# RESEARCH ARTICLE | JUNE 27 2023

# Projection to extract the perpendicular component (PEPC) method for extracting kinetics from time-resolved data

H. Ki 💿 ; J. Gu 💿 ; Y. Cha 💿 ; K. W. Lee 💿 ; H. Ihee 💌 💿

( Check for updates

*Struct Dyn* 10, 034103 (2023) https://doi.org/10.1063/4.0000189







Advance your science and career as a member of the AMERICAN CRYSTALLOGRAPHIC ASSOCIATION



AIP Publishing

# Projection to extract the perpendicular component (PEPC) method for extracting kinetics from time-resolved data

Cite as: Struct. Dyn. **10**, 034103 (2023); doi: 10.1063/4.0000189 Submitted: 6 April 2023 · Accepted: 22 May 2023 · Published Online: 27 June 2023

H. Ki,<sup>1,2</sup> 🕞 J. Gu,<sup>1,2</sup> 🕞 Y. Cha,<sup>1,2</sup> 🕞 K. W. Lee,<sup>1,2</sup> 🕞 and H. Ihee<sup>1,2,a)</sup> 🕞

### AFFILIATIONS

<sup>1</sup>Department of Chemistry and KI for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

<sup>2</sup>Center for Advanced Reaction Dynamics, Institute for Basic Science, Daejeon 34141, Republic of Korea

<sup>a)</sup>Author to whom correspondence should be addressed: hyotcherl.ihee@kaist.ac.kr

### ABSTRACT

Time-resolved x-ray liquidography (TRXL) is a potent method for investigating the structural dynamics of chemical and biological reactions in the liquid phase. It has enabled the extraction of detailed structural aspects of various dynamic processes, the molecular structures of intermediates, and kinetics of reactions across a wide range of systems, from small molecules to proteins and nanoparticles. Proper data analysis is key to extracting the information of the kinetics and structural dynamics of the studied system encrypted in the TRXL data. In typical TRXL data, the signals from solute scattering, solvent scattering, and solute–solvent cross scattering are mixed in the *q*-space, and the solute kinetics and solvent hydrodynamics are mixed in the time domain, thus complicating the data analysis. Various methods developed so far generally require prior knowledge of the molecular structures of candidate species involved in the reaction. Because such information is often unavailable, a typical data analysis often involves tedious trial and error. To remedy this situation, we have developed a method named projection to extract the perpendicular component (PEPC), capable of removing the contribution of solvent kinetics from TRXL data. The resulting data then contain only the solute kinetics, and, thus, the solute kinetics can be easily determined. Once the solute kinetics is determined, the subsequent data analysis to extract the structural information can be performed with drastically improved convenience. The application of the PEPC method is demonstrated with TRXL data from the photochemistry of two molecular systems:  $[Au(CN)_2^-]_3$  in water and CHI<sub>3</sub> in cyclohexane.

© 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http:// creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/4.0000189

# I. INTRODUCTION

Time-resolved x-ray liquidography (TRXL), also known as timeresolved x-ray solution scattering, has been established as a useful method for studying the structural dynamics and kinetics of chemical and biological reactions in the liquid solution phase.<sup>1–5</sup> Molecular structures of reaction intermediates and their kinetics have been investigated with TRXL at third generation synchrotrons for a variety of molecules spanning from diatomic molecules to proteins and nanoparticles.<sup>6–28</sup> Thanks to the recent development of x-ray free-electron lasers (XFELs), the time resolution of TRXL has reached the sub-picosecond regime<sup>29–38</sup> and has even allowed for tracking time-dependent positions of nuclear wavepackets.<sup>34,36,39</sup> Although the development of x-ray sources and facilities has greatly improved the time resolution and signal-tonoise ratio (SNR) of TRXL data, it is the development of appropriate data analysis methods that has enabled the expansion of the depth and breadth of information extracted from TRXL data.

The key step in the analysis of TRXL data is to determine the kinetics and structures of reaction intermediates, and many useful analysis methodologies have been developed.<sup>40–48</sup> In typical TRXL data, the signals from solute scattering, solvent scattering, and solute–solvent cross scattering are mixed in the *q*-space,<sup>43</sup> and the solute kinetics and solvent hydrodynamics (or solvent kinetics) are mixed in the time domain. For this reason, the data analysis of TRXL data is not trivial. Various methods that have been developed so far generally require prior knowledge of the molecular structures of candidate species involved in the reaction.<sup>7,10,13,43,49</sup> Since such information is often unavailable; a typical data analysis often involves tedious trial and error.

View Onlin

Export Citatio

In this work, we report a method to circumvent this situation. In this method named projection to extract the perpendicular component (PEPC), we take advantage of the fact that the scattering profiles of the solvent term are well-known. We show that by subtracting a scaled portion of the known solvent term from the TRXL signals, the resulting signals, i.e., the PEPC-treated signals, exhibit no contribution from the solvent kinetics, although they retain the intact solute kinetics. The absence of solvent kinetics in the PEPC-treated signals allows the straightforward determination of the solute kinetics without any prior knowledge of the molecular structures of the participating solute species.

# II. RESULTS AND DISCUSSION

#### A. TRXL data

Usually, experimental difference scattering curves,  $\Delta S(q, t)_{exp}$ , obtained by subtracting the scattering curve at a negative time delay  $(t_{ref})$  from the scattering curve at a positive time delay (*t*) of interest, as given in Eq. (1), are subject to data analysis,

$$\Delta S(q, t)_{exp} = S(q, t)_{exp} - S(q, t_{ref})_{exp}.$$
 (1)

To extract the kinetics and structural dynamics information from  $\Delta S(q, t)_{exp}$ , the theoretical  $\Delta S(q, t)$  and  $\Delta S(q, t)_{theory}$  need to be calculated and compared with  $\Delta S(q, t)_{exp}$ . For the purpose of analysis, we categorize the total scattering signal into the following three terms based on whether the atom belongs to the solute or solvent molecules. Accordingly,  $\Delta S(q, t)_{theory}$  comprises three components: (i) solute-only term [ $\Delta S(q, t)_{solute}$ ], (ii) solute–solvent cross term [ $\Delta S(q, t)_{cage}$ ], and (iii) solvent-only term [ $\Delta S(q, t)_{solvent}$ ], as given in the following equation:<sup>43</sup>

$$\Delta S(q, t)_{\text{theory}} = \Delta S(q, t)_{\text{solute}} + \Delta S(q, t)_{\text{cage}} + \Delta S(q, t)_{\text{solvent}}.$$
 (2)

The solute-only term contains information about the change in the molecular structure of solute molecules, whose concentrations change according to the solute kinetics. The cage term contains information about the change in the cage structure according to the solute kinetics. The solvent-only term contains the changes in the arrangement of solvent molecules in the bulk solvent, mainly due to the changes in temperature and density of the solvent, which are caused because the photon energy absorbed by the solute molecules is transferred to the solvent molecules. The solute-only term can be calculated using the Debye equation. For this calculation, the molecular structure of the solute molecules needs to be known. Often, the structure from density functional theory (DFT) calculations is used as the starting structure, and the structural parameters can be refined through a fitting process that employs those structural parameters as fitting parameters. The cage term can be calculated using pair distribution functions, g(r), from molecular dynamics (MD) simulations. The sum of the soluteonly term and the solute-solvent cross term can be categorized into the solute-related term, as follows:

$$\Delta S(q, t)_{\text{sol-rel}} = \Delta S(q, t)_{\text{solute}} + \Delta S(q, t)_{\text{cage}}, \quad (3)$$

then,  $\Delta S(q, t)_{\text{theory}}$  can be written as the sum of  $\Delta S(q, t)_{\text{sol-rel}}$  and  $\Delta S(q, t)_{\text{solvent}}$ 

$$\Delta S(q, t)_{\text{theory}} = \Delta S(q, t)_{\text{sol-rel}} + \Delta S(q, t)_{\text{solvent}}.$$
 (4)

The  $\Delta S(q, t)_{sol-rel}$  term can be expressed as the sum of the contributions of relevant solute species, i.e., the sum of difference scattering

curves of different solute species. More specifically, it can be written as follows:

$$\Delta S(q, t)_{\text{sol-rel}} = 1/R \times \sum_{k} (f_k(t) \times \text{SADS}_k(q)),$$
(5)

where *R* is the ratio of the number of solvent molecules with respect to that of solute molecules,  $f_k(t)$  is the molar fraction of the *k*th solute species among the total solute molecules at time *t*, and SADS<sub>k</sub>(*q*) is the solute-related term per unit concentration of the *k*th solute species. SADS stands for the species-associated difference scattering curve. As shown in Eq. (5), in a typical TRXL analysis, the solute-related term is multiplied by the factor of 1/R so that the term is scaled to indicate the amplitude of the signal per a solvent molecule, not per a solute molecule. In principle, Eq. (5) is applicable to general structural changes, even including the case of continuous structural changes. For such a continuous structural change, the structure at each time point would have its own SADS<sub>k</sub>(*q*) with its corresponding  $f_k(t)$  be a delta function. SADS<sub>k</sub>(*q*) can be expressed as follows:

$$SADS_k(q) = \Delta S_k(q)_{solute} + \Delta S_k(q)_{cage}, \qquad (6)$$

$$SADS_k(q) = S_k(q)_{solute} - S_0(q)_{solute} + S_k(q)_{cage} - S_0(q)_{cage}, \quad (7)$$

where  $\Delta S_k(q)_{solute}$  and  $\Delta S_k(q)_{cage}$  are the difference solute-only term and the difference solute-solvent cross term of the *k*th solute species, respectively;  $S_k(q)_{solute}$  and  $S_k(q)_{cage}$  are the solute-only term and the solute-solvent cross term of the kth solute species, respectively; and  $S_0(q)_{solute}$  and  $S_0(q)_{cage}$  are the solute-only term and the solute-solvent cross term of reactants, respectively.

Actually, as expressed in Eqs. (3) and (6), the solvent cages surrounding the solute molecules, as well as the solute molecules, contribute to  $\Delta S(q, t)_{sol-reb}$  thus to its component,  $SADS_k(q)$ . Here, for the simplicity of explanation, the term "solute species" is used as a term that includes both a solute species and its surrounding cage structure. We note that the kinetics of surrounding cage structure and solute molecules can be different, particularly at early time delays when non-equilibrium cage structures can be observed. For such a case, in Eq. (5), a large number of  $SADS_k(q)$ 's are required to describe the continuous structural changes accompanying the non-equilibrium dynamics. Also the corresponding  $f_k(t)$ 's would show more complex behavior on the *t*-axis to describe the kinetics of both solute molecules and surrounding cage structure, which are independent of each other.

The solvent-only term,  $\Delta S(q, t)_{solvent}$  given in Eq. (8), consists of two basis differentials  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_{T}$ ,<sup>43,49</sup> which can be obtained in a separate solvent heating experiment either by exciting the dye molecules in a dye solution<sup>50</sup> or by directly exciting the vibrational overtone of the solvent molecules in a neat solvent using near-infrared excitation,<sup>51</sup>

$$\Delta S(q, t)_{\text{solvent}} = \Delta T(t) \times (\partial S/\partial T)_{\rho} + \Delta \rho(t) \times (\partial S/\partial \rho)_{\text{T}}.$$
 (8)

In this equation,  $(\partial S/\partial T)_{\rho}$  is the change of the solvent scattering intensity in response to a temperature change at a constant density,  $(\partial S/\partial \rho)_T$  is the change of the solvent scattering intensity in response to a density change at a constant temperature, and  $\Delta T(t)$  and  $\Delta \rho(t)$  are the time-dependent changes in the temperature and density of the solvent at the time delay *t*, respectively.

A combination of Eqs. (4), (5), and (8) gives the following relationship:  $^{43,49}$ 

First of all, the concentrations,  $f_k(t)$ , are not independent of one another. Instead, they should obey the kinetic rate equations of the reactions. Second, the total energy of the system should be conserved. The sum of the electronic and vibrational energy of various solute spe-

$$\Delta S(q, t)_{\text{theory}} = 1/R \times \sum_{k} (f_k(t) \times \text{SADS}_k(q)) + \Delta T(t) \\ \times (\partial S/\partial T)_{\rho} + \Delta \rho(t) \times (\partial S/\partial \rho)_{T}.$$
(9)

The goal of data analysis is to determine (i) the kinetic framework of the reaction and related rate constants, which are manifested in  $f_k(t)$ , and (ii) the molecular structures of reaction intermediates, which are manifested in SADS<sub>k</sub>(q) that generate  $\Delta S(q, t)_{\text{theory}}$  in the best agreement with  $\Delta S(q, t)_{\text{exp}}$ .

