reliability of our approach is determined by the extent to which it can reproduce independently known ground truths.

Changed:
Lines 151-158 now read: Fourier analysis of the chronos before 615 fs, i.e., before the encounter with the conical intersection brings multiple species into play, reveals three prominent peaks at 4, 21, and 33.5 THz in the spectroscopically well explored region spanning the (3 – 30) THz range and its vicinity. These peaks have been previously observedby ultrafast time-domain Raman spectroscopy of PYP before the conical intersection 33.
The peak frequencies match those we observe to within 7% or better. (See Methods section entitled "Comparison with spectroscopically determined PYP frequency spectrum", and Table S1.) Such close agreement with independently known "ground truths" is the ultimate test of any machine learning approach.
Lines 1167 – 1181: A new Table S1 clarifies the above observations further.

Comment 11c:
Similarly, please discuss how the physical limitations on timing from using 40 fs XFEL pulses and 140 fs optical laser pulses can be overcome. Pushing the temporal resolution to extreme limits is the most important outcome from this work. But I need more evidence as given by the Frequency analysis in Fig S5. Is the system just so deterministic that faster times can be extracted? If yes would that still be an observation or is it already a simulation?

Response:
As the Reviewer notes, pump and probe pulses used in serial femtosecond crystallography are typically tens of femtoseconds long. In principle, simulation can be used to investigate the effect of the pulse characteristics on the achievable time resolution. However, extensivestudies of the spatial and temporal characteristics of incident pulses [43] has shown that it is extremely difficult, if not impossible, to determine the actual pulse characteristics in time- resolved serial femtosecond crystallography. Using experimental XFEL pump-probe data obtained with optical pulses as long as 75 fs, we have previously shown that the vibrational frequencies revealed by our approach accurately match those of well-known systems such as N_2, even when the incident pulse lengths are long compared with the vibrational frequencies [19]. This clearly demonstrates that reliable dynamical information can be extracted with few-femtosecond time resolution. See also Response to Comment 11b above.

Changed:
Lines 400 – 415: A new section entitled "Validating the achieved time resolution" has been addedto Methods.

Comment 12:
Please include a description of how the crystallographic snapshots have been sorted for FigS1. Is this the same sorting as in Pandey et al., or have the timing tool signals also be reevaluated to increase timing accuracy in the input data to the current analysis?

Response:
The sorting is the same as in Pandey et al. The timestamps were not reevaluated in any way.

Comment 13:
In Line 367 it should be "In order to confirm…"

Response:
Thank you for the correction, which has been implemented (line 419).

Comment 14:
Figure 1: Please include the cis model of the chromophore. Also, the supplementary movies would benefit from containing the refined structures to guide interpretation of the density fluctuations.

Response:
Figs. 2c-d now provide refined structures and the evolution of the torsional angle with time. The compilation of refined structural movies across the entire time range remains a future task.

Changed:
Lines 856 – 864: Additional figure and table inserted as Figs. 2c and d, respectively.

Comment 15:
Figure 2b: Please include the contour level at which the maps are shown.

Response: The contour level is 3 sigma throughout.

Changes:
Throughout: the contour level has been specified as 3 sigma.

Comment 16:
Figure 3a: Please add a unit to the color bar.

Response:
Thank you for alerting us to the missing color-bar scale.

Changes:
Appropriate figures throughout: The color-bar unit has been specified to be fs.

Comment 17:
Figure 4: I don't see how this Figure is relevant enough for a main figure of the manuscript. With such arbitrary axes it does poorly represent the new data that could be obtained withthe method. Combining the figure with Table 1 or other relevant parameters that describe the conical intersection to a wider audience would provide more specific information.

Response:
We regret that Fig. 4 in its previous form was inadequate. The figure has been revised to include the evolution of the two state populations as the conical intersection is traversed. The caption now includes a description of the key parameters of the conical intersection inmore accessible terms.

Changes:
New Fig. 4 shows the conical intersection and the evolution of the state populations. Thenew caption includes a more accessible description of the key parameters.

Comment 18:
Table 1: Please add units to the different parameters. Please explain or highlight the relevance of these parameters, if possible, in Figure 4 to make it more meaningful. Can some of these values be cross-checked with literature values obtained with orthogonal techniques?

Response:
We thank the Reviewer for these suggestions. Units have been added. To the best of our knowledge, there are no comparable studies on PYP.

Changes:
Unites have been added to the parameters of Table 1. Figure 4 has been revised.

Comment 19:
Figure S1: Please include reference to the data source. Please include 1.3 ps data used forFigure S4.

Response:
Thank you for the suggestions. Data sources have been provided. As noted earlier, "1.3 ps" was a typographical error, which has been corrected to "3 ps". The 3 ps data lie outside the time window of Fig. S4, and their inclusion would significantly compress the most pertinent section of the data along the time axis. We therefor ask for the Reviewer's indulgence.

Change:
Data sources provided in line 396ff. 1.3 ps has been changed to 3 ps throughout.

Comment 20:
Figure S3: For simplicity, please mention in the figure legend to which step synthetic and output 3D diffraction volumes belong in Figure S2.

Response:
Thank you for the suggestion.

Change:
Lines 1077 – 1080: Legend to Fig. S3 now includes the following: Pearson correlation and R-factor between synthetic (input) and output 3D diffraction volumes obtained from step 7 ofFig. S2.

Comment 21:
Figure S4: Please include panels presenting more than the chromophore region, so that noise levels in the maps can be compared. Please include the contour levels at which the maps are shown. Could be combined with Fig S10 to provide a convincing case for the validity of the methodology, especially if earlier time delays can be included as mentioned above.
Response:
Thank you for the suggestions.

Changes:
Fig. S4 now presents more than the immediate chromophore region. Fig. S4 caption specifies the contour level is 3 sigma. Figs. S4 and S10 have been combined into newFig. S4.

Comment 22:
Please provide resolution limits in Å instead of 1/Å to increase readability.

Response:
Thank you for the suggestion.

Changed:
Resolution limits given in Å.

Comment 23:
The material and methods section mentioned on line 354 that also the dark data yielded a single mode from the NLSA analysis. A supplementary figure comparing this dark mode the

light modes on the same scale may act as convincing control for the technique, especially ifno high frequency components are present.

Response:
Thank you for the suggestion. In Fig. S11, the modes obtained from the light data are compared with the modes obtained from the dark data. The first two "dark" chronos are identical, except for scale. This is a hallmark of a one-parameter process. Correlation analysis shows the single-parameter process correlates with the integrated Bragg spot intensity (Pearson correlation: 0.93), most likely due to slight drift in the incident beam intensity. The subsequent dark chronos represent noise.

Please note, however, that the dark data were not recorded as a function of pump-probe delay. The horizontal axis for the dark data in essence represents the nominal time elapsed since the start of the experiment, while the horizontal axis for the light data represents the pump-probe delay.

