MatchMaps: Non-isomorphous difference maps for X-ray crystallography

Dennis E. Brookner
Department of Molecular and Cellular Biology
Harvard University
Cambridge, MA

Doeke R. Hekstra
Department of Molecular and Cellular Biology
Harvard University
Cambridge, MA

Abstract

Conformational change mediates the biological functions of proteins. Crystallographic measurements can map these changes with extraordinary sensitivity as a function of mutations, ligands, and time. The isomorphous difference map remains the gold standard for detecting structural differences between datasets. Isomorphous difference maps combine the phases of a chosen reference state with the observed changes in structure factor amplitudes to yield a map of changes in electron density. Such maps are much more sensitive to conformational change than structure refinement is, and are unbiased in the sense that observed differences do not depend on refinement of the perturbed state. However, even minute changes in unit cell dimensions can render isomorphous difference maps useless. This is unnecessary. Here we describe a generalized procedure for calculating observed difference maps that retains the high sensitivity to conformational change and avoids structure refinement of the perturbed state. We have implemented this procedure in an open-source python package, MatchMaps, that can be run in any software environment supporting PHENIX and CCP4. Through examples, we show that MatchMaps "rescues" observed difference electron density maps for near-isomorphous crystals, corrects artifacts in nominally isomorphous difference maps, and extends to detecting differences across copies within the asymmetric unit, or across altogether different crystal forms.

1 Introduction

X-ray crystallography provides a powerful method for characterizing the changes in protein structure caused by a perturbation [13, 15, 3, 6]. For large structural changes and large changes in occupancy of states, it is usually sufficient to refine separate structural models for each dataset and draw comparisons between the refined structures. However, for many conformational changes, coordinate-based comparisons are inaccurate and insensitive.

In crystallography, electron density is not observed directly, but rather a diffraction pattern consisting of reflections with intensities proportional to the squared amplitudes of Fourier components of the electron density (structure factors). Unfortunately, the phases of these structure factors are not observable. These phases correspond in real space to shifts of the sinusoidal waves that add up to an electron density pattern. Accordingly, phases are usually calculated from a refined model. Since phases have a strong effect on the map appearance [19], naive electron density maps calculated using observed amplitudes and model-based phases will tend to resemble the model, a phenomenon known as model bias.

The gold standard for detecting conformational change in crystallographic data is the isomorphous difference map[20]. An isomorphous difference map is computed by combining differences in observed structure factor amplitudes with a single set of phases. The phases are usually derived from a model for one of the two states, chosen as a reference. That is, the "observed" difference density $\Delta\rho(\mathbf{x})$ is approximated as:

$$\Delta\rho(\mathbf{x}) = \left(|F_{\mathbf{h}}^{ON}| - |F_{\mathbf{h}}^{OFF}|\right)e^{i(2\pi\mathbf{h}^T\mathbf{x}+\phi_{c,\mathbf{h}}^{OFF})} \tag{1}$$

where $|F_{\mathbf{h}}^{ON}|$ and $|F_{\mathbf{h}}^{OFF}|$ are sets of observed structure factor amplitudes from the ON (perturbed) and OFF (reference) datasets respectively, $\mathbf{h}$ is shorthand for the triplet of Miller indices (h, k, l), $\mathbf{x}$ is shorthand for the point (x, y, z) (in fractional coordinates) and $e^{i\phi_{c,\mathbf{h}}^{OFF}}$ is a set of structure factor phases derived from refining a structural model to the OFF data. Crucially, therefore, isomorphous difference maps are computed without information regarding the ON structure. Any difference electron density relating to the ON data relative to the OFF data (e.g., positive difference density for a bound ligand) is thus guaranteed not to be biased by previous modeling of the ON state.

Many pairs of crystallographic datasets are not perfectly isomorphous, or not at all. Here we first discuss isomorphous difference maps and then introduce an algorithm, MatchMaps, that can calculate difference maps based on observed structure factor amplitudes without reliance on isomorphism. We illustrate the approach through examples for different cases: the near-isomorphous case; protein molecules within the same asymmetric unit; and the case of entirely different crystal forms. The software implementation of MatchMaps is open-source and readily extensible.

1.1 Limitations of isomorphous difference maps

The isomorphous difference map relies on the assumption that a set of "reference" structure factor phases can also be used as the phases of a difference map. Specifically, it must hold true that for structure factors with large differences in amplitude, the phases are similar; thus, the isomorphous difference structure factor is a good approximation of the true difference structure factor. Conversely, when the phase differences are large, the amplitudes need to be similar such that these terms contribute little to the isomorphous difference map[14, 20].