#### B. Previous data analysis methods

In general, two types of information are desired to be extracted: (i) kinetics and (ii) molecular structures. The kinetic information, which includes the number of participating species and their associated rate constants, can also be obtained from time-resolved optical or vibrational spectroscopies. Once the kinetic information is obtained, the next step is to extract the structural information hidden in the TRXL data. The conventional analysis protocols for TRXL data can be classified into three methods: (i) linear combination fitting (LCF) with iterative least squares minimization (or iterative weighted least squares minimization), (ii) non-orthogonal decomposition (NOD), which is LCF with the direct method, and (iii) global fitting analysis (GFA).

LCF<sup>52</sup> is one of the most straightforward approaches. In LCF, the difference curve at each time delay is fitted separately with a theoretical curve calculated using Eq. (9). Although  $f_k(t)$ s are linked to  $\Delta T(t)$  and  $\Delta \rho(t)$  via the energy conservation and hydrodynamics equation,<sup>44</sup> their relationships are ignored. Instead,  $f_k(t)$ ,  $\Delta T(t)$ , and  $\Delta \rho(t)$  are treated as independent fitting parameters. In the fitting, the chi-square  $(\chi^2)$  value representing the discrepancy between the experimental and calculated data is minimized via optimizing the fitting parameters, which are simply the proportionality coefficients of the contributing terms. Plotting the obtained proportionality coefficients of the solute terms, which correspond to  $f_k(t)$ s, as a function of time displays the kinetic profiles of the solute molecules, and plotting those of the solvent heating terms, which correspond to  $\Delta T(t)$  and  $\Delta \rho(t)$ , as a function of time delays displays how the temperature and density of the solvent changes with time, i.e., the hydrodynamic response of the solvent. This method requires prior knowledge on the identities of the solute molecules and their molecular structures because it is required to calculate  $SADS_k(q)$ 's based on the identities and structures for the fitting. If the solute species are not known, all plausible candidate species should be tested.4

In the LCF method, the least squares refinement is performed for the difference curve at each time delay. Because the difference curve is a linear combination of various contributing terms, one can use linear algebra to directly determine the proportionality coefficients (direct method).<sup>53,54</sup> In this way, the iterative least squares refinement can be bypassed. To distinguish this method from the LCF with the iterative minimization, we name the LCF with the direct method as "NOD."<sup>47</sup> In both LCF and NOD, the structural parameters of the candidate solute species can be varied, and in this way, the structure can also be refined. In terms of time and accuracy, NOD is advantageous over LCF with the least squares minimization, which always has a risk of being trapped in local minima.

In both LCF and NOD, the difference curve at each time delay is treated separately, with no constraints between the difference curves at different time delays. In reality, there are at least two constraints.

cies, the energy released through radiative emissions, and the thermal energy dissipated to the bulk solvent should be conserved throughout the progress of the photoreaction via energy conservation. In GFA,<sup>9,10,20,30,43</sup> these constraints can be taken advantage of by fitting the difference curves at all time delays simultaneously. Naturally, the fitting parameters are not the proportionality coefficients,  $f_k(t)$ ,  $\Delta T(t)$ , and  $\Delta \rho(t)$ , at individual time delays, but the rate constants and some parameters required to calculate the change in the temperature and density of the solvent (such as the relative energy values of each chemical species participating in the reaction, thermodynamic parameters of the solvent, and the quantum yield of the reaction). In the fitting, the sum of the  $\chi^2$  values at all time delays is minimized. Consequently, the total number of fitting parameters is greatly reduced. Nevertheless, this method requires prior knowledge of the solute species because  $SADS_k(q)$ s are needed for the fitting, as in LCF and NOD. Moreover, a kinetic model needs to be assumed, and many such candidate kinetic models need to be tested via trial and error. The kinetic profiles from preliminary LCF or NOD analyses can provide a starting point for a plausible kinetic model. C. Drawback of previous data analysis methods In all three methods discussed in Sec. II B, the theoretical scattering curves of solute molecules, which are used to calculate  $SADS_k(q)s$ , need to be known or at least be assumed. In some cases, the first guess of the plausible reaction pathways and the DFT structures of the corre-

of the plausible reaction pathways and the DFT structures of the corresponding reaction intermediates work properly. Nevertheless, as the unknowns and the complexity of the studied molecular system increase, a more systematic approach is required. The simplest and most straightforward way is to test all plausible scenarios of kinetic frameworks and structures. Such an approach often works suitably, but a more efficient approach is desirable.

At this point, it should be noted that the data analysis protocol for TRXL data from proteins is quite different from that for small molecules. For the protein data, the kinetics can be first determined using a method to extract the kinetic framework, even without knowing the structures of the protein intermediates, by treating the TRXL data as if they are time-resolved spectroscopic data.<sup>15,16,27,38,48</sup> Then, the resulting species-associated difference scattering curves (SADSs) consistent with the determined kinetics are subject to the subsequent structural analysis. The kinetics of proteins can be more complicated than those of small molecules. For example, the K30D mutant of homodimeric hemoglobin (HbI) exhibits one of the most complicated kinetic frameworks studied so far with TRXL.<sup>48</sup> Nevertheless, the kinetics can be determined owing to various methods developed for this purpose, such as singular value decomposition (SVD) with reduced time ranges (SVD-RTR)<sup>48</sup> and SVD-aided pseudo-principal-component analysis (SAPPA).<sup>45</sup> The reason why the kinetics can be readily determined without any prior knowledge or guess of the molecular structures of the solute species is as follows. In the case of the protein TRXL data, the solvent term can be easily subtracted from the experimental data to yield the solute-related term because the solvent term appears in a high-q region where the solute-related term does not show significant contribution in the high-q region.<sup>15,16,48</sup> A ramification from this

consideration of the protein TRXL data is that even for the smallmolecule data, if the solvent term can be removed, the kinetics can be readily determined from the resulting solute-related data. Toward this goal, we reported the SVD-aided non-orthogonal decomposition (SANOD) method,<sup>47</sup> where the components from singular value decomposition (SVD)<sup>55–57</sup> are used as the terms to be used in NOD instead of the calculated solute terms, circumventing the requirement to have prior knowledge of the solute species and their structures. The fact that the solvent term is known greatly facilitates SANOD. Nevertheless, to use the SANOD method, all necessary components need to be assembled, making its application not straightforward.

# D. A new method: Projection to extract the perpendicular component (PEPC)

In the discussion so far, it has been shown that the solute kinetic model can be easily determined if the solvent term can be removed from the TRXL data. Therefore, one straightforward approach would be to remove the solvent term. However, it is generally not possible to completely remove the contribution of the solvent term from the TRXL data even if the shape of the solvent term is known. To overcome this issue, our new strategy is to remove the solvent "kinetics" from the TRXL data instead of trying to remove the solvent term itself from the TRXL data. The schematic diagram of this new strategy is illustrated in Fig. 1(a). As long as the solvent kinetics can be removed without altering the solute kinetics, the solute kinetics can be readily determined from the resulting data as demonstrated in the case of the K30D HbI,<sup>48</sup> as explained in Sec. II C. In the following, we first prove that such a task not only is possible but also can be performed with a simple mathematical operation based on vector calculation.

To explain this concept, let us first consider TRXL data at a given time delay as a vector. For example, if the number of *q*-points in  $\Delta S(q, t)$  is  $n_q$ , then  $\Delta S(q, t)$  at a single time point (*t*) can be considered as a



FIG. 1. (a) Concept of the PEPC (projection to extract the perpendicular component) method. TRXL data obtained from an experiment  $[\Delta S(q, t]_{exp}, middle]$  are generally composed of multiple components. As a simple example, here, we assumed that  $\Delta S(q, t]_{exp}$  comprises a combination of the solvent-derived  $(\partial S/\partial T)_{\rho}$  term (left, red) and the species-associated difference scattering curve (SADS, left, blue) arising from a single transient solute species. Each component contributes to  $\Delta S(q, t]_{exp}$  with their own kinetics, making the analysis of  $\Delta S(q, t]_{exp}$  complicated. The PEPC method facilitates the kinetic analysis of TRXL data by extracting the solute kinetics from  $\Delta S(q, t]_{exp}$ . This is achieved by removing the kinetic contribution of  $(\partial S/\partial T)_{\rho}$  using prior knowledge on the shape of  $(\partial S/\partial T)_{\rho}$  in *q*-space. (b) The signal obtained from the TRXL experiment and its components are conceptually considered as vectors and expressed by arrows. A signal obtained from the TRXL experiment  $[\Delta S(q, t]_{exp}, magenta]$  is composed of two difference scattering  $(\Delta S/\partial T)_{\rho}$  using prior knowledge term [solute term plus solute–solvent cage term,  $\Delta S(q, t]_{escp}$ , magenta] is composed of two difference scattering of the vector) is well-known, and, therefore, the unit vector in this direction  $[(\partial S/\partial T)_{\rho}|_{(\partial S/\partial T)_{\rho}}|_{(\partial S/\partial T)_{\rho}}|_{(\partial S/\partial T)_{\rho}}|_{(\partial S/\partial T)_{\rho}|_{(\partial S/\partial T)_{\rho}}|_{(\partial S/\partial T)_$ 

vector in a space of  $n_q$  dimensions. Since  $\Delta S(q, t) = \Delta S(q, t)_{sol-rel} + \Delta S(q, t)_{solvent}$  both  $\Delta S(q, t)_{sol-rel}$  and  $\Delta S(q, t)_{solvent}$  can be also considered as vectors in the same space of  $n_q$  dimensions. For the sake of simplicity, we can first consider vectors in a space of two dimensions for convenient visualization (Fig. 1). In addition, let us consider a simplistic case where  $\Delta S(q, t)_{solvent}$  has only the  $(\partial S/\partial T)_{\rho}$  term, instead of both  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_{T}$  terms, and  $\Delta S(q, t)_{sol-rel}$  contains the kinetics of only a single solute species,

$$\Delta S(q, t)_{\text{solvent}} = \Delta T(t) \times (\partial S/\partial T)_{\rho}, \tag{10}$$

$$\Delta S(q, t)_{\text{sol-rel}} = 1/R \times f_1(t) \times \text{SADS}_1(q), \quad (11)$$

where  $\Delta T(t)$  is the solvent temperature change,  $(\partial S/\partial T)_{\rho}$  is the solvent scattering change per unit temperature change,  $f_1(t)$  is the molar fraction of the solute species at time *t*, and SADS<sub>1</sub>(*q*) is the solute-related term per unit concentration of the first solute species. In this simplistic case, the directions of  $\Delta S(q, t)_{sol-rel}$  and  $\Delta S(q, t)_{solvent}$  vectors do not change with time, whereas their magnitudes can change with time, making it much easier to explain the concept of our approach.