Changes:
Lines 374 – 377 now read: Applying the same preprocessing steps to the dark snapshots results in a data matrix of $D = 15{,}487$ rows and $N = 79{,}937$ columns. As no optical pump, and hence no pump-probe time delay was involved in recording dark snapshots, these datawere lexicographically sorted according to run numbers followed by event numbers.

New Fig. S11 inserted with the following captions (lines 1159 – 1164): Fig. S11 | a, The first five chronos obtained from light data ordered according to pump-probe delay. b, The firstfive chronos obtained from dark data lexicographically sorted according to run numbers followed by event numbers. The first two chronos are identical, except for scale. This is a hallmark of a one-parameter process. Correlation analysis shows the single-parameter process correlates with the integrated Bragg spot intensity (Pearson correlation: 0.93), mostlikely pertaining to drift in the incident beam intensity. The subsequent chronos represent noise.

Comment 24:
From line 529 onwards the authors explain how they fitted experimental trajectories to their experimental counterparts. This leaves me puzzled because (line 65-67) in the introduction the authors highlight that electronic structure calculations of PYP are unfeasible for the foreseeable future due to its size. However only the chromophore excitation has to be modeled and is relevant in the ultrafast time domain unless the authors found other significant motions further out in the protein as already mentioned above.
Response: We agree with the Reviewer that the photoisomerization reaction in PYP involves a relatively small region of the protein. Of course, such information is not availablefor a new system. More importantly, our analysis demonstrates that a few collective variables can describe a complex dynamical system, with each collective variable

representing a large ensemble of atoms. Our effective-Hamiltonian approach is applicableto all systems, which can be described in terms of a small number of collective modes, regardless of the total number of atoms involved. This means that our approach can, in principle, accurately model complex dynamical processes by identifying a small number ofcollective variables each involving a large number of atoms.

Changed:
Lines 223-227 read: The calculations employ an effective model Hamiltonian in a compressed collective-mode space to capture the properties of a conical intersection and itsvicinity. This allows numerically exact calculation of the nuclear quantum-dynamics as a function of a small number of molecule-specific model parameters.

Response to Reviewer #3
We are grateful to the distinguished Reviewer for finding our algorithmic approach sound.We agree that the approach we use is different from Geometric *Deep* Learning. However, our approach is based on the geometric properties of the manifold learned from the data.We therefore ask for the Reviewer's indulgence to use the term Geometric Machine Learning as we have in the past.

Reviewer Reports on the First Revision:

N/A

Author Rebuttals to First Revision:

Response to Reviewer Comments

Reviewer #1
We are grateful to the distinguished Reviewer for carefully reading our submission, and the insightful comments. Below, we discuss each item in turn, and list the changes made in the light of the comment.

Comment 1a:
The evidence for the claim of 1 fs resolution is the observation of an oscillatory signal with 10 fs period. It is difficult to picture how this signal could be produced and measured if the system was pumped by a 140 fs laser pulse. The first question is, what is producing this signal? This should be discussed in the manuscript. The only thing that comes to mind that matches the period would be a C-H vibration.

Response:
The evidence for the femtosecond temporal resolution stems primarily from two observations. First, the presence of sharp (~2 fs wide) turning points at 615 fs in all chronos (Fig. 2a). Second, a high signal-to-noise-ratio, oscillatory signal with a 10.5 fs

period, supporting single-femtosecond timing acuity via (non-uniform) Shannon-Nyquist sampling. (As shown by [Jerri, A. J. Proceedings of the IEEE 65, 1565-1596 (1977)], the clear observation of a frequency *f* in a noisy, non-uniformly sampled signal requires a timing acuity of about *f*/10.) These observations are now buttressed by close agreement between the peaks we observe in the Fourier spectra of the chronos and the peaks observed by ultrafast Raman spectroscopy in the experimentally investigated (3 - 30) THz range [Kuramochi et al., Nature Chemistry 9, 660 (2017)]. A new section has been added to Methods regarding this comparison.

We agree with the Reviewer that the 10.5 fs oscillatory signal may be associated with a C-H bond stretch, typically observed in the (10 – 12) fs range at moderate to high intensities [G. Socrates, "Infrared and Raman Characteristic Group Frequencies: Tables and Charts", 3rd Edition, John Wiley & Sons, 2007]. It is also possible that the oscillatory signal stems from an N-H vibrational bond stretch, which typically results in an absorption band in the (9.5 - 10.4 fs) range. A note to this effect has been inserted.

Changed:
Lines 162 – 183 now read: Fourier analysis of the chronos by multi-taper methods [30,31] reveals the clear presence of frequencies of up to 95 THz (10.5 fs) at signal-to-noise ratios of ~ 5 or higher (Supplementary Fig. S5). As the observation of a frequency component in an initially non-uniform set of timepoints requires a time resolution ~ 5 – 10 times shorter than the period of the component [32], the observation of a 10.5 fs signal validates the high time resolution of our approach. This is further supported by the clear presence of sharp (~2-fs-wide) turning points in all chronos at 615 fs (Fig. 2a), which we regard as a fiduciary marker for the conical intersection in PYP.

Fourier analysis of the chronos before 615 fs, i.e., before the encounter with the conical intersection brings multiple species into play, reveals three prominent peaks at 4, 21, and 33.5 THz in the spectroscopically well explored region spanning the (3 – 30) THz range and its vicinity. These peaks have been previously observed by ultrafast time-domain Raman spectroscopy of PYP before the conical intersection [33]. These spectroscopically measured PYP peak frequencies match those we observe to within 7% or better. (See Methods section entitled "Comparison with spectroscopically determined PYP frequency spectrum".) Such close agreement with independently known "ground truths" is the ultimate test of any machine learning approach.

The signal at 95 THz may be associated with a C-H bond stretch, typically observed in the 100 THz (10 – 12) fs band at moderate to high intensities. It is also possible that this oscillatory signal stems from an N-H vibrational bond stretch, which typically results in an absorption band in the (9.5 – 10.4) fs range [34].

See also new Table S1 and associated caption in lines 1212 – 1226, which read:

Source	Peak Position (± 2 THz)							
Chrono 2						70		

Chrono 3			37					
Chrono 4	3	21	30	58	64	72		98
Chrono 5	5				64	74	89	96
Average peak frequencies	4	21	33.5	58	64	72	89	97
Time-resolved Raman freq.	4	22.5	34.7	Not available				
Difference	0%	7%	3%					

Table S1. Peak positions obtained from multi-taper Fourier analysis of the chronos before the encounter with the conical intersection vs. peak positions obtained from time-resolved Raman spectra of PYP, all in THz. Each chrono displays a characteristic frequency spectrum, which sometimes includes a subset of the peaks observed in another chrono. The exact peak position can change by ~ 2 THz as multi-taper parameters are varied. Closely separated peaks are identified as one peak at the average frequency position. Using time-resolved Raman spectroscopy, Kuramochi et al.[33] have investigated the approximately 3 – 30 THz frequency range in PYP. As shown in the Table, all frequency peaks revealed by multi-taper analysis of the chronos before the encounter with the conical intersection can be identified with a peak observed by Kuramochi et al. with a frequency accuracy of 7% or better. However, the time-resolved Raman spectra contain additional peaks not observed in our analysis of the chronos. This suggests that not all spectroscopically observed frequencies pertain to the structure dynamical collective variables we have extracted from time-resolved scattering data.