Crystallographic phases are, however, highly sensitive to even small rotations and translations of protein molecules within the unit cell, as well as changes in solvent volume. These effects commonly lead to changes in the crystal's unit cell. Indeed, in crystallographic lore, a 1 percent variation of each cell axis is the cutoff for proper isomorphism. In

Figure 1, we present an example of datasets for *E. coli* dihydrofolate reductase. For the isomorphous case, we see that the phases are highly similar, and with few large differences in the amplitudes of structure factors that differ significantly in phase. Thus this pair of datasets is suitable for calculation of an isomorphous difference map. In contrast, even a 2% change along a single unit cell dimension ("Nonisomorphous" in Figure 1) is sufficient to obliterate the correlation among phases. Unfortunately, unit cell changes are often responsive to the exact changes one hopes to study (in the example of Figure 2b, a conformational change of the active site).

Of course, often an isomorphous difference map cannot be computed at all, for example due to differences in crystal form or, within an asymmetric unit, lack of crystallographic symmetry.

1.2 Rethinking isomorphous difference maps via the linearity of the Fourier transform

An isomorphous difference map is typically computed by first subtracting the structure factors (in reciprocal space) and then applying the Fourier transform to convert the structure factor differences into a real-space difference map. However, because the Fourier transform and subtraction are both linear operations, their order can be switched without changing the result: one might just as well calculate two electron density maps first and then subtract those maps voxel-by-voxel to yield an isomorphous difference map.

This reordering suggests how difference map computation can be generalized beyond the isomorphous case. Specifically, we see that the step in the algorithm most specific to the assumption of isomorphism is the creation of "hybrid" structure factors of the form:

$$|F_h^{ON}|e^{i\phi_{c,h}^{OFF}} \tag{2}$$

The OFF phases used in this calculation were computed with regards to a unit cell and molecular orientation that may be incompatible with the ON structure factor amplitudes. The method presented here improves these "hybrid" structure factors by computing phases that account for the (generally uninteresting) shifts in molecular position and orientation without removing any signal associated with "interesting" changes.

2 The MatchMaps algorithm

The goal of MatchMaps is to achieve the best possible real-space difference density map without utilizing a prior model of the structural changes of interest. To compute a real-space difference density map, one first needs to approximate structure factor phases for each dataset. As discussed above, the isomorphous difference map makes the simplifying assumption that the same set of structure factor phases can be used for both structures.

The key to MatchMaps is to improve phases for the ON data via rigid-body refinement of the OFF starting model against the ON structure factor amplitudes. This rigid-body refinement step improves phases by optimally placing the protein model in space. However, the restriction of this refinement to only whole-model rigid-body motion protects these new phases from bias towards modeled structural changes. The result is two sets of complex structure factors which make use of the information encoded in the structure factor amplitudes without relying on a second input model.

Next, each set of complex structure factors is Fourier-transformed into a real-space electron density map. These two real-space maps will not necessarily overlay in space. However, the rotation and translation necessary to overlay the maps can be obtained from the results of rigid-body refinement. Following real-space alignment, the maps can be subtracted voxel-wise to compute a difference map.

In the idealized case—similar structures, oriented identically in space, with identical unit cells—MatchMaps will perform essentially identically to an isomorphous difference map. However, as we show in the examples below, MatchMaps is more capable than a traditional isomorphous difference map of handling datasets that diverge from this ideal. Furthermore, even in seemingly simple cases where isomorphous difference maps perform well, the real-space MatchMaps approach can show distinct improvements.

2.1 Details of algorithmic implementation

The full MatchMaps algorithm is as follows. As inputs, the algorithm requires two sets of structure factor amplitudes (referred to as ON and OFF datasets, for simplicity) and a single starting model (corresponding to the OFF data).

1. Place both sets of structure factor amplitudes on a common scale using CCP4[1]'s `SCALEIT`[14] utility and truncate the data to the same resolution range.
2. Generate phases for each dataset via the `phenix.refine`[16] program. For each dataset, the OFF starting model is used, and only rigid-body refinement is permitted to prevent the introduction of model bias.
3. Fourier-transform each set of complex structure factors into a real-space electron density map using the python packages `reciprocalspaceship`[11] and `gemmi`[24].
4. Compute the translation and rotation necessary to overlay the two rigid-body refined models. Apply this translation-rotation to the ON real-space map such that it overlays with the OFF map. These computations are carried out using `gemmi`.
5. Place both real-space maps on a common scale.
6. Subtract real-space maps voxel-wise.
7. Apply a solvent mask to the final difference map.