Here, a two-dimensional vector  $\Delta S(q, t)_{exp}$  is the vector sum of a two-dimensional vector  $\Delta S(q, t)_{sol-rel}$  and a two-dimensional vector  $\Delta S(q, t)_{solvent}$ , as shown in Fig. 1(b). In fact, determining  $\Delta S(q, t)_{sol-rel}$ and  $\Delta S(q, t)_{solvent}$  vectors is our goal, but the issue is that it is impossible to determine  $\Delta S(q, t)_{sol-rel}$  and  $\Delta S(q, t)_{solvent}$  vectors without prior knowledge. Here, one critical piece of prior knowledge is that we know the direction of the  $\Delta S(q, t)_{solvent}$  vector, which is determined by the shape of the solvent term  $[(\partial S/\partial T)_{\rho}]$ . In other words, the solvent unit vector  $[(\partial S/\partial T)_{\rho}/|(\partial S/\partial T)_{\rho}|]$ , in the direction of the  $\Delta S(q, t)_{solvent}$  vector, is known from  $(\partial S/\partial T)_{\rho}$ . Therefore, we can decompose the original TRXL data  $[\Delta S(q, t)_{exp}]$  into components parallel  $[\Delta S(q, t)_{exp} \parallel solvent]$ and perpendicular  $[\Delta S(q, t)_{exp \perp solvent}]$  to the solvent unit vector using vector operations. For instance, the magnitude of the component  $\Delta S(q,$  $t_{exp \parallel solvent}$  can be obtained by taking the dot product of  $\Delta S(q, t)_{exp}$  and the solvent unit vector. Multiplying this magnitude to the solvent unit vector gives  $\Delta S(q, t)_{exp \parallel solvent}$ , which can be expressed as follows:

$$\Delta S(q, t)_{\exp \parallel \text{ solvent}} = \left( \Delta S(q, t)_{\exp} \cdot (\partial S/\partial T)_{\rho} / |(\partial S/\partial T)_{\rho}| \right) \\ \times (\partial S/\partial T)_{\rho} / |(\partial S/\partial T)_{\rho}| \\ = proj_{(\partial S/\partial T)_{\rho} / |(\partial S/\partial T)_{\rho}|} \Delta S(q, t)_{\exp} \\ = proj_{(\partial S/\partial T)_{\rho}} \Delta S(q, t)_{\exp},$$
(12)

where the symbol  $\cdot$  denotes the dot product, and  $proj_AB$  denotes the projection of a vector B to a vector A. This component parallel to the solvent unit vector  $[\Delta S(q, t)_{exp \parallel solvent}]$ , which is the projection of the original data  $[\Delta S(q, t)_{exp}]$  to the solvent unit vector  $[(\partial S/\partial T)_{\rho'}/((\partial S/\partial T)_{\rho})]$ , is then subtracted from the original data  $[\Delta S(q, t)_{exp}]$  to obtain the component perpendicular to the solvent unit vector  $[\Delta S(q, t)_{exp}]$ . For this reason, we named this method to extract the perpendicular component as the "projection to extract the perpendicular component (PEPC)." The resulting PEPC-treated data,  $\Delta S(q, t)_{PEPC}$ , can be expressed as follows:

$$\Delta S(q, t)_{PEPC} = \Delta S(q, t)_{exp \perp solvent}$$
  
=  $\Delta S(q, t)_{exp} - \Delta S(q, t)_{exp \parallel solvent}$   
=  $\Delta S(q, t)_{exp} - proj_{(\partial S/\partial T)_{e}} \Delta S(q, t)_{exp}.$  (13)

Let us now consider a series of  $\Delta S(q, t)_{exp}$  vectors at multiple time delays. For typical TRXL experimental data, the magnitudes of both  $\Delta S(q, t)_{sol-rel}$  and  $\Delta S(q, t)_{solvent}$  change with time. Figure 1(c) shows that, even for such a case, the solvent kinetics is completely removed in  $\Delta S(q, t)_{PEPC}$  and the solute kinetics, represented by the relative magnitudes of the original  $\Delta S(q, t)_{sol-rel}$  vectors, is retained in the  $\Delta S(q, t)_{PEPC}$  vectors.

So far, for heuristic purposes, the vectors were represented in two-dimensional space, and we considered simplistic cases where only the magnitudes of the  $\Delta S(q, t)_{solvent}$  and  $\Delta S(q, t)_{sol-rel}$  vectors change with time, whereas their directions do not because we intentionally supposed cases where each vector has only a single component. In real  $\Delta S(q, t)_{exp}$  vectors, the  $n_q$  dimensions need to be considered instead of the two dimensions. Moreover, because the  $\Delta S(q, t)_{solvent}$  vectors have the  $(\partial S/\partial \rho)_T$  component as well as the  $(\partial S/\partial T)_{\rho}$  component, both the direction and magnitude of the  $\Delta S(q, t)_{solvent}$  vectors change with time. Furthermore, because the  $\Delta S(q, t)_{solvent}$  vectors can contain contributions from multiple solute species, their directions as well as their magnitudes will change with time. In such a general case,  $\Delta S(q, t)_{exp}$  can be expressed as  $\Delta S(q, t)_{theory}$  given in Eq. (9), as follows:

$$\Delta S(q, t)_{exp} = 1/R \times \sum_{k} (f_{k}(t) \times SADS_{k}(q)) + \Delta T(t) \\ \times (\partial S/\partial T)_{\rho} + \Delta \rho(t) \times (\partial S/\partial \rho)_{T}.$$
(14)

Of course, at this stage,  $\Delta T(t)$ ,  $\Delta \rho(t)$ ,  $f_k(t)$ 's, and  $SADS_k(q)$ 's are not known. Still, conceptually,  $SADS_k(q)$  can be decomposed into the components parallel and perpendicular to  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_T$ , as follows:

$$\begin{aligned} \text{SADS}_{k}(q) &= \text{SADS}_{k,\parallel}(q) + \text{SADS}_{k,\perp}(q) \\ &= d_{k,\text{T}} \times (\partial S/\partial \text{T})_{\rho} + d_{k,\rho} \times (\partial S/\partial \rho)_{\text{T}} + \text{SADS}_{k,\perp}(q). \end{aligned}$$
(15)

Here, SADS<sub>k</sub>, ||(q) represents the projection of SADS<sub>k</sub>(q) onto a plane containing two vectors  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_{T}$ , and the coefficients  $d_{k,T}$  and  $d_{k,\rho}$  correspond to the contributions of  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_{T}$ , respectively, to SADS<sub>k</sub>, ||(q). SADS<sub>k</sub>,  $\perp(q)$  denotes the component of SADS<sub>k</sub>(q) perpendicular to both  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_{T}$ . A detailed description of the procedure and the mathematical background for decomposing SADS<sub>k</sub>(q) into SADS<sub>k</sub>, ||(q) and SADS<sub>k</sub>,  $\perp(q)$  as well as the process for determining the coefficients  $d_{k,T}$  and  $d_{k,\rho}$  are provided in section "Mathematical background for the decomposition of SADS<sub>k</sub> into SADS<sub>k</sub>, || and SADS<sub>k</sub>, $\perp$ " of the supplementary material. Substituting Eq. (15) into Eq. (14) gives the following equation:

$$\Delta S(q, t)_{\exp} = 1/R \times \sum_{k} (f_{k}(t) \times SADS_{k,\perp}(q)) + \left(\Delta T(t) + 1/R \times \sum_{k} (f_{k}(t) \times d_{k,T})\right) \times (\partial S/\partial T)_{\rho} + \left(\Delta \rho(t) + 1/R \times \sum_{k} (f_{k}(t) \times d_{k,\rho})\right) \times (\partial S/\partial \rho)_{T}.$$
(16)

Via PEPC, the vector component parallel to  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_{T}$ , which is  $(\Delta T(t) + 1/R \times \sum_{k} (f_{k}(t) \times d_{k,T})) \times (\partial S/\partial T)_{\rho} + (\Delta \rho(t) + 1/R \times \sum_{k} (f_{k}(t) \times d_{k,\rho})) \times (\partial S/\partial \rho)_{T}$ , is subtracted from the 28 June 2023 23:58:25

 $\Delta S(q, t)_{exp}$  vector, yielding the vector component perpendicular to  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_{T}$ , as follows:

$$\Delta S(q, t)_{\text{PEPC}} = 1/R \times \sum_{k} (f_{k}(t) \times \text{SADS}_{k, \perp}(q))$$
  
=  $1/R \times \sum_{k} (f_{k}(t) \times (\text{SADS}_{k}(q) - d_{k, T})$   
 $\times (\partial S/\partial T)_{\rho} - d_{k, \rho} \times (\partial S/\partial \rho)_{T}).$  (17)

In  $\Delta S(q, t)_{PEPC}$ , the amounts of the subtracted solvent term are  $\Delta T(t) + 1/R \times \sum_k (f_k(t) \times d_{k,T})$  and  $\Delta \rho(t) + 1/R \times \sum_k (f_k(t) \times d_{k,\rho})$  for  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_T$ , respectively. These amounts deviate from the correct ones,  $\Delta T(t)$  and  $\Delta \rho(t)$ , by  $1/R \times \sum_k (f_k(t) \times d_{k,\Gamma})$  and  $1/R \times \sum_k (f_k(t) \times d_{k,\rho})$ , respectively. In other words, the solvent term is subtracted more than necessary, making the shapes of  $\Delta S(q, t)_{PEPC}$  in the *q*-space altered from the true vector,  $1/R \times \sum_k (f_k(t) \times SADS_k(q))$ . Nevertheless, as can be seen in Eq. (17), the terms contributing to the solvent kinetics, such as  $\Delta T(t)$  and  $\Delta \rho(t)$ , are completely removed, and only the terms contributing to the solute kinetics,  $f_k(t)$ , are left in  $\Delta S(q, t)_{PEPC}$ . Therefore, a kinetic analysis such as the one applied to the TRXL data for HbI K30D<sup>48</sup> can be used to extract kinetics from  $\Delta S(q, t)_{PEPC}$  without any prior knowledge of the solute species and their molecular structures.

### E. Extracting the solute kinetics from $\Delta S(q, t)_{PEPC}$

Once the  $\Delta S(q, t)_{PEPC}$  data are obtained via PEPC, the subsequent kinetic analysis is straightforward. Here, we provide a brief description. The first step is to use SVD on  $\Delta S(q, t)_{PEPC}$  to decompose  $\Delta S(q, t)_{PEPC}$  into the time-independent components in the q-space (left singular vectors, LSVs) and their temporal profiles in the time axis (right singular vectors, RSVs). If the experimental data have  $n_q q$ points at each time delay and  $n_t$  time delays, the SVD of these data gives a total of  $n_t$  LSVs and a total of  $n_t$  RSVs. However, in general, most of these LSVs and RSVs express only the noise of the experimental data, and the number of components expressing a meaningful shape on the experimental data is limited to only a few. For this reason, we first use the low-rank approximation to determine the number of meaningful components and discard the remaining noise from the experimental data prior to the kinetic analysis. The relative contributions of the decomposed vectors are represented by their singular values. Based on these singular values and the autocorrelation values, the number of significant components is determined.

The significant RSVs that survive the low-rank approximation are then fitted simultaneously with a sum of exponential functions sharing common time constants. The number of kinetic components required to satisfactorily fit RSVs and the number of significant singular vectors from the low-rank approximation play important roles as constraints in the subsequent kinetic analysis to determine the kinetic frameworks. A good example can be found in recent work on the K30D mutant of HbI.