Comment 1b:
The second question is whether the signal could be detected and not blurred out by the duration of the excitation laser pulse. Let us assume that a 10 fs vibration is excited by the laser pulse. This excitation could start at any point within the laser excitation, so it will essentially be blurred (convoluted) in time with the 140 fs duration (probably of Gaussian or similar shape) of the pump pulse. The authors have verified the performance of their algorithm using synthetic data created using as a model the structures retrieved from their analysis. This is of course important, but what would be needed here is a model where the duration of the excitation pules is taken into account.

Response:
We appreciate the Reviewer's comment, and regret that the issue was not treated adequately in the previous manuscript.

As the Reviewer notes, pump and probe pulses used in serial femtosecond crystallography are typically tens of femtoseconds long. In principle, simulation can be used to investigate the effect of the pulse characteristics on the achievable time resolution. However, extensive studies of the spatial and temporal characteristics of incident pulses [43] has shown that it is extremely difficult, if not impossible, to determine the actual pulse characteristics in time-resolved serial femtosecond crystallography. We therefore follow a data-driven machine-learning approach, whereby the veracity of our approach is judged by the extent to which the approach reproduces independently known "ground-truths". Using experimental XFEL pump-probe data obtained with

optical pulses as long as 75 fs, we have previously shown that the vibrational frequencies revealed by our approach accurately match those of well-known systems such as N_2, even when the incident pulse lengths are long compared with the vibrational frequencies [19]. The relevant table from this reference is reproduced here for the Reviewer's convenience. Similarly, each of the PYP vibrational frequencies revealed by our present work has been independently observed (see Table S1). This clearly demonstrates that accurate dynamical information can be extracted by our technique.

d **Vibrational Periods (fs)**

Measured	Theory	State	Remarks
15.2	15.1	$X^2\Sigma_g^+$	Previously observed in time domain
16.7	16.7	$X^1\Sigma_g^+$	
19.6	17.7	$A^2\Pi_u$	Overlap between peaks degrades accuracy
20.2	22.4	$a^3\Pi_u$	
23.6	23.7	$A^1\Pi_u$	

Changed:

Lines 460 – 474 now read: The pump and probe pulses used in serial femtosecond crystallography are typically tens of femtoseconds long. In principle, simulation can be used to investigate the effect of the pulse characteristics on the achievable time resolution. However, extensive studies of the spatial and temporal characteristics of incident pulses [43] has shown that it is extremely difficult, if not impossible, to determine the actual pulse characteristics in time-resolved serial femtosecond crystallography. We therefore follow a data-driven machine-learning approach, whereby the veracity of our results is determined by the extent to which our algorithm reproduces independently known ground-truths. Using experimental XFEL pump-probe data obtained with optical pulses as long as 75 fs, we have previously shown that the vibrational frequencies revealed by our approach accurately match those of well-known systems such as N_2, even when the incident pulse lengths are long compared with the vibrational frequencies [19]. Similarly, in the spectroscopically examined frequency range, each of the PYP vibrational frequencies revealed by our present work has been independently observed by time-resolved Raman techniques [33]. (See also Table S1). This clearly demonstrates that reliable dynamical information can be extracted with few-femtosecond time resolution.

Comment 2a:

It seems that a weakness of the machine learning approach is that there is no clear connection between the output of the analysis and the physical evolution of the system, i.e. how to combine the modes to build the actual trajectory of the system. The output of the algorithm needs to be fitted to simulated data in order to retrieve the trajectory through the conical intersection. Why is that? Isn't that information also encoded in the diffraction patterns?

Response:
As described in [Henry & Hofrichter, Methods in Enzymology, 210, 129 (1992)] and [Schmidt et al. Biophys J 84, 2112 (2003)], additional information is needed to reconstruct the physical evolution of a system from the modes obtained by standard (linear) singular value decomposition (SVD). This stems from noise attenuating the singular value (relative weight) of each mode by an unknown amount. As such, additional information (for example a kinetic model) is needed to reconstruct the physical behavior of the system. Fung et al. [Fung et al., Nature 532, 471 (2016)] demonstrate how this issue can be resolved for dynamical systems with minimal additional information.

Changes:
Lines 264 - 267 now read: As in standard singular value decomposition, additional information is needed to accomplish this task [Henry & Hofrichter (1992), Schmidt et al. (2003), Fung et al. (2016)]. Away from the conical intersection, we determine the appropriate mode combinations as previously described in [37], and extract structures from the resulting DED's as outlined in [35].

Comment 2b:
A related question, what is the physical significance of the very fast turning point in the chronos, as seen in Fig. 2a?

Response:
The sharp turning points reflect the rapid changes in each of the dynamical collective modes as the conical intersection is traversed. This assignment is now explicitly stated in the manuscript.

Changed:
Lines 170 - 171 read: Fourier analysis of the chronos before 615 fs, i.e., before the encounter with the conical intersection ff.

Comment 3a:
The simulation assumes a trajectory in a two-dimensional PES. As far as I understood, the machine learning algorithm output indicates that the dynamics takes place on a multidimensional space. What is the justification to use a two-dimensional PES to fit the data to?

Response:
We agree that the dynamics could indeed unfold on a multi-dimensional PES manifold. In principle, this space could be described by an effective Hamiltonian involving all experimentally observed modes. Such an approach would, however, substantially increase the number of free parameters. To circumvent this problem, we adopt a two-dimensional PES as the minimal model needed to describe the vicinity of the conical

intersection. This approximation significantly reduces the danger of overfitting, and enables us to solve the time-dependent Schrödinger equation for a sufficiently dense set of points in parameter space. The appropriateness of this approach is supported by the observation that all five pairwise combinations of experimentally determined structure-dynamical modes yield two-dimensional PES parameter-sets, which are the same to within uncertainty. This indicates that, in the vicinity of the conical intersection, all dynamical trajectories unfold on the same two-dimensional PES manifold. Further away from the conical intersection, our analysis reveals the presence of at least five distinct trajectories corresponding to different conduits to and from the vicinity of conical intersection.

Changed:
Lines 285 - 292 now read: The PES's are approximated by a second-order Taylor expansion, known as the vibronic coupling model [38,39,] and are assumed to be symmetric with respect to both normal modes, possibly differing in their respective vibrational frequencies. Inter-state coupling is mediated via one mode only. To determine the dynamics in the space spanned by two modes, we numerically solve the time-dependent Schrödinger equation for a wave packet initially occupying the excited electronic state.