We note that MatchMaps is structured such that step 2 can be generalized to refinement of any uninteresting features, not just rigid-body motion, if the user provides a custom PHENIX parameter file, as specified in the online documentation. For example, if the starting model contains multiple protein chains, each chain could be rigid-body-refined separately.

2.2 Installation

MatchMaps can be installed using the `pip` python package manager (`pip install matchmaps`). The various pure-python dependencies of MatchMaps are handled by `pip`. Additionally, MatchMaps requires installation of the popular

CCP4 and Phenix software suites for crystallography. Once installed, the above protocol can be run in a single step from the command line.

In addition to the base MatchMaps command-line utility, the utilities `matchmaps.ncs` and `matchmaps.mr` provide additional functionalities explored in the examples below and the online documentation. MatchMaps is fully open-source and readily extensible for novel use cases.

For more information, read the MatchMaps documentation at https://rs-station.github.io/matchmaps.

3 Examples

The following examples explore the benefits and functionalities offered by MatchMaps. All examples make use of published crystallographic data available from the Protein Data Bank. Scripts and data files for reproducing the figures can be found on Zenodo.

3.1 MatchMaps map where isomorphous difference map falters

The enzyme Dihydrofolate Reductase (DHFR) is a central model system for understanding the role of conformational change in productive catalytic turnover[21, 5, 4]. Specifically, the active-site Met20 loop of *E. coli* DHFR can adopt several different conformations, each stabilized by specific bound ligands and crystal contacts[21]. DHFR bound to $NADP^+$ and substrate analog folate adopts a "closed" Met20 loop (PDB ID 1RX2), whereas DHFR bound to $NADP^+$ and product analog (dideazatetrahydrofolate) adopts an "occluded" Met20 loop (PDB ID 1RX4). These structures are highly similar, other than the relevant changes at the active site (Figure 2a, structural changes shown in boxes; RMSD 0.37 Å for protein C-alpha atoms excluding Met20 loop).

Importantly, the presence of the occluded loop conformation leads to altered crystal packing wherein the crystallographic *b* axis increases by 2%, from 98.91 Å to 100.88 Å (Figure 2b). This means that structures 1RX2 and 1RX4, though extremely similar, cannot be effectively compared by an isomorphous difference map (Figure 2c,g). We illustrated the striking change in phases between these structures in Figure 1. MatchMaps is able to correct for this non-isomorphism and recover the expected difference signal.

First, we focused on ligand rearrangement in the active site. In the occluded-loop structure, the cofactor (Figure 2c-f, left) leaves the active site while the substrate (Figure 2c-f, right) slides laterally within the active site. MatchMaps shows this expected signal, with negative (red) difference density for the cofactor and positive (blue) difference density for the substrate (Figure 2e-f). By contrast, an isomorphous difference map (Figure 2c) is unable to recover this signal. The structural model for the occluded-loop structure is shown for clarity in Figure 2f as blue sticks and clearly matches with the positive difference density. Importantly, this ON model is never used in the computation of the MatchMaps map.

We find a similar result around residues 21-25 of the Met20 loop (Figure 2g-j). Again, MatchMaps shows readily interpretable difference signal for the change in loop conformation between the closed-loop (red) and occluded-loop

(blue) structures (Figure 2i-j). The isomorphous difference map, on the other hand, contains no interpretable signal in this region of strong structural change (Figure 2g). The occluded-loop model is shown for visual comparison in Figure 2j but was not used for computation of the MatchMaps map.

Finally, we visualize the "expected" structural changes by first aligning the refined models and placing them in the same unit cell, then computing F_{model} phases for each and subtracting them. The MatchMaps difference maps correspond well with these maps (Figure 2d,h).

3.2 Unbiased modeling of ligand binding

The enzyme Protein Tyrosine Phosphatase 1B (PTP1B) plays a key role in insulin signaling[8], making it a long-standing target for the treatment of diabetes using ortho- and allosteric by drug design [23, 15, 7]. For illustration, we compare recent high-quality room-temperature structures of the apo protein (PDB ID 7RIN) with the protein bound to the competitive inhibitor TCS401 (PDB ID 7MM1)[12]. In addition to the presence/absence of signal for the ligand itself, the apo structure exhibits an equilibrium between "open" and "closed" active-site loops[22], whereas the bound structure shows only the closed loop.

The datasets 7RIN and 7MM1 are sufficiently isomorphous that an isomorphous difference map reveals the main structural changes. MatchMaps performs similarly. Strong positive difference density (blue mesh) is seen for the TCS401 ligand (grey sticks) in both the isomorphous difference map (Figure 3a) and the MatchMaps difference map (Figure 3b). Around residues 180-182 of the active-site loop (known as the WPD loop), both the isomorphous difference map (Figure 3c) and MatchMaps difference map (Figure 3d) show strong signal for the decrease in occupancy (red mesh) of the open loop conformation (red sticks) and an increase in occupancy (blue mesh) of the closed loop conformation (blue sticks).