### F. Application of PEPC to real data

To demonstrate the effectiveness of the PEPC method, we applied the technique to actual experimental data obtained from the photoreaction of the gold trimer complex (GTC),  $[Au(CN)_2^{-1}]_3$  in water.<sup>30</sup> The results of GFA of the data have been reported.<sup>30</sup>

The photoreaction pathway revealed from the GFA, including the molecular structure of the reactants and intermediates, as well as the time constants of the reaction pathways, is shown in Fig. 2(a). We employed the PEPC technique to process the original  $\Delta S(q, t)_{exp}$ ,



FIG. 2. (a) Photoreaction mechanism of  $[Au(CN)_2^-]_3$  (gold trimer complex, GTC) in water, adapted from previous work.<sup>30</sup> The molecular structures of the reactants and intermediates involved in the photoreaction are shown, along with the time constants and their errors. (b)  $\Delta S(q, t)_{exp}$  from a TRXL experiment on  $[Au(CN)_2^{-1}]_3$  in water,<sup>3</sup> and  $\Delta S(q, t)_{PEPC}$  obtained by applying the PEPC method to  $\Delta S(q, t)_{exp}$ . The kinetic contribution of  $(\partial S/\partial T)_{\alpha}$  of water was eliminated using PEPC. (c) The three right singular vectors (RSVs) obtained from the SVD analysis of  $\Delta S(q, t)_{PEPC}$ . The dots represent the RSVs from experimental data. The three RSVs were globally fitted using a sum of three exponential functions convoluted with a Gaussian instrumental response function (IRF), yielding time constants of 1.7 ps, 1.0 ns, and 114 ns. The lines represent the theoretical fitting curves. (d) The three SADSs obtained from the kinetic analysis of  $\Delta S(q, d)$ t) PEPC (SADS PEPC, black solid lines) are compared with those from the global fitting analysis (GFA) of  $\Delta S(q, t)_{exp}$  (SADS<sub>real</sub>, red solid lines). (e) Three corrected SADSs obtained by correcting the amount of  $(\partial S/\partial T)_{\rho}$  excessively removed in SADS<sub>PEPC</sub>'s (SADS<sub>PEPC, corr</sub> black dash lines) are compared with those from the GFA of  $\Delta S(q, t)_{exp}$  (SADS<sub>real</sub>, red solid lines). Each SADS<sub>PEPC, corr</sub> is obtained by optimizing the weight,  $\alpha$ , of  $(\partial S/\partial T)_{\rho}$  for each SADS<sub>PEPC</sub> so that to minimize the discrepancy between the sum of SADS<sub>PEPC</sub> and  $\alpha \times (\partial S / \partial T)_{\rho}$  and its corresponding SADS<sub>real</sub>. (f) Time-dependent changes of solvent temperature obtained by fitting  $\Delta S(q, t)_{exp}$  for each *t* with the sum of the contributions of three SADS<sub>PEPC, corr</sub> and  $\Delta T(t) \times (\partial S/\partial T)_{\rho}$ , where  $\Delta T(t)$  is the fitting parameter for each t.

which led to the generation of PEPC-treated  $\Delta S(q, t)_{PEPC}$  data as depicted in Fig. 2(b). For PEPC, the signal component which is parallel to  $(\partial S/\partial T)_{\rho}$  of water is removed from  $\Delta S(q, t)_{exp}$ . Normally, the solvent term consists of two signal components  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_{T}$ , and, thus, the signal component which is parallel to either of the two signal components should be removed in the PEPC method. Nevertheless, here, only one of the two components  $(\partial S/\partial \rho)_T$  is considered and removed in the PEPC method because  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial T)_{\rho}$  $\partial \rho$ )<sub>T</sub> for water have similar shapes.<sup>30</sup> The shape of  $(\partial S/\partial T)_{\rho}$  is obtained from a separate experiment using a dye solution (FeCl<sub>3</sub> dissolved in water). It should be noted that, in Fig. 2 and in all subsequent figures, the difference scattering curves,  $\Delta S(q, t)$ , are displayed in the form of  $q\Delta S(q, t)$ , i.e.,  $\Delta S(q, t)$  multiplied by the magnitude of the momentum-transfer vector, q, to better visualize the small signals at large q values. SVD analysis was performed to investigate the soluteonly kinetics in  $\Delta S(q, t)_{PEPC}$ . Examination of the singular values, autocorrelation values, LSVs, and RSVs indicates that the first three components make a substantial contribution to  $\Delta S(q, t)_{PEPC}$  (Fig. S1). The three major RSVs obtained from the SVD analysis of  $\Delta S(q, t)_{PEPC}$  are shown in Fig. 2(c). A satisfactory fit of the RSVs of  $\Delta S(q, t)_{PEPC}$  gives three time constants of  $1.7 \pm 0.1$  ps,  $1.0 \pm 0.1$  ns, and  $114 \pm 3$  ns. The time constants are in excellent agreement with those identified from the previous study using GFA ( $1.6 \pm 0.1$  ps,  $3.0 \pm 0.5$  ns, and  $100 \pm 20 \text{ ns}$ )<sup>30</sup> except for some discrepancy in the second time constant. The three components and three time constants immediately suggest a kinetic model with three species in sequential transitions. Hence, we used this kinetic model and applied kinetics-constrained analysis (KCA) to obtain the SADSs of the three species. We note that when KCA is applied to extract SADSs, a specific kinetic model is applied, and SADSs compatible with the applied kinetic model are extracted. The extracted SADSs are not equal to LSVs. Instead, each SADS must be a linear combination of LSVs. The resulting SADSs  $(SADS_{PEPC})$  are shown in Fig. 2(d) (black solid lines). When compared to the SADSs for the three intermediates identified from the previous GFA [SADS<sub>real</sub>, Fig. 2(d), red solid lines],<sup>30</sup> at first glance, the shape of each SADS<sub>PEPC</sub> is noticeably different from its corresponding SADS<sub>real</sub>. Nevertheless, the determined kinetic framework and the time constants are consistent with the reported ones obtained from the GFA. The excellent agreement shows that the solute kinetics can be easily obtained from  $\Delta S(q, t)_{PEPC}$ . One might notice that the contour plot of the PEPC-treated data in Fig. 2(b) [and Figs. 3(a) and 4(c) to be presented later] exhibits some horizontal stripes. We note that these stripes are not generated by the PEPC procedure but rather are present in the original data,  $\Delta S(q, t)_{exp}$ , except for the case of Fig. 3(a). The stripes are less visible in  $\Delta S(q, t)_{exp}$  due to its higher overall signal amplitude compared to that of  $\Delta S(q, t)_{PEPC}$ . In the case of Fig. 3(a), the horizontal stripes were significantly enhanced with the PEPC procedure. However, the increase in the amplitude of the artifact features is not inherent to the PEPC method itself but is a result of the poor signal-to-noise ratio of the signal components used in PEPC, particularly  $\Delta S(q, t = 100 \text{ ps})_{exp}$ . A more in-depth discussion of this issue is presented in the "Horizontal stripes in the contour plot of the PEPCtreated data" section of the supplementary material.

# G. Structural analysis using the SADSs from $\Delta S(q, t)_{PEPC}$

When the kinetic model is determined in this way, SADSs corresponding to each species suitable for the model can be obtained.



**FIG. 3.** (a)  $\Delta S(q, t)_{exp}$  from a TRXL experiment on  $[Au(CN)_2^{--}]_3$  in water,<sup>30</sup> and  $\Delta S(q, t)_{PEPC}$  obtained by applying the PEPC method to  $\Delta S(q, t)_{exp}$ . The contour plot shown for  $\Delta S(q, t)_{exp}$  is the same as that shown in Fig. 2(b). Here, the kinetic contributions of two components,  $(\partial S/\partial T)_{\rho}$  of water and  $\Delta S(q, t = 100 \text{ ps})_{exp}$ , were eliminated using PEPC, leaving only the signal component perpendicular to the two components. The PEPC method is applied to two subsets of  $\Delta S(q, t)_{exp}$ , each corresponding to  $t < 100 \text{ ps} [\Delta S(q, t)_{XFEL}]$  and  $t \ge 100 \text{ ps} [\Delta S(q, t)_{synch}]$ , yielding  $\Delta S(q, t)_{XFEL}$ , PEPC and  $\Delta S(q, t)_{synch}$ , PEPC, respectively. (b) and (c) The first RSVs obtained from the SVD analysis of (b)  $\Delta S(q, t)_{XFEL}$ , PEPC and (c)  $\Delta S(q, t)_{synch}$ , PEPC. For (b), the first RSV was fitted using an exponential decay function with zero amplitude at t < 0, convoluted with a Gaussian IRF, yielding a time constant for the decay of 1.54 ps. For (c), the first RSV was fitted by the sum of an exponential rise function and an exponential decay function with time constants of 1.14 and 81 ns, respectively.

By applying structural analysis to these SADSs, information on the structure of each species can be obtained. It should be noted that the PEPC process preserves the solute kinetics but alters the shapes of SADSs in *q*-space, as mentioned in Sec. II D. Therefore, the alteration of SADSs needs to be taken into account in the structural analysis based on the SADSs obtained from  $\Delta S(q, t)_{PEPC}$ , denoted as  $SADS_k(q)_{PEPC}$ . The alteration of the shapes of  $SADS_k(q)_{PEPC}$  is caused because the solvent term is excessively removed in the PEPC process, as given in Eq. (17). Here,  $SADS_k(q)_{PEPC}$  and  $SADS_k(q)_{real}$  correspond to  $SADS_{k, \perp}(q)$  and  $SADS_k(q)$  in Eq. (17), respectively.

This relation is expressed in the following equation:

$$\begin{aligned} \operatorname{SADS}_{k}(q)_{\operatorname{PEPC}} &= \operatorname{SADS}_{k,\perp}(q) \\ &= \operatorname{SADS}_{k}(q) - d_{k,\mathrm{T}} \times (\partial \mathrm{S}/\partial \mathrm{T})_{\rho} - d_{k,\rho} \times (\partial \mathrm{S}/\partial \rho)_{\mathrm{T}} \\ &= \operatorname{SADS}_{k}(q)_{\operatorname{real}} - d_{k,\mathrm{T}} \times (\partial \mathrm{S}/\partial \mathrm{T})_{\rho} - d_{k,\rho} \times (\partial \mathrm{S}/\partial \rho)_{\mathrm{T}}. \end{aligned}$$

$$(18)$$

This equation shows that in SADS<sub>k</sub>(q)<sub>PEPC</sub>, the  $(\partial S/\partial T)_{\rho}$  term is further removed by  $d_{k, T} \times (\partial S/\partial T)_{\rho}$ , and the  $(\partial S/\partial \rho)_{T}$  term is further removed by  $d_{k, \rho} \times (\partial S/\partial \rho)_{T}$ . Considering this, it is necessary to conduct structural analysis using the amount of solvent term excessively removed in the PEPC process as a fitting parameter. The corrected SADSs can be expressed as follows:

$$SADS_k(q)_{corr} = SADS_k(q)_{PEPC} + g_k \times (\partial S/\partial T)_{\rho} + h_k \times (\partial S/\partial \rho)_T.$$
(19)



FIG. 4. (a) Reaction pathways of CHI<sub>3</sub> photolysis in methanol and cyclohexane, adapted from the literature.9 (b) Potential signal components for CHI<sub>3</sub> photolysis in cyclohexane, predicted based on the null hypothesis that the reaction pathway for cyclohexane is the same as that assigned for methanol. (c)  $\Delta S(q, t)_{exp}$  from a TRXL experiment on CHI<sub>3</sub> in cyclohexane,<sup>20</sup> and  $\Delta S(q, t)_{PEPC}$  obtained by applying the PEPC method to  $\Delta S(q, t)_{exp}$ . The contributions of the four signal components shown in (b) are removed from  $\Delta S(q, t)_{exp}$  via PEPC. The remaining signal in  $\Delta S(q, t)_{PEPC}$  indicates that the null hypothesis is not true, and the reaction pathways are different for cyclohexane and methanol. (d) The first RSV obtained from the SVD analysis of  $\Delta S(q, t)_{PEPC}$  (dots). The kinetics of the first RSV agrees well with the time-dependent concentration profile of iso-CHI2-I retrieved using GFA (solid line), indicating that the remaining signal in  $\Delta S(q, t)_{PEPC}$  is due to the contribution of the reaction pathway yielding iso-CHI2-I. (e) The first LSV obtained from the SVD analysis of  $\Delta S(q, t)_{PEPC}$  (black solid line). The LSV is compared with the SADS corresponding to the formation of iso-CHI2-I [SADS(iso-CHI2-I), red solid line]. Although the first LSV did not correspond well with SADS(iso-CHI2-I), the LSV<sub>corr</sub> (black dashed line), which was adjusted for the contributions of the signal components excessively removed during the PEPC procedure, showed excellent agreement with SADS(iso-CHI2-I). The LSVcorr is described as the sum of five terms, LSV,  $\alpha \times (\partial S / \partial T)_{\rho}$ ,  $\beta \times (\partial S / \partial \rho)_{T}$ ,  $\gamma \times SADS(CHI_{3} \rightarrow CHI_{2} + I)$ , and  $\delta \times \text{SADS}(\text{CHI}_3 \rightarrow \text{CHI}_2 + \frac{1}{2} \text{ I}_2)$  where  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  are the fitting parameters. (f) and (g) The time traces of  $\Delta T(f)$  and  $\Delta \rho(g)$ , retrieved from the correct kinetic framework that includes the contribution of iso-CHI2-I (dots). Once the kinetics of the reaction is determined using PEPC and the solute-related term for each relevant chemical species is determined through a subsequent structural analysis, the reconstruction of the hydrodynamic response of the solvent is straightforward.  $\Delta T$ and  $\Delta \rho$  for each time delay can be retrieved as described in the text. The resulting  $\Delta T$  and  $\Delta \rho$  match those determined using GFA (solid lines).