Comment 3b:
The fitting procedure was not entirely clear to me. The manuscript states "we fit each of the simulated trajectories to the six experimental trajectories." But what exactly is meant here by fitting the trajectories? Is it the structures vs time, or is the machine learning algorithm also run on the simulation to extract the modes? The manuscript also states "In this, we consider both temporal and spatial translations of the simulated trajectories, and allow the simulated trajectories to be linearly scaled." What is meant by linear scaling of the trajectories, in time, in space? Why is this scaling of trajectories needed, and what is the physical significance of this scaling?

Response:
We apologize for the lack of clarity on this point. The purpose of identifying simulated trajectories with experimental ones is to determine the set of physical parameters best able to describe each of the experimental trajectories. This is performed by comparing simulated dynamical trajectories with experimental ones. In principle, one would simply select the simulated trajectories most closely resembling (have the smallest chi-squared value with respect to) each of the experimental trajectories. In practice, the axes describing the experimental and simulated trajectories may be rotated and/or rigidly shifted with respect to each other. Attenuation due to noise and timing jitter may also change the "length" of each experimental axis by an unknown amount. The best-fit search must therefore allow rigid shifts in the origin, axis rotations, and scaling of the simulated trajectories.

Changed:
Lines 631 – 641: A new section entitled "Identifying each experimental trajectory with a simulated counterpart" has been inserted to elucidate these points.

Comment 3c:
A more technical question: How is it possible to simulate 500k trajectories, when most quantum chemical simulation can only produce a few hundred trajectories in a reasonable time?

Response:
Most quantum chemical simulations are computationally expensive because their complexity increases rapidly with the number of degrees of freedom. We circumvent this problem by employing an effective two-modes model, which meets the minimal requirements of a conical intersection, and for which the time-dependent Schrödinger equation can be solved efficiently. This enables us to simulate a large number of trajectories, each of which requires only tractably small computational resources.
Changed:
Lines 273 - 276 now read: The calculations employ an effective model Hamiltonian in a compressed collective-mode space to capture the properties of a conical intersection and its vicinity. This allows numerically exact calculation of the nuclear quantum-dynamics as a function of a small number of molecule-specific model parameters.

Response to Reviewer #2

We are grateful to the distinguished Reviewer for the careful reading of our submission, and the insightful comments and questions. Below, we address each comment, and list the changes made in the light of that comment.

Comment 1:
I need to be convinced by how well the results relate to previous crystallographic, spectroscopic and computational work on the PYP system. Without this comparison it is unclear what specific new mechanistic insights have been obtained with the new method. Much of the manuscript reads like the introduction to a problem without a clear conclusion leaving me somewhat unsatisfied with the analysis. Before publication can be considered I recommend a major revision taking the specific recommendations and general guidelines below into account. I really want to believe such temporal resolution is possible, but more work needs to be put into the analysis and the manuscript.

Response:
We are grateful to the distinguished Reviewer for the opportunity to revise our submission in the light of her/his comments.

Comment 2:
Too often the manuscript replaces clear information with vague general statements. This imprecision already starts in the abstract where neither the protein under investigation nor the experimental method that generated the evaluated data is named. Instead the

authors chose the vague "2'000-atom protein" and "extracted from experimental data" where photoactive yellow protein and time-resolved serial crystallography would have been much more precise. I understand that the authors want to highlight the general applicability of their method but still the abstract and text needs to contain the specifics and needs rewriting. The introduction could be shorter and better targeted.

Response:
We regret the lack of specifics. We have revised manuscript to provide more specific details without sacrificing accessibility.

Changes:
Throughout.

Comment 3:
The title is somewhat misleading. Single-femtosecond resolution is likely overstated. Because of the physical limits of the experimental method it should be 5-10 fs at best (see discussion below on crystal size and homogenous excitation). Without refinements there is no atomic-resolution observation (see discussion on crystallographic evaluation below). And not the "2'000-atom protein" transverses a conical intersection but rather the 11 atoms of the p-coumaric acid. Of course, the chromophore is covalently bound to the protein but there is no description of the wider protein throughout the manuscript. I recommend to rephrase the title.

Response:
The title, abstract, and text have been changed to reflect a time resolution in the few-femtosecond regime. We agree that only a specific part of the chromophore undergoes major changes in traversing the conical intersection. However, in general, it is not possible to determine this region objectively without performing the analysis we have outlined. As it turns out, only a small region of the chromophore participates in the isomerization reaction, but the nature and extent of this region are a priori unknown. Indeed, although PYP has been the subject of study for decades, our results now reveal that "peripheral" side-chains host previously unanticipated charge oscillations, whose importance remains to be investigated. In sum, we agree that only a part of the chromophore actively participates in the isomerization process, but exactly which part must be identified by the type of whole-chromophore analysis we have presented.

Changes:
Throughout, including the title and abstract.

Comment 4:
Line 59 introduces p-cinnamic acid as the chromophore in PYP. This should be p-coumaric acid. The difference is only a singly hydroxy group, which however forms H-bonds with the protein and should not be neglected when describing the de-excitation

and isomerization mechanism. Likely this is just a typo though as the chromophore shown in Figure 1 is correct.

Response:

We thank the Reviewer for this correction, which has been implemented.

Changed:

Line 59: "p-cinnamic acid" has been changed to "p-coumaric acid".

Comment 5:

Further down in the abstract, the authors promise "single-femtosecond, atomic resolution movies". But in the end the 3D diffraction volumes obtained by the new method have not been used to define molecular structures to allow an atomistic interpretation. Figures S4 and S10 contain a comparison of electron density maps from conventional time-resolved crystallography and the new algorithm used in this paper. The similarity is convincing and provides validity to the new approach. But why were no structures refined and compared?

Response:

We apologize for this omission. Fig 2c now provides the refined cis structures at 800 fs, 900 fs and 3 ps. These were obtained as outlined in the Supplementary Information section 7 of [Fung et al., Nature 532, 471 (2016)] and in [Pandey et al. Nature Methods 17, 73 (2020)]. In principle, a movie of many refined structures can be compiled. That remains a future task.

Changed:

Fig. 2c now includes refined cis structures at 285 fs, 800 fs, 900 fs, and 3 ps. Additional references ([36] and [37]) have been provided. Finally, Table 2d provides the changes in the torsional angle with time. This angle is often used as a structurally intuitive isomerization reaction coordinate.

Comment 6:

Also, why only the 1.3 ps time delay which according to Figure S1 was not part of the data set obtained by subsampling of the experimental data. A series of electron density maps and a comparison of the refined structures at time delays that are discussed in the main manuscript and/or Pande et al would be much better. Both works have overlapping authors temporal resolution as a highlight of the current manuscript.

Response:

Unfortunately, "1.3 ps" was a typographical error, and should have read "3 ps". The "continuous" experimental data extend only to ~ 900 fs, with the 3-ps data forming a separate cluster in time. We have, however, applied our approach to the 3-ps cluster of jittered data, and the results are shown in Figs. 2c and S4.

Changed:

1.3 ps has been changed to 3 ps throughout.

Comment 7:
Another promise at the end of the abstract is the key parameters of the conical intersection and the de-excitation process. But again, I was not able to find these key parameters beside some poorly explained values in Table 1.