However, even in this seemingly straightforward case, we find that the isomorphous difference map is susceptible to an artifact resulting from a slight (1.37 degrees) rotation of the protein. The displacement between the original refined structural coordinates of each structure is especially strong around residues 22-25 (Figure 3e, apo model in gray, bound model in blue). In this region, an isomorphous difference map picks up on this artifactual difference between the datasets and displays strong difference signal (blue and red mesh). Remarkably, this signal is similar in magnitude to the "true" signal described above. In contrast, MatchMaps internally aligns the models after the computation of phases and before subtraction. Figure 3f shows residues 22-25 following whole-molecule alignment of the protein models (apo model in red, bound model in blue). Following global alignment of the refined models, it is clear that this region does not contain any "interesting" signal. Sure enough, the MatchMaps difference map contains no strong signal in this region.

3.3 `matchmaps.mr`: comparing data from different spacegroups

For many protein systems, careful analysis of electron density change is stymied for pairs of similar structures which crystallize in different crystal forms. The MatchMaps algorithm can be further generalized to allow comparison of

datasets in entirely different crystal packings or spacegroups. Specifically, the OFF model can serve as a search model for molecular replacement for the ON data. Following this extra step, the algorithm proceeds identically. We implement this modified algorithm in the commandline utility `matchmaps.mr`.

One such example is the enzyme DHFR, which has been crystallized in many spacegroups[21]. Here, we examine two structures of the enzyme bound to $NADP^+$, in spacegroups $P2_12_12_1$ (PDB ID 1RX1) and $C2$ (PDB ID 1RA1), visualized in Figure 4a. These structures are overall similar, but differ in the active site (Figure 4b-d). Here, we visualize these structural changes directly in electron density without introducing model bias.

Specifically, in the $P2_12_12_1$ structure, the active site Met20 loop adopts a closed conformation. In the $C2$ structure, the Met20 loop adopts an "open" conformation, which is stabilized by a crystal contact in this crystal form[21]. The difference between the open and closed loops is exemplified by residues 17-24 (Figure 4c). The closed loop is stabilized by the formation of a key hydrogen bond between the Asn23 backbone and the Ser148 sidechain. In the open conformation, Asn23 is too far from Ser148 to form a hydrogen bond (Figure 4d).

Remarkably, the positive difference density (red) for the closed loop is strong and readily interpretable in Figures 4c-d. The MatchMaps map was computed only using the $P2_12_12_1$ (blue) closed-loop model. This means that the signal for the open loop conformation is derived only from the observed structure factor amplitudes for the open-loop state in an unrelated crystal form!

3.4 `matchmaps.ncs`: comparing NCS-related molecules

The real-space portion of the MatchMaps algorithm can be repurposed to create "internal" difference maps across non-crystallographic symmetry (NCS) operations. As an example, we examined the crystal structure of the fifth PDZ domain (PDZ5) from the *Drosophila* protein Inactivation, no after-potential D (INAD). This domain plays an essential role in terminating the response of photoreceptors to absorbed photons by modulation of its ability to bind ligands [17]. In particular, the binding cleft of PDZ5 can be locked by formation of a disulfide bond between residues C606 and C645. PDZ5 was found to crystallize in a form with three molecules in the asymmetric unit (Figure 5a). Remarkably, each of these molecules adopts a different state. Specifically, chain C contains a disulfide bond between residues Cys606 and Cys645, whereas chain B does not. Chain A adopts a bound state by binding the C terminus of chain C. MatchMaps enables calculation of an internal difference map, yielding a clearly interpretable difference map of the formation of the disulfide bond (Figure 5c).

4 Discussion

The isomorphous difference map has been the gold standard for detecting conformational change for many years [14, 20]. However, we show above that the same inputs—one structural model and two sets of structure factor amplitudes—can be combined to compute a difference map that shares the strength of an isomorphous difference map while ameliorating a key weakness. Specifically, structure factor phases are highly sensitive not only to structural changes ("interesting" signal) but also to changes in unit cell dimensions and model pose ("uninteresting" signal). The introduction of

rigid-body refinement minimizes the contribution of this uninteresting signal to the final difference map. In Figure 2, we illustrate a case where a loss of isomorphism significantly degrades the signal of an isomorphous difference map. In this case, MatchMaps is still able to recover the expected difference signal. Figure 3 shows a situation where the isomorphous difference map performs well for the strongest difference signal. However, even in this seemingly straightforward use case, the isomorphous difference map is still susceptible to "uninteresting signal". MatchMaps is able to successfully remove this artifact.