Here,  $g_k$  and  $h_k$  are the coefficients for  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_T$ , respectively, and ideally should be the same as  $d_{k, T}$  and  $d_{k, \rho}$ , respectively. Then, in the fitting process, the  $\chi^2$  value representing the discrepancy between SADS<sub>k</sub>(q)<sub>corr</sub> and SADS<sub>k</sub>(q) is minimized using  $g_k$  and  $h_k$  as well as the structural parameters of intermediates and reactants which are used to calculate  $S_k(q)_{solute}$ ,  $S_0(q)_{solute}$ ,  $S_k(q)_{cage}$ , and  $S_0(q)_{cage}$  in Eq. (7), as fitting parameters. By correcting the contribution of the solvent term that is excessively removed in the PEPC process, specifically  $g_k \times (\partial S/\partial T)_{\rho}$  and  $h_k \times (\partial S/\partial \rho)_T$ , the shape of the difference scattering curve for each species obtained from the GFA technique<sup>30</sup> can be reproduced [Fig. 2(f)]. Here, the contribution of  $h_k$  $\times (\partial S/\partial \rho)_{T}$  is ignored for this specific example with water as a solvent. It is noteworthy that, despite adjusting  $g_k$  and  $h_k$  to minimize the discrepancy between  $SADS_k(q)_{corr}$  and  $SADS_k(q)$ , the satisfactory fitting of  $SADS_k(q)_{corr}$  to  $SADS_k(q)$  is only achieved for the correct structure of the species, thus enabling the identification of the species and optimization of the detailed molecular structure. For instance, Fig. S5 highlights that  $SADS_1(q)_{corr}$  is only consistent with SADS calculated for the structure corresponding to the T<sub>1</sub>' state and exhibits a noticeable deviation from those calculated for the structures corresponding to the T1 state or the tetramer (Fig. S5, left panels). Similarly,  $SADS_2(q)_{corr}$  and  $SADS_3(q)_{corr}$  match only to SADS calculated for the structures corresponding to the T1 state and the tetramer, respectively (Fig. S5, middle and right panels). These findings unequivocally demonstrate that  $SADS_k(q)_{corr}$  accurately matches SADS calculated only for the correct molecular structure of the species. In other words, despite the altered shape of  $SADS_k(q)_{PEPC}$  in *q*-space, it remains feasible to extract the relevant structural information corresponding to each species by analyzing  $SADS_k(q)_{PEPC}$ .

With this, both the kinetic analysis and the structural analysis for the solute term are completed. Here, we note that consideration of  $g_k$ and  $h_k$  in the structural fitting using PEPC-treated data does not increase the complexity of the structural fitting. In other words,  $g_k$  and  $h_k$  are not independent fitting parameters in the structural fitting but are functions expressed by independent fitting parameters. Accordingly, the structural analysis of a SADS<sub>k</sub>(q)<sub>PEPC</sub> has the same level of complexity (the number of independent fitting parameters) as the structural analysis of a correct SADS, a SADS<sub>k</sub>(q), that is not altered by the PEPC process. The method for determining the coefficients  $g_k$  and  $h_k$  is described in section "Determination of coefficients  $g_k$  and  $h_k$  during structural analysis" of the supplementary material.

The last remaining step is to reconstruct the information about the hydrodynamic response of the solvent (solvent kinetics). The analysis of solvent kinetics is straightforward once the solute-related term,  $\Delta S(q, t)_{sol-rel}$ , is determined from the kinetic analysis and structural analysis. At this stage,  $f_k(t)$  for the solute-related term in Eq. (5) is obtained through kinetic analysis. The subsequent structural analysis refines the molecular structures and determines  $SADS_k(q)$ . The solvent-only term,  $\Delta S(q, t)_{solvent} = \Delta T(t) \times (\partial S/\partial T)_{\rho} + \Delta \rho(t) \times (\partial S/\partial T)_{\rho}$  $\partial \rho$ <sub>T</sub>, can be determined by subtracting  $\Delta S(q, t)_{sol-reb}$  i.e., 1/  $R \times \Sigma_k[f_k(t) \times \text{SADS}_k(q)]$ , from  $\Delta S(q, t)_{exp}$ , as shown in Eq. (14).  $\Delta T(t)$  and  $\Delta \rho(t)$  can be retrieved from the  $\Delta S(q, t)_{solvent}$  by simply performing NOD of  $\Delta S(q, t)_{solvent}$  as a linear combination of two terms  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_{T}$ . In the case of the  $[Au(CN)_{2}]_{3}$  in water shown earlier, the profile of  $\Delta T(t)$  reconstructed through the corresponding process is shown in Fig. 2(f). Note that the two terms ( $\partial S$ /  $\partial T$ )<sub> $\rho$ </sub> and  $(\partial S/\partial \rho)_T$  for the solvent response have similar shapes in *q*space, especially in the case of water, making it difficult to distinguish the contributions of the two terms in  $\Delta S(q, t)_{solvent}$ . Accordingly, here, the hydrodynamics response is described with only the  $(\partial S/\partial T)_{\rho}$  term, as in Eq. (10), instead of the linear combination of the two terms, and

the resulting  $\Delta T(t)$  is indicated. In general, except for water, the two terms  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_{T}$  for the solvent response have distinct shapes in *q*-space, and, therefore, the contribution of each term in the solvent response can be easily determined. An example of such a general case, CHI<sub>3</sub> in cyclohexane, is given in Figs. 4(f) and 4(g).

# H. PEPC for simultaneous removal of the contributions of multiple signal components

As demonstrated in Sec. II F, PEPC is a powerful tool for eliminating the contribution of a single component, such as  $(\partial S/\partial T)_{\rho}$ . In fact, PEPC can also be used to simultaneously remove multiple signal components, making it a versatile method with applications extending beyond solvent subtraction. For example, PEPC can be employed to eliminate the contributions of known, trivial reaction intermediates, allowing for the isolation of potential contributions from unknown intermediates that may be hidden in experimental signals. Additionally, PEPC can be applied to eliminate the kinetic contributions of an intermediate dominating at a specific time delay, enabling the extraction of contributions from other intermediates in the reaction.

Prior to discussing applications of PEPC, we provide an overview of its theoretical principles. We describe how PEPC can eliminate the contributions of multiple signal components, which we refer to as "trivial components" for brevity, and provide a mathematical representation of the resulting PEPC-treated signal. The vector space spanned by these trivial components is referred to as the "trivial space (TS)." Assuming the presence of *m* trivial components, denoted as  $\{trv_1, trv_2, ..., trv_m\}$ , PEPC eliminates signal components parallel to these *m* trivial components and retains only the component perpendicular to all *m* components. The relation between  $\Delta S(q, t)_{exp}$  and  $\Delta S(q, t)_{PEPC}$  can be expressed as follows:

$$\Delta S(q, t)_{exp} = \sum_{i=1}^{m} w_i \times trv_i + \Delta S(q, t)_{PEPC}.$$
 (20)

Here,  $w_i$  represents the weight of *i*th trivial component,  $trv_i$ , removed via PEPC. As illustrated in Fig. 1(c), the kinetics of a trivial component does not contribute in the direction perpendicular to that component. Therefore, by finding  $\Delta S(q, t)_{PEPC}$  that is perpendicular to all *m* trivial components, the kinetics of the signal components other than the *m* trivial components can be extracted. This can be achieved by satisfying the following equation for all *i*:

$$\Delta S(q, t)_{\text{PEPC}} \cdot trv_i = 0. \tag{21}$$

The main concern is whether a set  $\{w_1, w_2, ..., w_n\}$  for Eq. (20) that satisfies Eq. (21) for all *i* can be readily calculated.

In fact, the set { $w_1$ ,  $w_2$ , ...,  $w_n$ } that satisfies Eq. (21) for all *i* can be readily obtained by calculating the mathematical projection. The projection of a vector *v* onto a subspace U, spanned by a set of vectors { $u_1$ ,  $u_2$ , ...,  $u_n$ }, denoted by  $proj_Uv$ , provides the vector component that lies within the subspace, while the residual of the projection, i.e.,  $v - proj_Uv$ , gives the component perpendicular to all vectors,  $u_1$ ,  $u_2$ , ..., and  $u_n$ .<sup>58</sup> Thus, the first term on the right-hand side of Eq. (20) can be obtained by calculating the projection,  $proj_{TS}\Delta S(q, t)_{exp}$ , and the second term,  $\Delta S(q, t)_{PEPC}$ , that satisfies Eq. (21) can be obtained by calculating the residual of the projection,  $\Delta S(q, t)_{exp} - proj_{TS}\Delta S(q, t)_{exp}$ . Mathematically, this can be expressed as follows:

$$\Delta S(q, t)_{PEPC} = \Delta S(q, t)_{exp} - \sum_{i=1}^{m} w_i \times trv_i$$
$$= \Delta S(q, t)_{exp} - proj_{TS}\Delta S(q, t)_{exp}.$$
(22)

Equation (22) shows how PEPC can eliminate the contributions of multiple trivial components, retaining only the component perpendicular to the trivial components by removing signal components parallel to the trivial components. Mathematically, this PEPC treatment can be expressed using a projection operator. However, the projection formula may not be very intuitive in practice. To address this, we propose a practical and intuitive method based on a mathematical property of the projection operator: the projection,  $proj_Uv$ , is given by the linear combination of  $\{u_1, u_2, ..., u_n\}$  that minimizes the magnitude of the residual,  $v - proj_Uv$ .<sup>58</sup> By utilizing this property, we can obtain the projection,  $proj_{TS}\Delta S(q, t)_{exp}$ , by finding the linear combination of  $\{trv_1, trv_2, ..., trv_m\}$  that minimizes the magnitude of the residual of the projection,  $\Delta S(q, t)_{PEPC}$ . Based on this, the projection,  $proj_{TS}\Delta S(q, t)_{exp}$ , can be obtained by finding the least squares solution of the following system of linear equations:<sup>59–61</sup>

$$\Delta S(q, t)_{\exp} = \sum_{i=1}^{m} x_i \times tr v_i.$$
(23)

Here,  $x_i$  represents the weight of *i*th trivial component,  $trv_i$ . Note that the least squares solution, which we denote as  $\{w_{LS, 1}, w_{LS, 2}, ..., w_{LS, m}\}$ , can be obtained unambiguously based on well-established numerical algorithms.<sup>62</sup> Once the least squares solution is determined,  $\Delta S(q, t)_{PEPC}$  can be derived as follows:

$$proj_{\rm TS}\Delta S(q, t)_{\rm exp} = \sum_{i=1}^{m} w_{LS,i} \times trv_i, \qquad (24)$$

$$\Delta S(q, t)_{PEPC} = \Delta S(q, t)_{exp} - proj_{TS}\Delta S(q, t)_{exp}$$
$$= \Delta S(q, t)_{exp} - \sum_{i=1}^{m} w_{LS,i} \times trv_{i}.$$
 (25)

This approach based on least squares fitting provides a straightforward and practical method for numerically calculating the PEPC-treated signal,  $\Delta S(q, t)_{PEPC}$ , for multiple trivial components, { $trv_1, trv_2, ..., trv_m$ }.

The kinetic analysis of  $\Delta S(q, t)_{PEPC}$  resulting from the PEPC treatment with multiple trivial components can be performed in the same way as for  $\Delta S(q, t)_{PEPC}$  treated with a single trivial component, namely,  $(\partial S/\partial T)_{\rho}$ . Once the kinetic framework is established from the kinetic analysis, the SADS corresponding to the kinetic framework can be obtained using KCA. To extract molecular structural information, the resulting SADS<sub>k</sub>(q)<sub>PEPC</sub> can be analyzed similarly, with corrections similar to those in Eq. (19) but with the equation modified as follows:

$$SADS_k(q)_{corr} = SADS_k(q)_{PEPC} + \sum_{i=1}^m \alpha_{i,k} \times trv_i.$$
(26)

Here,  $\alpha_{i_{b}k}$  is the weight of the correction for the *i*th trivial component, *trv<sub>i</sub>*. By employing Eq. (26), the molecular structure of each species can be optimized by minimizing the deviation between the experimentally obtained SADS<sub>k</sub>(*q*)<sub>corr</sub>, which has been corrected for the distortion from PEPC treatment, and the calculated SADS<sub>k</sub>(*q*)<sub>real</sub> derived from the molecular structure.