Response:
We regret that the role and importance of the key parameters were insufficiently clear. The physically important parameters are the vibrational frequencies of the potential energy surfaces, and the coupling strength, i.e., the transition probability between the participating electronic states. In combination, the vibrational frequencies and the coupling define the topography of the CI. A new paragraph elucidates the meaning and significance of these parameters.

Changed:
Lines 318 - 323 now read: Specifically, the parameter ω characterizes the kinetic energy, and (r, θ) the initial position of the wave packet, respectively. The shape of the potential energy surfaces is determined mainly by the vibrational frequencies γ_1, γ_2. The coupling parameter λ defines the probability of a transition between the two electronic states as a result of the wave packet moving through, or close to the conical intersection. Vibrational frequencies and coupling strength together define the topography of the conical intersection.

Comment 8:
When does the trans-cis isomerization happen according to the new data? Figure 2 seems to imply that the features at 615 fs in Chrono 2-5 resemble the isomerization, which would fit well to the ~550 fs in Pande et al. Several times throughout the manuscript it is mentioned that the features around the C3-C4 bonds have "previously" or "normally" been associated with the trans-to-cis isomerization. If the authors want to interpret these features differently please be clear and less ambiguous in wording.

Response:
The answer to the question regarding the "time of isomerization" depends somewhat on definitions. Based on the sharp changes in the topos and chronos of modes 1 - 5, isomerization occurs at 615 fs. The point of closest approach to the conical intersection (indicated by a red circle in Fig. 3c) is reached (25±3) fs later. As noted in the text, inter-state transition is a statistical process. Our estimates therefore represent weighted averages. The qualifiers "previously" and "normally" were used to clarify the consensus understanding of the isomerization process in PYP, not to call this consensus into question. In order not to confuse the issue, we have reduced the use of these qualifiers as far as possible.

Changes: Throughout, as appropriate.

Comment 8:
I don't understand why Mode 1 is not shown since it represents the moving average of all modes. As such it should be the most important and the electron density maps obtained by it should be most similar to the ones obtained by conventional analysis. If electron density maps of mode 1 alone can't be shown because modes always have to be used in combination please think about a clearer nomenclature, for example relabel the maps to M1+M2, M1+M3 etc.

Response:
As the Reviewer notes, Mode 1 represents a moving average over the data. As the excitation-induced changes are small compared with the mean, this mode is not particularly informative. However, Mode 1 has now been included in Figs. 2a, and S11. The Reviewer is correct that Mode 1 is added to all difference electron density maps. This point has been clarified.

Changed:
Line 883: Caption to Fig. 2 now makes it clear that Mode 1 has been added.
Lines 185 - 187 now read: As Mode 1 represents the moving average of the signal, its topo is added to each of the subsequent modes to facilitate comparison with DED's obtained by conventional means.

Comment 9:
On page 8-9 the modes 2-5 are compared in more detail. Please discuss how these modes relate to the reaction pathway.

Response:
The modes constitute the collective variables involved in isomerization via the conical intersection. These variables define the axes of the space in which the reaction pathways (aka dynamical trajectories) are traversed. Every point on the reaction pathway can thus be described in terms of the contribution of two collective variables.

Changed:
Lines 154 - 159 read: More specifically, each topo represents a characteristic difference electron density map, evolving in time as prescribed by its corresponding chrono (Fig. 2a). Each topo-chrono pair thus represents a characteristic dynamical mode of the charge distribution. In essence, these modes constitute the basis functions, which may be used in combination to describe the dynamical trajectories ("the reaction paths") of the system (see Fig. 3c, and below).

Comment 10a:
Photoactivation of PYP has a high quantum efficiency but still a significant proportion of the chromophore will relax back to the dark state while others will proceed towards the cis configuration within the investigated time frame. It might further be expected from

QM/MM simulations that not all chromophores move fully synchronized which could also be due to experimental limitations (see below). Please interpret the modes in a common model of PYP photoactivation.

Response:
As noted in response to the previous question, each mode represents an axis of the space in which a reaction pathway (dynamical trajectory) unfolds. Just as a dynamic trajectory in two-dimensional space can be defined as {x(t), y(t)}, a reaction pathway can be described as a succession of {x, y} values, thus combining the contribution of two collective variables. Our work identifies the collective variables relevant to the isomerization process. While the modes are mutually orthogonal, the choice of modes is not unique, and a single mode need not correspond to an experimental observable. The observables correspond to dynamical trajectories, rather than a specific mode. Our work is able to capture these trajectories on the femtosecond scale, and, in the vicinity of the conical intersection, relate them to population transfer between states (Fig. 3).

As shown in Fig. 2c, our results can also be used to determine refined structures, and thus make contact with traditional interpretations of PYP isomerization. Our work extends this possibility to the deep femtosecond range. The primary focus of the present paper, however, is the crossing of conical intersections.

Changed:
These points are clarified in lines 152ff.

Comment 10b:
Does the symmetry of Mode 5 for example reflect unproductive isomerizations?

Response:
Yes. The fact that chrono-5 starts and ends at about the same value indicates a collective variable corresponding to unproductive isomerization.

Changed:
Lines 249-251 now read: Taken together, the modes described above reveal the structure-dynamical motifs involved in PYP de-excitation, including ultrafast collective motions leading to successful and unsuccessful isomerization attempts.

Comment 10c:
What is the relevance of the rapid charge oscillations in Mode 2 or the oscillatory charge redistribution in mode 3?

Response:
We have highlighted these rapid charge oscillations in order to move away from the notion that the isomerization consists of a smooth transition from one well-ordered state to another. Our results show that the reality is more complex, involving, for example, rapid charge oscillations.

Comment 10d:
How can it be that the same sequence of events is repeated after 615 fs in Mode 4? The manuscript needs some justification and discussion with respect to the literature why these observations are relevant.

Response:
The chronos display two types of approximate symmetry about 615 fs: a) mirror symmetry; and, b) inversion symmetry. These symmetries necessarily entail repeating the same or reverse sequence of events after 615 fs as before this timepoint. As noted earlier, the dynamical trajectories, rather than the modes themselves, correspond to experimental observables. Our observations are potentially significant, because they elucidate, in unprecedented detail, the structure-dynamical trajectories of de-excitation via a conical intersection, and yield the properties of the relevant PES manifold.

Comment 10d:
In line 183 it should be Tyr98 instead of Tye98.

Response:
We are grateful for the correction.

Changed:
Appropriately corrected (line 203).

Comment 11a:
Please include a Supplementary figure and discussion of the achievable time resolution taking the experimental parameters under which the data was collected into account.