In our experience, the results from crystallographic perturbation experiments are often shelved due to changes in unit cell constants. Interesting conformational changes often slightly alter the unit cell. Unit cells constants are also sensitive to temperature[10], radiation damage[18], pressure[2], and humidity[9], meaning that even data collected on the same crystal may not be isomorphous. MatchMaps removes, in principle, the requirement for isomorphism and allows for the analysis of far more crystallographic differences.

Furthermore, the computation of an isomorphous difference map is entirely incompatible with data from different crystal forms. The `matchmaps.mr` extension of MatchMaps allows for model-bias-free comparisons of electron densities regardless of crystal form, opening up a new world of structural comparisons. For instance, an isomorphous difference map cannot characterize the impacts of crystal packing. As shown above, MatchMaps can create such a map and thus allows enhanced understanding of the often subtle role of crystal packing on protein structure.

MatchMaps depends only on the common CCP4 and Phenix crystallographic suites along with various automatically installed pure-python dependencies. MatchMaps runs in minutes on a modern laptop computer. The only required input files are a PDB or mmCIF file containing the protein model, two MTZ files containing structure factor amplitudes and uncertainties, and any CIF ligand restraint files necessary for refinement. These are the same inputs required for many common purposes (such as running `phenix.refine`) and would likely already be on hand. As outputs, MatchMaps produces real-space maps in the common MAP/CCP4/MRC format which can be readily opened in molecular visualization software such as PyMOL or Coot. For these reasons, MatchMaps should slot naturally into the crystallographer's workflow for analysis of related datasets. Additionally, MatchMaps is open source and can be easily modified for a new use case by an interested developer. The authors welcome issues and pull requests on GitHub for the continued improvement of the software.

5 Acknowledgements

We thank Harrison Wang (Harvard University) for testing of the MatchMaps code. We thank Marcin Wojdyr (Global Phasing Ltd.) for assistance with the gemmi library. This work was supported by the NIH Director's New Innovator Award (DP2-GM141000, to D.R.H.).

References

[1] Jon Agirre et al. "The CCP4 suite: integrative software for macromolecular crystallography". In: *Acta crystallographica. Section D, Structural biology* 79.March (2023), pp. 449–461. ISSN: 20597983. DOI: 10.1107/S2059798323003595.

[2] Buz Barstow et al. "Alteration of citrine structure by hydrostatic pressure explains the accompanying spectral shift". In: *Proceedings of the National Academy of Sciences of the United States of America* 105.36 (2008), pp. 13362–13366. ISSN: 00278424. DOI: 10.1073/pnas.0802252105.

[3] Gira Bhabha, Justin T. Biel, and James S. Fraser. "Keep on moving: Discovering and perturbing the conformational dynamics of enzymes". In: *Accounts of Chemical Research* 48.2 (Feb. 2015), pp. 423–430. ISSN: 15204898. DOI: 10.1021/ar5003158.

[4] Gira Bhabha et al. "A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis". In: *Science* 332.6026 (Apr. 2011), pp. 234–238. ISSN: 00368075. DOI: 10.1126/science.1198542. URL: www.sciencemag.org/cgi/content/full/science.1199412/DC1.

[5] David D. Boehr et al. "The dynamic energy landscape of dihydrofolate reductase catalysis". In: *Science* 313.5793 (Sept. 2006), pp. 1638–1642. ISSN: 00368075. DOI: 10.1126/science.1130258.

[6] Gisela Brändén and Richard Neutze. "Advances and challenges in time-resolved macromolecular crystallography". In: *Science* 373.6558 (Aug. 2021). ISSN: 0036-8075. URL: https://doi.org/10.1126/science.aba0954%20https://science.sciencemag.org/content/373/6558/eaba0954.

[7] Meng S. Choy et al. "Conformational Rigidity and Protein Dynamics at Distinct Timescales Regulate PTP1B Activity and Allostery". In: *Molecular Cell* 65.4 (2017), pp. 644–658. ISSN: 10974164. DOI: 10.1016/j.molcel.2017.01.014. URL: http://dx.doi.org/10.1016/j.molcel.2017.01.014.

[8] Mounib Elchebly et al. "Increased Insulin Sensitivity and Obesity Resistance in Mice Lacking the Protein Tyrosine Phosphatase-1B Gene". In: *Science* 283.5407 (Mar. 1999), pp. 1544–1548. ISSN: 00368075. DOI: 10.1126/SCIENCE.283.5407.1544. URL: https://www.science.org/doi/10.1126/science.283.5407.1544.