28 June 2023 23:58:25

# I. Application of PEPC to determine the kinetic framework

To showcase the versatility of PEPC, particularly in handling multiple trivial components, we utilized it in analyzing two distinct datasets: one for the photoreaction of [Au(CN)2-]3 in water30 and the other for the photoreaction of CHI3 in cyclohexane.<sup>20</sup> For [Au(CN)<sub>2</sub><sup>-</sup>]<sub>3</sub> in water, the TRXL data consist of two sub-datasets measured at two different facilities: a synchrotron and an XFEL.<sup>30</sup> The data corresponding to  $t \ge 100$  ps [ $\Delta S(q, t)_{synch}$ ] and t < 100 ps [ $\Delta S(q, t)_{synch}$ ] t)<sub>XFEL</sub>] were obtained from the experiments at the synchrotron and XFEL, respectively. Typically, a synchrotron experiment using the same sample is performed before the XFEL experiment. In this case, the preliminary information,  $\Delta S(q, t)_{synch}$ , can be used when analyzing  $\Delta S(q, t)_{XFEL}$  to identify any contributions from short-lived intermediates that were not identified in  $\Delta S(q, t)_{synch}$ . Moreover, when analyzing  $\Delta S(q, t)_{synch}$ , PEPC can be used to identify any contributions from long-lived intermediates other than those present at the earliest time delay, t = 100 ps, by removing the contribution of the intermediate present at t = 100 ps from the entire  $\Delta S(q, t)_{synch}$ . Here, in contrast to the PEPC process illustrated in Fig. 2(b), the contribution of  $\Delta S(q,$  $t = 100 \text{ ps})_{exp}$  is removed, in addition to the contribution of solvent heating signal  $(\partial S/\partial T)_{\rho}$ . As a result, only the signal component perpendicular to these two components remains after removing the signal components parallel to  $(\partial S/\partial T)_{\rho}$  and  $\Delta S(q, t = 100 \text{ ps})_{\text{exp}}$ . The resulting  $\Delta S(q, t)_{PEPC}$  from  $\Delta S(q, t)_{XFEL}$  and  $\Delta S(q, t)_{synch}$  is shown in Fig. 3(a) as  $\Delta S(q, t)_{XFEL}$ , PEPC and  $\Delta S(q, t)_{synch}$ , PEPC, respectively. The observation of a remaining signal in  $\Delta S(q, t)_{XFEL}$ , PEPC and  $\Delta S(q, t)_{XFEL}$ t)<sub>synch, PEPC</sub> suggests the presence of reaction intermediates other than those exists at t = 100 ps, which was previously assigned as an intermediate in the T<sub>1</sub> state based on GFA.<sup>30,34</sup> SVD analysis revealed that each of  $\Delta S(q, t)_{XFEL, PEPC}$  and  $\Delta S(q, t)_{synch, PEPC}$  exhibits only one significant LSV (Figs. S6 and S7), indicating the presence of an additional early intermediate before the intermediate in the T<sub>1</sub> state within the t < 100 ps range and the other late intermediate in the t > 100 ps range. The significant RSVs are shown in Figs. 3(b) and 3(c). Kinetic analysis of the RSV of  $\Delta S(q, t)_{XFEL}$ , PEPC reveals that the RSV rises instantaneously within the instrument response function (IRF, 480 fs) of the experiment and decays with a single time constant of  $1.54 \pm 0.07$  ps. Obviously, this observation indicates that the early intermediate (an intermediate in the T<sub>1</sub>' state)<sup>34</sup> forms within the IRF and transforms to the intermediate in the T1 state with a 1.54 ps time constant. For  $\Delta S(q, t)_{synch, PEPC}$ , the RSV rises with a  $1.14 \pm 0.07$  ns time constant and decays with an  $81 \pm 4$  ns time constant, indicating that the T<sub>1</sub> intermediate transforms to the other late intermediate (a tetramer)<sup>30</sup> with 1.14 ns, and the late intermediate decays to ground state with 81 ns.

From the description of the kinetic analysis earlier, one would notice that the kinetic analysis of two sub-datasets,  $\Delta S(q, t)_{XFEL, PEPC}$ and  $\Delta S(q, t)_{synch, PEPC}$ , is even more straightforward than that of  $\Delta S(q, t)_{PEPC}$  demonstrated in Sec. II F. This is due to two reasons. First, the data are divided into two subsets and analyzed separately, thereby reducing the numbers of signal components and time constants in each sub-dataset. Specifically,  $\Delta S(q, t)_{XFEL, PEPC}$  has one signal component and one time constant, and  $\Delta S(q, t)_{synch, PEPC}$  has one signal component and two time constants, compared to the three signal components and three time constants in the entire  $\Delta S(q, t)_{PEPC}$  shown in Fig. 2(b). However, the advantage of dividing the data into subsets for kinetic analysis will not be further discussed in this study, as it is not relevant to the benefit of using PEPC. Second, and more importantly, PEPC eliminates the contribution of an additional trivial component, i.e.,  $\Delta S(q, t = 100 \text{ ps})_{exp}$ , reducing the number of signal components, i.e., the number of relevant chemical species contributing to the signal. Thus, applying PEPC reduces the number of solute signal components in  $\Delta S(q, t)_{XFEL}$  and  $\Delta S(q, t)_{synch}$  from two  $[T_1' \text{ and } T_1 \text{ in}$  $\Delta S(q, t)_{XFEL}$  and  $T_1$  and tetramer in  $\Delta S(q, t)_{synch}]$  to one  $[T_1' \text{ only in}$  $\Delta S(q, t)_{XFEL}$  and tetramer only in  $\Delta S(q, t)_{synch}]$ . Consequently, the kinetics of the remaining solute species,  $T_1'$  and tetramer, is more clearly exposed in the resulting significant RSVs of  $\Delta S(q, t)_{XFEL}$ , PEPC and  $\Delta S(q, t)_{synch}$ , PEPC, respectively, facilitating the kinetic analysis.

The photoreaction of CHI<sub>3</sub> presents a fundamental question concerning whether the reaction pathway depends on the solvent, and if so, how. Therefore, it is crucial to determine whether the reaction pathways differ in different solvents. Previous studies examined the reaction pathway in methanol using TRXL [Fig. 4(a), pathways wrapped in a blue dashed line], identifying that only the dissociation channel yielding the CHI<sub>2</sub> radical and I radical or the iodine molecule is active.<sup>9</sup> Afterward, the same reaction was examined in a different solvent, cyclohexane, using TRXL.<sup>20</sup> In the study, the reaction pathway in cyclohexane was identified by using GFA, confirming that the pathway is different in cyclohexane [Fig. 4(a), pathways wrapped in a red dashed line].

If one were to analyze the TRXL data for cyclohexane, again, an essential question to address is whether the reaction pathway differs from that observed in methanol. To address this question using GFA, one would need to verify whether the fitting using the kinetic framework determined for methanol works well with the TRXL data for cyclohexane, which requires optimizing several parameters such as rate constants and excitation ratio. However, such a verification task can be tedious and time-consuming. In this study, we demonstrate that PEPC offers a more straightforward approach for confirming any differences in reaction pathways between solvents.

The key is that, by applying PEPC, the contributions of the signal components corresponding to the reaction pathways identified for methanol can be removed from the TRXL data for cyclohexane, allowing for an assessment of whether the reaction pathway differs in the two solvents. Under the null hypothesis that the reaction pathway for cyclohexane is the same as that assigned for methanol, it can be predicted that the two solute-related terms corresponding to the formation of  $CHI_2$  and the iodine radical [Fig. 4(b), blue solid line] and the subsequent transformation of the iodine radical to the iodine molecule [Fig. 4(b), green solid line] as well as the two solvent terms of cyclohexane [Fig. 4(b), black and red solid lines for  $(\partial S/\partial T)_{\rho}$  and  $(\partial S/\partial \rho)_{T}$ of cyclohexane, respectively] would contribute to the TRXL signal for the photolysis of CHI3 in cyclohexane. PEPC enables removal of the contributions of these four signal components from  $\Delta S(q, t)_{exp}$ , and the resulting  $\Delta S(q, t)_{PEPC}$  can be used to test the null hypothesis. If the null hypothesis is true, then there would be no meaningful signal in the  $\Delta S(q, t)_{PEPC}$ . However, if the reaction pathway is different and a signal component not identified for methanol exists, then a significant amplitude of the signal would remain in the  $\Delta S(q, t)_{PEPC}$ , indicating the presence of a distinct signal component that was not be completely subtracted during the PEPC process.

The contour plots of the measured  $\Delta S(q, t)_{exp}$  for cyclohexane and the  $\Delta S(q, t)_{PEPC}$  are depicted in Fig. 4(c). The presence of a

remaining signal in  $\Delta S(q, t)_{PEPC}$  suggests that the null hypothesis is invalid, and the reaction pathways differ for cyclohexane and methanol. Further investigation, involving kinetic and structural analysis of the residual signal in  $\Delta S(q, t)_{PEPC}$ , reveals that the residual signal originates from the pathway that produces an isomer of CHI<sub>3</sub>, iso- $CHI_2-I_2^{20}$  a pathway that is inactive in methanol<sup>9</sup> [Fig. 4(a)]. SVD analysis of  $\Delta S(q, t)_{PEPC}$  results in a single major component (Fig. S8). Regarding kinetics, the major RSV obtained from the SVD analysis [Fig. 4(d), dots] matches with the time-dependent concentration profile of iso-CHI<sub>2</sub>-I, obtained using GFA<sup>20</sup> [Fig. 4(d), solid line], indicating that the residual signal in  $\Delta S(q, t)_{PEPC}$  is due to iso-CHI<sub>2</sub>-I. Furthermore, regarding structural analysis, the LSV<sub>corr</sub> [Fig. 4(e), black dashed line] calculated based on Eq. (26), i.e.,  $LSV_{corr} = LSV + \alpha$  $\times (\partial S/\partial T)_{\rho} + \beta \times (\partial S/\partial \rho)_{T} + \gamma \times SADS(CHI_{3} \rightarrow CHI_{2} + I) + \delta$ × SADS(CHI<sub>3</sub>→CHI<sub>2</sub> + 1/2 I<sub>2</sub>), where  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  are the fitting parameters, agrees excellently with the SADS calculated for iso- $CHI_2-I^{20}$  [SADS(*iso*-CHI<sub>2</sub>-I), Fig. 4(e), red solid line], thus confirming that the residual signal in  $\Delta S(q, t)_{PEPC}$  originates from *iso*-CHI<sub>2</sub>-I.

The demonstration of PEPC to the TRXL data for the photoreaction of CHI<sub>3</sub> shows how PEPC can be utilized to determine the kinetic framework of the reaction and to analyze the structural dynamics of the reaction. The significant residual signal observed in  $\Delta S(q, t)_{PEPC}$ confirms the difference in the reaction pathways for cyclohexane and methanol. Further analysis of  $\Delta S(q, t)_{PEPC}$  reveals (1) a reaction intermediate, *iso*-CHI<sub>2</sub>–I, that is involved in the reaction pathways in cyclohexane, along with (2) the kinetics of *iso*-CHI<sub>2</sub>–I, especially, when and how *iso*-CHI<sub>2</sub>–I forms and decays, and (3) the molecular structure of *iso*-CHI<sub>2</sub>–I. The determined kinetics and molecular structures can be utilized for the analysis of the hydrodynamics of the solvent, cyclohexane, as explained in Sec. II G. The resulting  $\Delta T(t)$  and  $\Delta \rho(t)$  are shown in Figs. 4(f) and 4(g), respectively.

Although we have presented and demonstrated the PEPC method as a powerful tool for analyzing the kinetic behavior of TRXL signals, it is worth noting that a conventional method, referred to as the "nodal-point method (NP method)," can also provide similar benefits. However, as discussed in the "Comparison of the PEPC method and the nodal-point method (NP method)" section of the supplementary material, our comparison shows that the PEPC method outperforms the NP method in several crucial aspects.