Response:
To assess the veracity of our approach, we rely on a data-driven machine-learning approach, whereby the veracity of our results is determined by the extent to which our algorithm reproduces independently known ground-truths. […], we have previously shown that the vibrational frequencies as high as 50 THz revealed by our approach accurately match those of well-known systems such as N_2, even when the incident pulse lengths are long compared with the vibrational frequencies [19]. Similarly, in the spectroscopically examined frequency range, each of the PYP vibrational frequencies revealed by our present work has been independently observed by time-resolved Raman techniques [33]. (See also table inserted below and Table S1). The table below clearly demonstrates that reliable dynamical information can be extracted with few-femtosecond time resolution. Theoretically, the achievable time resolution is determined by the total number of snapshots, the timing jitter and pixel noise, and the average time interval between snapshots. Application of the expression (8) of the Supplementary section 3 of Fung et al. to the present PYP dataset leads to a time resolution estimate in the sub-femtosecond regime. Such estimates, of course, represent the rarely achievable best case.

d **Vibrational Periods (fs)**

Measured	Theory	State	Remarks
15.2	15.1	$X^2\Sigma_g^+$	Previously observed in time domain
16.7	16.7	$X^1\Sigma_g^+$	
19.6	17.7	$A^2\Pi_u$	Overlap between peaks degrades accuracy
20.2	22.4	$a^3\Pi_u$	
23.6	23.7	$A^1\Pi_u$	

Changed:
Lines 465 – 474 now read: We therefore follow a data-driven machine-learning approach, whereby the veracity of our results is determined by the extent to which our algorithm reproduces independently known ground-truths. […] we have previously shown that the vibrational frequencies revealed by our approach accurately match those of well-known systems such as N_2 [19]. Similarly, in the spectroscopically well examined frequency range of about (3 – 30) THz, each of the PYP vibrational frequencies revealed by our present work has been independently observed by time-resolved Raman techniques [33] (see Table S1). This clearly demonstrates that reliable dynamical information can be extracted with few-femtosecond time resolution.

Comment 11b:
Beside the jitter mentioned in the main text, an important factor should be the size of the used crystals. Typical for such experiments is a crystal size of about 10 um which can be crossed by the activating pump laser in 33 fs or longer as the speed of light is reduced by ~25% in a watery environment. As the diffraction signal originates from an average of PYP molecules activated within this time frame there is a timing uncertainty inherent to the physical properties of the system. Even if PYP molecules in the first micrometer dominate the light signal because of the high optical density in a crystal, it is not clear to me how a time resolution of 1fs can be achieved as the described method is not sorting individual molecules.

Response:
We agree that the crystal size and illumination conditions likely play an important role. Some of these effects have been examined in detail in [Grünbein et al. Nat. Methods 17, 681 (2020)]. The primary message of such analyses is that it is extremely difficult, if not impossible, to characterize, let alone simulate the details of the experimentally achieved illumination conditions. As noted above, we therefore adopt a data-driven machine-learning approach, in which the reliability of our approach is determined by the extent to which it can reproduce independently known ground truths.

Changed:
Lines 170-178 now read: Fourier analysis of the chronos before 615 fs, i.e., before the encounter with the conical intersection brings multiple species into play, reveals three

prominent peaks at 4, 21, and 33.5 THz in the spectroscopically well explored region spanning the (3 – 30) THz range and its vicinity. These peaks have been previously observed by ultrafast time-domain Raman spectroscopy of PYP before the conical intersection 33. These spectroscopically measured PYP peak frequencies match those we observe to within 7% or better. (See Methods section entitled "Comparison with spectroscopically determined PYP frequency spectrum", and Table S1.) Such close agreement with independently known "ground truths" is the ultimate test of any machine learning approach.
Lines 1212 – 1214: A new Table S1 clarifies the above observations further.

Comment 11c:
Similarly, please discuss how the physical limitations on timing from using 40 fs XFEL pulses and 140 fs optical laser pulses can be overcome. Pushing the temporal resolution to extreme limits is the most important outcome from this work. But I need more evidence as given by the Frequency analysis in Fig S5. Is the system just so deterministic that faster times can be extracted? If yes would that still be an observation or is it already a simulation?

Response:
As the Reviewer notes, pump and probe pulses used in serial femtosecond crystallography are typically tens of femtoseconds long. In principle, simulation can be used to investigate the effect of the pulse characteristics on the achievable time resolution. However, extensive studies of the spatial and temporal characteristics of incident pulses [43] has shown that it is extremely difficult, if not impossible, to determine the actual pulse characteristics in time-resolved serial femtosecond crystallography. Using experimental XFEL pump-probe data obtained with optical pulses as long as 75 fs, we have previously shown that the vibrational frequencies revealed by our approach accurately match those of well-known systems such as N_2, even when the incident pulse lengths are long compared with the vibrational frequencies [19]. This clearly demonstrates that reliable dynamical information can be extracted with few-femtosecond time resolution. See also Response to Comment 11b above.

Changed:
Lines 459 – 474: A new section entitled "Validating the achieved time resolution" has been added to Methods.

Comment 12:
Please include a description of how the crystallographic snapshots have been sorted for Fig S1. Is this the same sorting as in Pandey et al., or have the timing tool signals also be reevaluated to increase timing accuracy in the input data to the current analysis?

Response:
The sorting is the same as in Pandey et al. The timestamps were not reevaluated in any way.

Comment 13:
In Line 367 it should be “In order to confirm…”

Response:
Thank you for the correction, which has been implemented (line 477).

Comment 14:
Figure 1: Please include the cis model of the chromophore. Also, the supplementary movies would benefit from containing the refined structures to guide interpretation of the density fluctuations.

Response:
Figs. 2c-d now provide refined structures and the evolution of the torsional angle with time. The compilation of refined structural movies across the entire time range remains a future task.

Changed:
Lines 894 – 903: Additional figure and table inserted as Figs. 2c and d, respectively.

Comment 15:
Figure 2b: Please include the contour level at which the maps are shown.

Response: The contour level is 3 sigma throughout.

Changes:
Throughout: the contour level has been specified as 3 sigma.

Comment 16:
Figure 3a: Please add a unit to the color bar.

Response:
Thank you for alerting us to the missing color-bar scale.

Changes:
Appropriate figures throughout: The color-bar unit has been specified to be fs.

Comment 17:
Figure 4: I don’t see how this Figure is relevant enough for a main figure of the manuscript. With such arbitrary axes it does poorly represent the new data that could be obtained with the method. Combining the figure with Table 1 or other relevant parameters that describe the conical intersection to a wider audience would provide more specific information.

Response:

We regret that Fig. 4 in its previous form was inadequate. The figure has been revised to include the evolution of the two state populations as the conical intersection is traversed. The caption now includes a description of the key parameters of the conical intersection in more accessible terms.

Changes:
New Fig. 4 shows the conical intersection and the evolution of the state populations. The new caption includes a more accessible description of the key parameters.

Comment 18:
Table 1: Please add units to the different parameters. Please explain or highlight the relevance of these parameters, if possible, in Figure 4 to make it more meaningful. Can some of these values be cross-checked with literature values obtained with orthogonal techniques?

Response:
We thank the Reviewer for these suggestions. Units have been added. To the best of our knowledge, there are no comparable studies on PYP.

Changes:
Unites have been added to the parameters of Table 1. Figure 4 has been revised.