[9] Christopher Farley et al. "Improved reproducibility of unit-cell parameters in macromolecular cryocrystallography by limiting dehydration during crystal mounting". In: *Acta Crystallographica Section D: Biological Crystallography* 70.8 (2014), pp. 2111–2124. ISSN: 13990047. DOI: 10.1107/S1399004714012310.

[10] James S. Fraser et al. "Accessing protein conformational ensembles using room-temperature X-ray crystallography". In: *Proceedings of the National Academy of Sciences of the United States of America* 108.39 (Sept. 2011), pp. 16247–16252. ISSN: 00278424. DOI: 10.1073/pnas.1111325108.

[11] J.B. Greisman, K.M. Dalton, and D.R. Hekstra. "reciprocalspaceship: a Python library for crystallographic data analysis". In: *Journal of Applied Crystallography* 54.5 (Sept. 2021), pp. 1521–1529. ISSN: 1600-5767. DOI: 10.1107/S160057672100755X. URL: http://scripts.iucr.org/cgi-bin/paper?te5079%20https://journals.iucr.org/j/issues/2021/05/00/te5079/.

[12] Jack B. Greisman et al. "Native SAD phasing at room temperature". In: *Acta Crystallographica Section D: Structural Biology* 78.8 (Aug. 2022), pp. 986–996. ISSN: 20597983. DOI: 10.1107/S2059798322006799/RR5221SUP1.PDF. URL: http://scripts.iucr.org/cgi-bin/paper?rr5221%20https://journals.iucr.org/d/issues/2022/08/00/rr5221/.

[13] Doeke R. Hekstra et al. "Electric-field-stimulated protein mechanics". In: *Nature* 540.7633 (Dec. 2016), pp. 400–405. ISSN: 14764687. DOI: 10.1038/nature20571.

[14] R. Henderson and J. K. Moffat. "The difference Fourier technique in protein crystallography: errors and their treatment". In: *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry* 27.7 (1971), pp. 1414–1420. ISSN: 0567-7408. DOI: 10.1107/s0567740871004060.

[15] Daniel A. Keedy et al. "An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering". In: *eLife* 7 (2018), pp. 1–36. ISSN: 2050084X. DOI: 10.7554/eLife.36307.

[16] Dorothee Liebschner et al. "Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix". In: *Acta Crystallographica Section D: Structural Biology* 75 (2019), pp. 861–877. ISSN: 20597983. DOI: 10.1107/S2059798319011471.

[17] Prashant Mishra et al. "Dynamic Scaffolding in a G Protein-Coupled Signaling System". In: *Cell* 131.1 (2007), pp. 80–92. ISSN: 00928674. DOI: 10.1016/j.cell.2007.07.037.

[18] Raimond B.G. Ravelli and Sean M. McSweeney. "The 'fingerprint' that X-rays can leave on structures". In: *Structure* 8.3 (2000), pp. 315–328. ISSN: 09692126. DOI: 10.1016/S0969-2126(00)00109-X.

[19] R. J. Read. "Improved Fourier coefficients for maps using phases from partial structures with errors". In: *Acta Crystallographica Section A* 42.3 (1986), pp. 140–149. ISSN: 16005724. DOI: 10.1107/S0108767386099622.

[20] Mark A. Rould and Charles W. Carter. "Isomorphous Difference Methods". In: *Methods in Enzymology* 374 (2003), pp. 145–163. ISSN: 00766879. DOI: 10.1016/S0076-6879(03)74007-5.

[21] Michael R. Sawaya and Joseph Kraut. "Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: Crystallographic evidence". In: *Biochemistry* 36.3 (Jan. 1997), pp. 586–603. ISSN: 00062960. DOI: 10.1021/bi962337c.

[22] Sean K. Whittier, Alvan C. Hengge, and J. Patrick Loria. "Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases". In: *Science* 341.6148 (Aug. 2013), pp. 899–903. ISSN: 10959203. DOI: 10.1126/science.1241735. URL: www.sciencemag.org/cgi/content/full/science.1239951/DC1.

[23] Christian Wiesmann et al. "Allosteric inhibition of protein tyrosine phosphatase 1B". In: *Nature Structural & Molecular Biology 2004 11:8* 11.8 (July 2004), pp. 730–737. ISSN: 1545-9985. DOI: 10.1038/nsmb803. URL: https://www.nature.com/articles/nsmb803.

[24] Marcin Wojdyr. "GEMMI: A library for structural biology". In: *Journal of Open Source Software* 7.73 (2022), p. 4200. ISSN: 2475-9066. DOI: 10.21105/joss.04200.