# J. Potential of the PEPC method for other types of data

In this study, we demonstrated the ability of the PEPC method to remove the solvent kinetics from the TRXL data by utilizing the prior knowledge of the shape of the solvent term. In principle, the application of PEPC is not limited to TRXL data and can be extended to other types of data where a certain component is known. By using the PEPC method, the kinetics of the known component can be removed regardless of the type of the data. Furthermore, the PEPC method can be applied to datasets other than time-resolved data. In time-resolved data, spectra change with time. For other types of data, the spectra change with experimental variables other than time, such as temperature or pressure. In such cases, if the spectrum of a species is known and its magnitude changes with the experimental variable, the PEPC method can be used to subtract its variable-dependent changes from the original data. For example, for a set of spectra measured as a function of temperature, the temperature-dependent changes in the spectrum of a species can be removed using the PEPC method. Overall, the PEPC method has a great potential as a versatile and powerful tool for the analysis of various types of data, extending beyond the scope of TRXL data analysis.

### **III. CONCLUSION**

In this work, we introduced PEPC, a new analytic method for the analysis of time-resolved data, and demonstrated its application through various examples. By removing the kinetic contributions of known signal components, such as solvent heating or well-known intermediates or products, the PEPC technique greatly facilitates the kinetic analysis of the treated signal. As a representative example, we showed that the PEPC technique can be used to easily extract kinetic information of solute-related terms from TRXL data by removing the kinetic contributions of solvent terms. Although the solvent term being excessively removed in this process alters the shape of the solute-related term in q-space, structural analysis can still be performed on the resulting PEPC-treated data to obtain structural information. As the PEPC process is fully arithmetic, it has a fast computation time, allowing for real-time monitoring of the status of the data being obtained during the experiment. Overall, the PEPC technique has the potential to serve as a valuable tool for both postexperiment data analysis and real-time monitoring of time-resolved data, with the ability to extend beyond the scope of TRXL data analysis. Further research is required to explore and demonstrate the full range of applications of this powerful and versatile method in various fields of science.

#### SUPPLEMENTARY MATERIAL

See the supplementary material for detailed discussions on various topics, including the mathematical background for the decomposition of  $SADS_k$  into  $SADS_k$  and  $SADS_k$ , the appearance of horizontal stripes in the contour plot of the PEPC-treated data, the determination of coefficients  $g_k$  and  $h_k$  during structural analysis, and a comparison of the PEPC method and the nodal-point method.

#### ACKNOWLEDGMENTS

This work was supported by the Institute for Basic Science (IBS-R033).

# AUTHOR DECLARATIONS Conflict of Interest

The authors have no conflicts to disclose.

#### **Author Contributions**

Hosung Ki: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). Jain Gu: Investigation (supporting); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). Yongjun Cha: Investigation (supporting); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). Kyung Won Lee: Investigation (supporting); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). Hyotcherl Ihee: Conceptualization (equal); Funding acquisition (lead); Supervision (lead); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal).

# DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

#### REFERENCES

- <sup>1</sup>H. Ihee, "Visualizing solution-phase reaction dynamics with time-resolved x-ray liquidography," Acc. Chem. Res. **42**, 356–366 (2009).
- <sup>2</sup>H. Ihee, M. Wulff, J. Kim, and S. Adachi, "Ultrafast x-ray scattering: Structural dynamics from diatomic to protein molecules," Int. Rev. Phys. Chem. 29, 453–520 (2010).
- <sup>3</sup>K. H. Kim, J. Kim, J. H. Lee, and H. Ihee, "Topical review: Molecular reaction and solvation visualized by time-resolved x-ray solution scattering: Structure, dynamics, and their solvent dependence," Struct. Dyn. 1, 011301 (2014).
- <sup>4</sup>J. Kim, K. H. Kim, K. Y. Oan, J. H. Lee, K. Hong, H. Cho, N. Huse, R. W. Schoenlein, T. K. Kim, and H. Ihee, "Tracking reaction dynamics in solution by pump-probe x-ray absorption spectroscopy and x-ray liquidography (solution scattering)," Chem. Commun. **52**, 3734–3749 (2016).
- <sup>5</sup>H. Ki, K. Y. Oang, J. Kim, and H. Ihee, "Ultrafast x-ray crystallography and liquidography," Annu. Rev. Phys. Chem. **68**, 473–497 (2017).
- <sup>6</sup>R. Neutze, R. Wouts, S. Techert, J. Davidsson, M. Kocsis, A. Kirrander, F. Schotte, and M. Wulff, "Visualizing photochemical dynamics in solution through picosecond x-ray scattering," Phys. Rev. Lett. 87, 195508 (2001).
- <sup>7</sup>J. Davidsson, J. Poulsen, M. Cammarata, P. Georgiou, R. Wouts, G. Katona, F. Jacobson, A. Plech, M. Wulff, G. Nyman, and R. Neutze, "Structural determination of a transient isomer of CH<sub>2</sub>I<sub>2</sub> by picosecond x-ray diffraction," Phys. Rev. Lett. **94**, 245503 (2005).
- <sup>8</sup>A. Plech, V. Kotaidis, M. Lorenc, and M. Wulff, "Thermal dynamics in laser excited metal nanoparticles," Chem. Phys. Lett. 401, 565–569 (2005).
- <sup>9</sup>J. H. Lee, J. Kim, M. Cammarata, Q. Kong, K. H. Kim, J. Choi, T. K. Kim, M. Wulff, and H. Ihee, "Transient x-ray diffraction reveals global and major reaction pathways for the photolysis of iodoform in solution," Angew. Chem., Int. Ed. 47, 1047–1050 (2008).
- <sup>10</sup>J. H. Lee, T. K. Kim, J. Kim, Q. Kong, M. Cammarata, M. Lorenc, M. Wulff, and H. Ihee, "Capturing transient structures in the elimination reaction of haloalkane in solution by transient x-ray diffraction," J. Am. Chem. Soc. 130, 5834–5835 (2008).
- <sup>11</sup>M. Andersson, E. Malmerberg, S. Westenhoff, G. Katona, M. Cammarata, A. B. Wohri, L. C. Johansson, F. Ewald, M. Eklund, M. Wulff, J. Davidsson, and R. Neutze, "Structural dynamics of light-driven proton pumps," Structure 17, 1265–1275 (2009).
- <sup>12</sup>M. Christensen, K. Haldrup, K. Bechgaard, R. Feidenhans'l, Q. Y. Kong, M. Cammarata, M. Lo Russo, M. Wulff, N. Harrit, and M. M. Nielsen, "Time-resolved x-ray scattering of an electronically excited state in solution. Structure of the <sup>3</sup>A<sub>2u</sub> state of tetrakis-μ-pyrophosphitodiplatinate(II)," J. Am. Chem. Soc. 131, 502–508 (2009).
- <sup>13</sup>Q. Kong, J. H. Lee, K. H. Kim, J. Kim, M. Wulff, H. Ihee, and M. H. J. Koch, "Ultrafast x-ray solution scattering reveals different reaction pathways in the photolysis of triruthenium dodecacarbonyl (Ru<sub>3</sub>(CO)<sub>12</sub>) after ultraviolet and visible excitation," J. Am. Chem. Soc. **132**, 2600–2607 (2010).
- <sup>14</sup>P. L. Ramachandran, J. E. Lovett, P. J. Carl, M. Cammarata, J. H. Lee, Y. O. Jung, H. Ihee, C. R. Timmel, and J. J. van Thor, "The short-lived signaling state of the photoactive yellow protein photoreceptor revealed by combined structural probes," J. Am. Chem. Soc. 133, 9395–9404 (2011).
- <sup>15</sup>K. H. Kim, S. Muniyappan, K. Y. Oang, J. G. Kim, S. Nozawa, T. Sato, S. Y. Koshihara, R. Henning, I. Kosheleva, H. Ki, Y. Kim, T. W. Kim, J. Kim, S. Adachi, and H. Ihee, "Direct observation of cooperative protein structural dynamics of homodimeric hemoglobin from 100 ps to 10 ms with pump-probe x-ray solution scattering," J. Am. Chem. Soc. **134**, 7001–7008 (2012).
- <sup>16</sup>T. W. Kim, J. H. Lee, J. Choi, K. H. Kim, L. J. van Wilderen, L. Guerin, Y. Kim, Y. O. Jung, C. Yang, J. Kim, M. Wulff, J. J. van Thor, and H. Ihee, "Protein structural dynamics of photoactive yellow protein in solution revealed by pump-probe x-ray solution scattering," J. Am. Chem. Soc. **134**, 3145–3153 (2012).
- <sup>17</sup>K. H. Kim, J. H. Lee, J. Kim, S. Nozawa, T. Sato, A. Tomita, K. Ichiyanagi, H. Ki, J. Kim, S. Adachi, and H. Ihee, "Solvent-dependent molecular structure of

ionic species directly measured by ultrafast x-ray solution scattering," Phys. Rev. Lett. **110**, 165505 (2013).

- <sup>18</sup>J. H. Lee, M. Wulff, S. Bratos, J. Petersen, L. Guerin, J.-C. Leicknam, M. Cammarata, Q. Kong, J. Kim, K. B. Møller, and H. Ihee, "Filming the birth of molecules and accompanying solvent rearrangement," J. Am. Chem. Soc. 135, 3255–3261 (2013).
- <sup>19</sup>H. Takala, A. Bjorling, O. Berntsson, H. Lehtivuori, S. Niebling, M. Hoernke, I. Kosheleva, R. Henning, A. Menzel, J. A. Ihalainen, and S. Westenhoff, "Signal amplification and transduction in phytochrome photosensors," Nature 509, 245–248 (2014).
- <sup>20</sup>C. W. Ahn, H. Ki, J. Kim, J. Kim, S. Park, Y. Lee, K. H. Kim, Q. Kong, J. Moon, M. N. Pedersen, M. Wulff, and H. Ihee, "Direct observation of a transiently formed isomer during iodoform photolysis in solution by time-resolved x-ray liquidography," J. Phys. Chem. Lett. 9, 647–653 (2018).
- <sup>21</sup>D. Leshchev, T. C. B. Harlang, L. A. Fredin, D. Khakhulin, Y. Liu, E. Biasin, M. G. Laursen, G. E. Newby, K. Haldrup, M. M. Nielsen, K. Wärnmark, V. Sundström, P. Persson, K. S. Kjaer, and M. Wulff, "Tracking the picosecond deactivation dynamics of a photoexcited iron carbene complex by time-resolved x-ray scattering," Chem. Sci. 9, 405–414 (2018).
- <sup>22</sup>D. Khakhulin, L. M. Lawson Daku, D. Leshchev, G. E. Newby, M. Jarenmark, C. Bressler, M. Wulff, and S. E. Canton, "Visualizing the coordination-spheres of photoexcited transition metal complexes with ultrafast hard x-rays," Phys. Chem. Chem. Phys. 21, 9277–9284 (2019).
- <sup>23</sup>Q. Kong, D. Khakhulin, I. A. Shkrob, J. H. Lee, X. Zhang, J. Kim, K. H. Kim, J. Jo, J. Kim, J. Kang, V.-T. Pham, G. Jennings, C. Kurtz, R. Spence, L. X. Chen, M. Wulff, and H. Ihee, "Solvent-dependent complex reaction pathways of bromoform revealed by time-resolved x-ray solution scattering and x-ray transient absorption spectroscopy," Struct. Dyn. 6, 064902 (2019).
- <sup>24</sup>D. Leshchev, D. Khakhulin, G. Newby, H. Ki, H. Ihee, and M. Wulff, "Subnanosecond secondary geminate recombination in mercury halides HgX<sub>2</sub> (X = I, Br) investigated by time-resolved x-ray scattering," J. Chem. Phys. 151, 054310 (2019).
- <sup>25</sup>H. Ki, S. Park, S. Eom, J. Gu, S. Kim, C. Kim, C. W. Ahn, M. Choi, S. Ahn, D. S. Ahn, J. Choi, M. H. Baik, and H. Ihee, "Gold nanoparticle formation via x-ray radiolysis investigated with time-resolved x-ray liquidography," Int. J. Mol. Sci. 21, 7125 (2020).
- <sup>26</sup>T. W. Kim, S. J. Lee, J. Jo, J. G. Kim, H. Ki, C. W. Kim, K. H. Cho, J. Choi, J. H. Lee, M. Wulff, Y. M. Rhee, and H. Ihee, "Protein folding from heterogeneous unfolded state revealed by time-resolved x-ray solution scattering," Proc. Natl. Acad. Sci. U. S. A. 117, 14996–15005 (2020).
- <sup>27</sup>S. J. Lee, Y. Kim, T. W. Kim, C. Yang, K. Thamilselvan, H. Jeong, J. Hyun, and H. Ihee, "Reversible molecular motional switch based on circular photoactive protein oligomers exhibits unexpected photo-induced contraction," Cell Rep. Phys. Sci. 2, 100512 (2021).
- <sup>28</sup>H. Jeong, H. Ki, J. G. Kim, J. Kim, Y. Lee, and H. Ihee, "Sensitivity of timeresolved diffraction data to changes in internuclear distances and atomic positions," Bull. Korean Chem. Soc. **43**, 376–390 (2022).
- <sup>29</sup>D. Arnlund, L. C. Johansson, C. Wickstrand, A. Barty, G. J. Williams, E. Malmerberg, J. Davidsson, D. Milathianaki, D. P. DePonte, R. L. Shoeman, D. J. Wang, D. James, G. Katona, S. Westenhoff, T. A. White, A. Aquila, S. Bari, P. Berntsen, M. Bogan, T. B. van Driel, R. B. Doak, K. S. Kjaer, M. Frank, R. Fromme, I. Grotjohann, R. Henning, M. S. Hunter, R. A. Kirian, I. Kosheleva, C. Kupitz, M. N. Liang, A. V. Martin, M. M. Nielsen, M. Messerschmidt, M. M. Seibert, J. Sjohamn, F. Stellato, U. Weierstall, N. A. Zatsepin, J. C. H. Spence, P. Fromme, I. Schlichting, S. Boutet, G. Groenhof, H. N. Chapman, and R. Neutze, "Visualizing a protein quake with time-resolved x-ray scattering at a free-electron laser," Nat. Methods 11, 923–926 (2014).
- <sup>30</sup>K. H. Kim, J. G. Kim, S. Nozawa, T. Sato, K. Y. Oang, T. W. Kim, H. Ki, J. Jo, S. Park, C. Song, T. Sato, K. Ogawa, T. Togashi, K. Tono, M. Yabashi, T. Ishikawa, J. Kim, R. Ryoo, J. Kim, H. Ihee, and S. Adachi, "Direct observation of bond formation in solution with femtosecond x-ray scattering," Nature 518, 385–389 (2015).
- <sup>31</sup>M. Levantino, G. Schiro, H. T. Lemke, G. Cottone, J. M. Glownia, D. L. Zhu, M. Chollet, H. Ihee, A. Cupane, and M. Cammarata, "Ultrafast myoglobin structural dynamics observed with an x-ray free-electron laser," Nat. Commun. 6, 6772 (2015).