Comment 19:
Figure S1: Please include reference to the data source. Please include 1.3 ps data used for Figure S4.

Response:
Thank you for the suggestions. Data sources have been provided. As noted earlier, "1.3 ps" was a typographical error, which has been corrected to "3 ps". The 3 ps data lie outside the time window of Fig. S4, and their inclusion would significantly compress the most pertinent section of the data along the time axis. We therefor ask for the Reviewer's indulgence.

Change:
Data sources provided in line 455ff. 1.3 ps has been changed to 3 ps throughout.

Comment 20:
Figure S3: For simplicity, please mention in the figure legend to which step synthetic and output 3D diffraction volumes belong in Figure S2.

Response:
Thank you for the suggestion.

Change:

Lines 1122 – 1125: Legend to Fig. S3 now includes the following: Pearson correlation and R-factor between synthetic (input) and output 3D diffraction volumes obtained from step 7 of Fig. S2.

Comment 21:
Figure S4: Please include panels presenting more than the chromophore region, so that noise levels in the maps can be compared. Please include the contour levels at which the maps are shown. Could be combined with Fig S10 to provide a convincing case for the validity of the methodology, especially if earlier time delays can be included as mentioned above.
Response:
Thank you for the suggestions.

Changes:
Fig. S4 now presents more than the immediate chromophore region. Fig. S4 caption specifies the contour level is 3 sigma. Figs. S4 and S10 have been combined into new Fig. S4.

Comment 22:
Please provide resolution limits in Å instead of 1/Å to increase readability.

Response:
Thank you for the suggestion.

Changed:
Resolution limits given in Å.

Comment 23:
The material and methods section mentioned on line 354 that also the dark data yielded a single mode from the NLSA analysis. A supplementary figure comparing this dark mode the light modes on the same scale may act as convincing control for the technique, especially if no high frequency components are present.

Response:
Thank you for the suggestion. In Fig. S11, the modes obtained from the light data are compared with the modes obtained from the dark data. The first two "dark" chronos are identical, except for scale. This is a hallmark of a one-parameter process. Correlation analysis shows the single-parameter process correlates with the integrated Bragg spot intensity (Pearson correlation: 0.93), most likely due to slight drift in the incident beam intensity. The subsequent dark chronos represent noise.

Please note, however, that the dark data were not recorded as a function of pump-probe delay. The horizontal axis for the dark data in essence represents the nominal time elapsed since the start of the experiment, while the horizontal axis for the light data represents the pump-probe delay.

Changes:
Lines 433 – 436 now read: Applying the same preprocessing steps to the dark snapshots results in a data matrix of $D = 15{,}487$ rows and $N = 79{,}937$ columns. As no optical pump, and hence no pump-probe time delay was involved in recording dark snapshots, these data were lexicographically sorted according to run numbers followed by event numbers.

New Fig. S11 inserted with the following captions (lines 1204 – 1209): Fig. S11 | a, The first five chronos obtained from light data ordered according to pump-probe delay. b, The first five chronos obtained from dark data lexicographically sorted according to run numbers followed by event numbers. The first two chronos are identical, except for scale. This is a hallmark of a one-parameter process. Correlation analysis shows the single-parameter process correlates with the integrated Bragg spot intensity (Pearson correlation: 0.93), most likely pertaining to drift in the incident beam intensity. The subsequent chronos represent noise.

Comment 24:
From line 529 onwards the authors explain how they fitted experimental trajectories to their experimental counterparts. This leaves me puzzled because (line 65-67) in the introduction the authors highlight that electronic structure calculations of PYP are unfeasible for the foreseeable future due to its size. However only the chromophore excitation has to be modeled and is relevant in the ultrafast time domain unless the authors found other significant motions further out in the protein as already mentioned above.
Response: We agree with the Reviewer that the photoisomerization reaction in PYP involves a relatively small region of the protein. Of course, such information is not available for a new system. More importantly, our analysis demonstrates that a few collective variables can describe a complex dynamical system, with each collective variable representing a large ensemble of atoms. Our effective-Hamiltonian approach is applicable to all systems, which can be described in terms of a small number of collective modes, regardless of the total number of atoms involved. This means that our approach can, in principle, accurately model complex dynamical processes by identifying a small number of collective variables each involving a large number of atoms.

Changed:
Lines 273-276 read: The calculations employ an effective model Hamiltonian in a compressed collective-mode space to capture the properties of a conical intersection and its vicinity. This allows numerically exact calculation of the nuclear quantum-dynamics as a function of a small number of molecule-specific model parameters.

Response to Reviewer #3
We are grateful to the distinguished Reviewer for finding our algorithmic approach sound. We agree that the approach we use is different from Geometric *Deep* Learning.

However, our approach is based on the geometric properties of the manifold learned from the data. We therefore ask for the Reviewer's indulgence to use the term Geometric Machine Learning as we have in the past.

Reviewer Reports on the Second Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors have done a careful job in responding to my earlier comments. All in all, they have provided sufficient evidence to show that the method works in retrieving information about the dynamics and clarified several points, however, there is not sufficient evidence to back the claim that structural dynamics at or below 10 fs can be retrieved.
1a. The sharp turning point in the Cronos, with a width of 2 fs is interpreted as a marker for the wavepacket going through the conical intersection. However, if the wavepacket is created with a 140 fs laser pulse, then it should take around 140 fs for it to go through the conical intersection. It is still not clear what is the physical picture in which excitation with a 140 fs laser pulse could create a sharp 2 fs feature in time. In terms of vibrational frequencies, the authors reinforce their point by comparing with reference 33 (Kuramochi et al), unfortunately no vibrational with a period shorter than 30 fs are available to be compared.
1b. The authors have chosen not to include the duration of the excitation laser pulse in the synthetic data that they use to verify the performance of the algorithm. The reason that they cite is the difficulty in determining the pulse characteristics in femtosecond crystallography experiments. I disagree with the reason given. All that is needed to make a reasonable and valuable simulation is to include the duration of the excitation laser pulse, which is very easily measured, and in fact cited in the original study of this molecule. Without this, the main question of whether sub-10 fs dynamics can be extracted after excitation with a 140 fs laser pulse cannot be addressed.
All of my comments have been adequately addressed by the authors.

Referee #2 (Remarks to the Author):

Thank you for the revised version of your manuscript "Few-femtosecond atomic-resolution observation of a protein traversing a conical intersection". It is much improved and I am satisfied with the implemented changes. It would be great to see the manuscript published in Nature.

Reducing the claim from 1fs time-resolution to Few-femtoseconds circumvents many of the uncertainties in terms of pulse length and timing without compromising the main claim of the manuscript to improve time-resolution by re-evaluating the data using geometric machine learning. I would have liked to see a more thorough discussion on how the pulse length and crystal size effects the achievable time resolution but understand that this might be beyond the scope of this manuscript focusing more on the data evaluation procedure.