6 Figures and figure legends

6.1 Figure 1

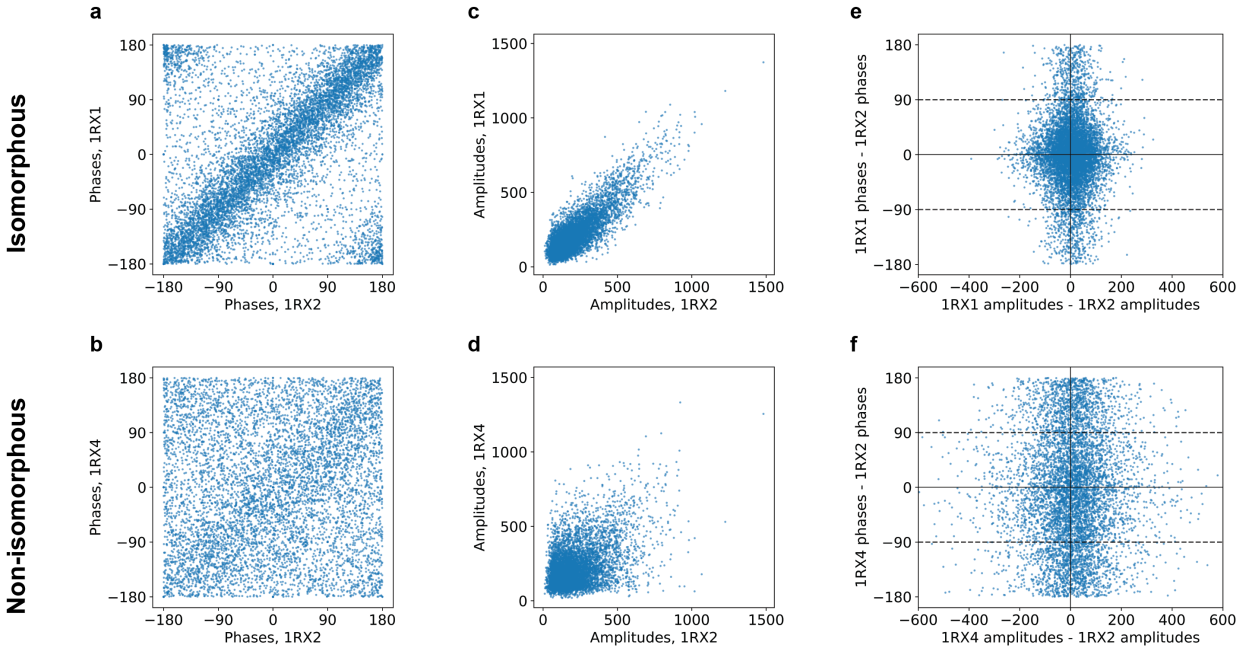

Figure 1: Structure factors depend sensitively on unit cell dimensions Here, we compare the *E. coli* DHFR dataset 1RX2 with an isomorphous structure (1RX1) and with a non-isomorphous structure (1RX4, see also figure 2). (a-b) Structure factor phases are (a) highly correlated for the isomorphous structures, but (b) mostly uncorrelated between non-isomorphous structures. In both cases, computed phases are obtained from the "PHIC" column of the deposited MTZ files. (c-d) Structure factor amplitudes are less correlated in the non-isomorphous case. (e-f) For an isomorphous difference map to be meaningful, structure factor amplitudes should only differ when the phase difference is small, and structure factor phases should only differ when the amplitude difference is small (see ref. [20], figure 2). This requirement is met in the isomorphous case (e), but not in the non-isomorphous case (f): large amplitude differences are also seen when phase differences are large.