- <sup>32</sup>E. Biasin, T. B. van Driel, K. S. Kjaer, A. O. Dohn, M. Christensen, T. Harlang, P. Chabera, Y. Liu, J. Uhlig, M. Papai, Z. Nemeth, R. Hartsock, W. Liang, J. Zhang, R. Alonso-Mori, M. Chollet, J. M. Glownia, S. Nelson, D. Sokaras, T. A. Assefa, A. Britz, A. Galler, W. Gawelda, C. Bressler, K. J. Gaffney, H. T. Lemke, K. B. Moller, M. M. Nielsen, V. Sundstrom, G. Vanko, K. Warnmark, S. E. Canton, and K. Haldrup, "Femtosecond x-ray scattering study of ultrafast photoinduced structural dynamics in solvated [Co(terpy)<sub>2</sub>]<sup>2+</sup>," Phys. Rev. Lett. 117,013002 (2016).
- 33K. H. Kim, J. G. Kim, K. Y. Oang, T. W. Kim, H. Ki, J. Jo, J. Kim, T. Sato, S. Nozawa, S. Adachi, and H. Ihee, "Femtosecond x-ray solution scattering reveals that bond formation mechanism of a gold trimer complex is independent of excitation wavelength," Struct. Dyn. 3, 043209 (2016).
- 34J. G. Kim, S. Nozawa, H. Kim, E. H. Choi, T. Sato, T. W. Kim, K. H. Kim, H. Ki, J. Kim, M. Choi, Y. Lee, J. Heo, K. Y. Oang, K. Ichiyanagi, R. Fukaya, J. H. Lee, J. Park, I. Eom, S. H. Chun, S. Kim, M. Kim, T. Katayama, T. Togashi, S. Owada, M. Yabashi, S. J. Lee, S. Lee, C. W. Ahn, D.-S. Ahn, J. Moon, S. Choi, J. Kim, T. Joo, J. Kim, S. Adachi, and H. Ihee, "Mapping the emergence of molecular vibrations mediating bond formation," Nature 582, 520-524 (2020).
- <sup>35</sup>E. Biasin, Z. W. Fox, A. Andersen, K. Ledbetter, K. S. Kjaer, R. Alonso-Mori, J. M. Carlstad, M. Chollet, J. D. Gaynor, J. M. Glownia, K. Hong, T. Kroll, J. H. Lee, C. Liekhus-Schmaltz, M. Reinhard, D. Sokaras, Y. Zhang, G. Doumy, A. M. March, S. H. Southworth, S. Mukamel, K. J. Gaffney, R. W. Schoenlein, N. Govind, A. A. Cordones, and M. Khalil, "Direct observation of coherent femtosecond solvent reorganization coupled to intramolecular electron transfer," Nat. Chem. 13, 343-349 (2021).
- <sup>36</sup>E. H. Choi, J. G. Kim, J. Kim, H. Ki, Y. Lee, S. Lee, K. Yoon, J. Kim, J. Kim, and H. Ihee, "Filming ultrafast roaming-mediated isomerization of bismuth triiodide in solution," Nat. Commun. 12, 4732 (2021).
- 37 H. Ki, S. Choi, J. Kim, E. H. Choi, S. Lee, Y. Lee, K. Yoon, C. W. Ahn, D.-S. Ahn, J. H. Lee, J. Park, I. Eom, M. Kim, S. H. Chun, J. Kim, H. Ihee, and J. Kim, "Optical Kerr effect of liquid acetonitrile probed by femtosecond timeresolved x-ray liquidography," J. Am. Chem. Soc. 143, 14261-14273 (2021).
- <sup>38</sup>Y. Lee, J. G. Kim, S. J. Lee, S. Muniyappan, T. W. Kim, H. Ki, H. Kim, J. Jo, S. R. Yun, H. Lee, K. W. Lee, S. O. Kim, M. Cammarata, and H. Ihee, "Ultrafast coherent motion and helix rearrangement of homodimeric hemoglobin visualized with femtosecond x-ray solution scattering," Nat. Commun. 12, 3677 (2021).
- <sup>39</sup>J. G. Kim, E. H. Choi, Y. Lee, and H. Ihee, "Femtosecond x-ray liquidography visualizes wavepacket trajectories in multidimensional nuclear coordinates for a bimolecular reaction," Acc. Chem. Res. 54, 1685-1698 (2021).
- <sup>40</sup>D. I. Svergun, "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing," Biophys. J. 76, 2879-2886 (1999).
- <sup>41</sup>D. Svergun, "Advanced solution scattering data analysis methods and their applications," J. Appl. Crystallogr. 33, 530-534 (2000).
- <sup>42</sup>D. I. Svergun, M. V. Petoukhov, and M. H. J. Koch, "Determination of domain structure of proteins from x-ray solution scattering," Biophys. J. 80, 2946-2953 (2001).
- 43H. Ihee, M. Lorenc, T. K. Kim, Q. Y. Kong, M. Cammarata, J. H. Lee, S. Bratos, and M. Wulff, "Ultrafast x-ray diffraction of transient molecular structures in solution," Science 309, 1223-1227 (2005).
- <sup>44</sup>A. Panjkovich and D. I. Svergun, "Deciphering conformational transitions of proteins by small angle x-ray scattering and normal mode analysis," Phys. Chem. Chem. Phys. 18, 5707-5719 (2016).

- <sup>45</sup>K. Y. Oang, C. Yang, S. Muniyappan, J. Kim, and H. Ihee, "SVD-aided pseudo principal-component analysis: A new method to speed up and improve determination of the optimum kinetic model from time-resolved data," Struct. Dyn. 4,044013 (2017).
- <sup>46</sup>D. Franke, C. M. Jeffries, and D. I. Svergun, "Machine learning methods for xray scattering data analysis from biomacromolecular solutions," Biophys. J. 114, 2485-2492 (2018).
- 47H. Ki, Y. Lee, E. H. Choi, S. Lee, and H. Ihee, "SVD-aided non-orthogonal decomposition (SANOD) method to exploit prior knowledge of spectral components in the analysis of time-resolved data," Struct. Dyn. 6, 024303 (2019).
- <sup>48</sup>M. Choi, J. G. Kim, S. Muniyappan, H. Kim, T. W. Kim, Y. Lee, S. J. Lee, S. O. Kim, and H. Ihee, "Effect of the abolition of intersubunit salt bridges on allosteric protein structural dynamics," Chem. Sci. 12, 8207-8217 (2021).
- 49Q. Y. Kong, J. H. Lee, A. Plech, M. Wulff, H. Ihee, and M. H. J. Koch, "Ultrafast x-ray solution scattering reveals an unknown reaction intermediate in the photolysis of [Ru-3(CO)12]," Angew. Chem., Int. Ed. 47, 5550-5553 (2008).
- 50 K. S. Kjaer, T. B. v Driel, J. Kehres, K. Haldrup, D. Khakhulin, K. Bechgaard, M. Cammarata, M. Wulff, T. J. Sørensen, and M. M. Nielsen, "Introducing a standard method for experimental determination of the solvent response in laser pump, x-ray probe time-resolved wide-angle x-ray scattering experiments on system in solution," Phys. Chem. Chem. Phys. 15, 15003-15016 (2013).
- <sup>51</sup>M. Cammarata, M. Lorenc, T. K. Kim, J. H. Lee, Q. Y. Kong, E. Pontecorvo, M. Lo Russo, G. Schiro, A. Cupane, M. Wulff, and H. Ihee, "Impulsive solvent heating probed by picosecond x-ray diffraction," J. Chem. Phys. 124, 124504 (2006).
- <sup>52</sup>M. Bierlaire, P. L. Toint, and D. Tuyttens, "On iterative algorithms for linear least squares problems with bound constraints," Linear Algebra Appl. 143, 111-143 (1991).
- <sup>53</sup>Å. Björck, "Solving linear least squares problems by Gram-Schmidt orthogonalization," BIT Numer. Math. 7, 1-21 (1967).
- 54G. H. Golub, P. Manneback, and P. L. Toint, "A comparison between some direct and iterative methods for certain large scale geodetic least squares problems," SIAM J. Sci. Stat. Comput. 7, 799-816 (1986).
- <sup>55</sup>G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," Numer, Math. 14, 403-420 (1970).
- <sup>56</sup>M. Schmidt, S. Rajagopal, Z. Ren, and K. Moffat, "Application of singular value" decomposition to the analysis of time-resolved macromolecular x-ray data," Biophys. J. 84, 2112–2129 (2003).
- 57Y. O. Jung, J. H. Lee, J. Kim, M. Schmidt, K. Moffat, V. Srajer, and H. Ihee, "Volume-conserving trans-cis isomerization pathways in photoactive yellow protein visualized by picosecond x-ray crystallography," Nat. Chem. 5, 212-220 (2013).
- 58G. Strang, Introduction to Linear Algebra, 5th ed. (Cambridge University Press, 2021), pp. 206-213.
- 59C. D. Meyer, Matrix Analysis and Applied Linear Algebra (SIAM, 2000),
- p. 439. <sup>60</sup>N. J. Salkind, *Encyclopedia of Measurement and Statistics* (SAGE Publications, 2006), pp. 530-536.
- <sup>61</sup>L. Hogben, Handbook of Linear Algebra (CRC Press, 2007), p. 86.
- <sup>62</sup>G. Golub, "Numerical methods for solving linear least squares problems," Numer. Math. 7, 206-216 (1965).

28