The revised version is also much more precise in its comparisons to the initial claims in Pande et al. and brings the results into better context with what is known on PYP. Personally, I am less interested in the descriptions of the conical intersections and am not sure to what extent the principles can be generalized to other systems like retinal, DNA or azobenzenes. But the technology is in place now to find out and I am sure many will want to make use of it to extract the maximum possible out of their ultrafast TR-SFX data. Certainly, the general discussion will be of interest to many readers given the broad scope of Nature.

A minor change

Line 133: "pertaining to a timepoint determined with an accuracy of about 1 fs."

Please change to "few fs" to be consistent with title and the rest of the manuscript.

Author Rebuttals to Second Revision:

Response to Reviewer Comments

Reviewer #1

We are grateful to the distinguished Reviewer, and appreciate her/his comments.

Comment:
The authors have done a careful job in responding to my earlier comments. All in all, they have provided sufficient evidence to show that the method works in retrieving information about the dynamics and clarified several points,
Response:
Thank you.

Comment:
There is not sufficient evidence to back the claim that structural dynamics at or below 10 fs can be retrieved.
Response:
We have clarified that the claim of structural dynamics at or below 10 fs is based on the following four observations.

1. Accurate recovery of femtosecond dynamics from a synthetic input signal resulting in an input/output correlation coefficient of better than 0.99, and an R factor below 15%. (See lines 454 – 463 of manuscript.)
2. Accurate recovery of independently known femtosecond dynamics from pump-probe experimental data of N_2 obtained using a 70fs long pump pulse (Ref. 37).
3. Clear observation of a high-SNR oscillatory signal at 95 THz (10.5 fs) in PYP, indicating, via the Shannon Nyquist sampling theorem (Ref. 32), a time resolution of about 1 fs. (See lines 162-167 of manuscript.)
4. Accurate validation of the recovered oscillatory signals by oscillations independently recovered from PYP by time-resolved Raman spectroscopy, up to the highest frequency accessed spectroscopically. (See Table S1 of manuscript.)
5. As noted in lines 179 – 181 of the manuscript, the recovery of the 33.5 THz (30 fs) oscillatory signal by both Raman spectroscopy and by our data-driven technique already validates a time resolution better than 10 fs. Specifically, Shannon-Nyquist sampling of a noisy jittered signal (Ref. 32) requires the ability to sample the signal at ~ 5x – 10x the signal frequency, i.e., a time resolution of (3 – 6) fs. (See lines 179 – 181 of manuscript.)

We respectfully submit that the above observations constitute sufficient evidence for a time resolution of a few femtoseconds.

Comment:
1a. The sharp turning point in the Cronos, with a width of 2 fs is interpreted as a marker for the wave packet going through the conical intersection.
Response:
The revised paper no longer uses the sharp turning points to estimate the time resolution.

Comment:
1b. The authors have chosen not to include the duration of the exciting laser pulse in the synthetic data that they use to verify the performance of the algorithm. The reason that they cite is the difficulty in determining the pulse characteristics in femtosecond crystallography experiments. I disagree with the reason given. All that is needed to make a reasonable and valuable simulation is to include the duration of the excitation laser pulse, which is very easily
measured, and in fact cited in the original study of this molecule.
Response:
We agree with the Reviewer that the pump laser pulse can be accurately measured ex situ. However, as shown in detail in Ref. 43 of the manuscript, the spatial (and hence temporal) profiles of the optical excitation reaching the protein microcrystals suspended in a liquid jet are much more complex, and poorly known. Even if these complications could be circumvented, the assumption of a linear mapping of the temporal shape of the excitation pulse to the time-resolved diffraction signal obtained by serial femtosecond crystallography has been called into question, and remains to be resolved (Ref. 43). These issues are well beyond the scope of our paper.

Reviewer Reports on the Third Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors have clarified, and I do not dispute, that the analysis brings up features as fast as 10 fs from the data. My main issue was with how this features could be generated with an excitation laser pulse that is much longer than 10 fs. I think the last point by the author sheds some light on this:
"We agree with the Reviewer that the pump laser pulse can be accurately measured ex situ. However, as shown in detail in Ref. 43 of the manuscript, the spatial (and hence temporal) profiles of the optical excitation reaching the protein microcrystals suspended in a liquid jet are much more complex, and poorly known. Even if these complications could be circumvented, the assumption of a linear mapping of the temporal shape of the excitation pulse to the timeresolved diffraction signal obtained by serial femtosecond crystallography has been called into question, and remains to be resolved (Ref. 43). These issues are well beyond the scope of our paper."
The manuscript referenced by the authors "Gruenbein et al., Nature Methods 17, 681 (2020)," states that under the experimental conditions for the excitation each molecule absorbed more than 10 photons. I believe this is the key point. Under a multi-photon excitation it would be reasonable to expect that frequency components much faster than the duration of the laser pulse can be triggered and observed. If this is what he authors are implying then I would be OK with this. But this needs to be more explicitly stated and described in the manuscript.
If I understood correctly, the capability of the method to extract dynamics faster than the duration of the excitation pulse relies on the nonlinearity (multi-photon) nature of the excitation. This is an

important limitation of the work, since in many investigations of time-resolved biological and chemical dynamics it is essential to stay in the range of 1-photon excitation. The results are nevertheless impressive and I would support publication of the manuscript, but it is imperative to clearly state under what condition one can expect to extract dynamics faster than the excitation pulse. If my interpretation is incorrect then I ask the authors to clarify under what conditions one can retrieve dynamics faster than the excitation pulse. Clarifying this issue would only enhance the impact of the work since it would be easier for researchers to determine how best to design their experiments to take advantage of this method.

Author Rebuttals to Third Revision:

Response to Reviews
Aug. 5, 2021

We have addressed the remaining reviewer comments below, taking care that no peer reviewed data are removed altogether when making these changes.

REQUEST: Reviewer #1 asks to much better clarify in the manuscript the process by which components faster than the pulse can be triggered and observed. Please address this comment, which was also raised by Reviewer #2 previously.
RESPONSE: We thank the reviewer for the opportunity to clarify this issue. Lines 165-174 now read: "Fourier analysis of the chronos by multi-taper methods [27,28] reveals the clear presence of frequencies of up to 95 THz (10.5 fs) at signal-to-noise ratios of ~ 5 or higher (Supplementary Fig. 5). As the observation of a frequency component in an initially non-uniform set of timepoints requires a time resolution ~ 5 – 10 times shorter than the period of the component [29], the clear observation of a 10.5 fs signal validates the few-femtosecond time resolution of our approach. This high time-resolution is particularly remarkable, because the data were obtained with a 140-fs optical pump pulse [10]. This suggests that in the experiment, a temporal gating effect may have played a role, as a consequence of, e.g., nonlinear multiphoton processes [30], or light-induced structural disorder [31]." Please also note the additional reference [31].

REQUEST: Reviewer #3 originally mentioned that the term 'geometric machine learning' was not accurate in this context. Please simply remove 'geometric'.
RESPONSE: We have replaced "geometric" machine learning with "manifold-based" machine learning in the first instance, and with "machine learning" subsequently.