6.2 Figure 2

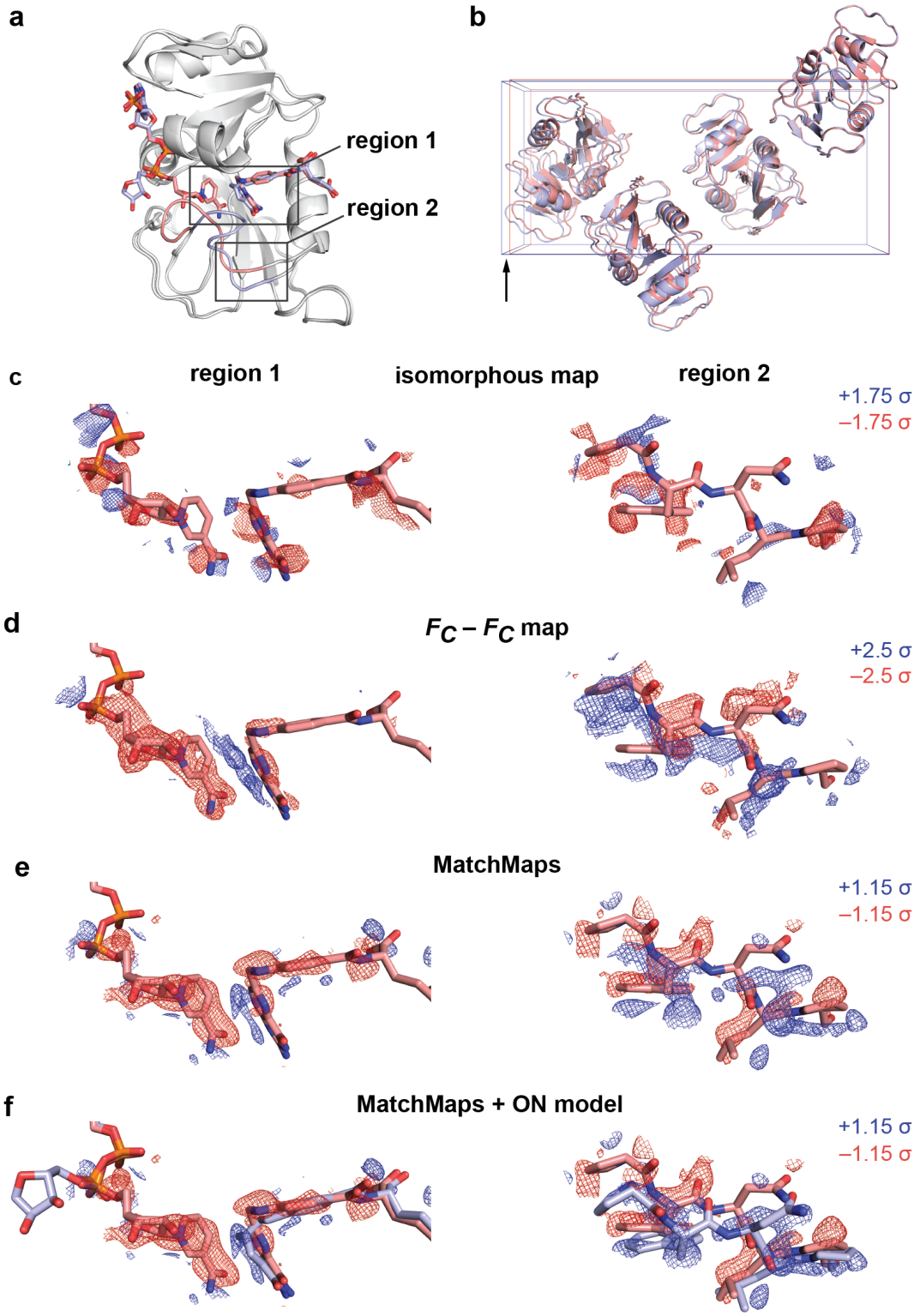
a
region 1
region 2
b
c
region 1
isomorphous map
region 2
+1.75 σ
−1.75 σ
d
$F_C - F_C$ map
+2.5 σ
−2.5 σ
e
MatchMaps
+1.15 σ
−1.15 σ
f
MatchMaps + ON model
+1.15 σ
−1.15 σ

Figure 2: MatchMaps for near-isomorphous structures (a) The structures 1RX2 and 1RX4 are overall similar (gray cartoons). The structures differ mainly at the active site loop (1RX2, pink cartoon; 1RX4, blue cartoon) and in the positions of the active site ligands (1RX2, pink sticks; 1RX4, blue sticks). (b) The unit cells of 1RX2 and 1RX4 differ by 2% along the longest dimension (left to right in this figure) from 98.912 Åto 100.879 Å. (c-f) Visualization of the change in ligand position between 1RX2 (red sticks) and 1RX4 (f, blue sticks). Positive difference density is shown as blue mesh; negative different density is shown as red mesh. Importantly, the 1RX4 structural coordinates were not used in the creation of the the isomorphous or MatchMaps maps. The isomorphous difference map contains essentially no interpretable signal. In contrast, the MatchMaps map (e-f) contains clear signal for disappearance of the cofactor and the lateral sliding of the substrate. Panel (d) contains an Fc-Fc map computed as a "ground truth" map of the expected signal. (g-j) Visualization of the change in loop conformation between 1RX2 and 1RX4. Only protein residues 21-25 are shown. Coloring is as in panels c-f. (g) Again, the isomorphous difference map is not interpretable. (i-j) The MatchMaps positive difference density clearly corresponds with the 1RX4 structural model, which was not used in the creation of the map.

6.3 Figure 3

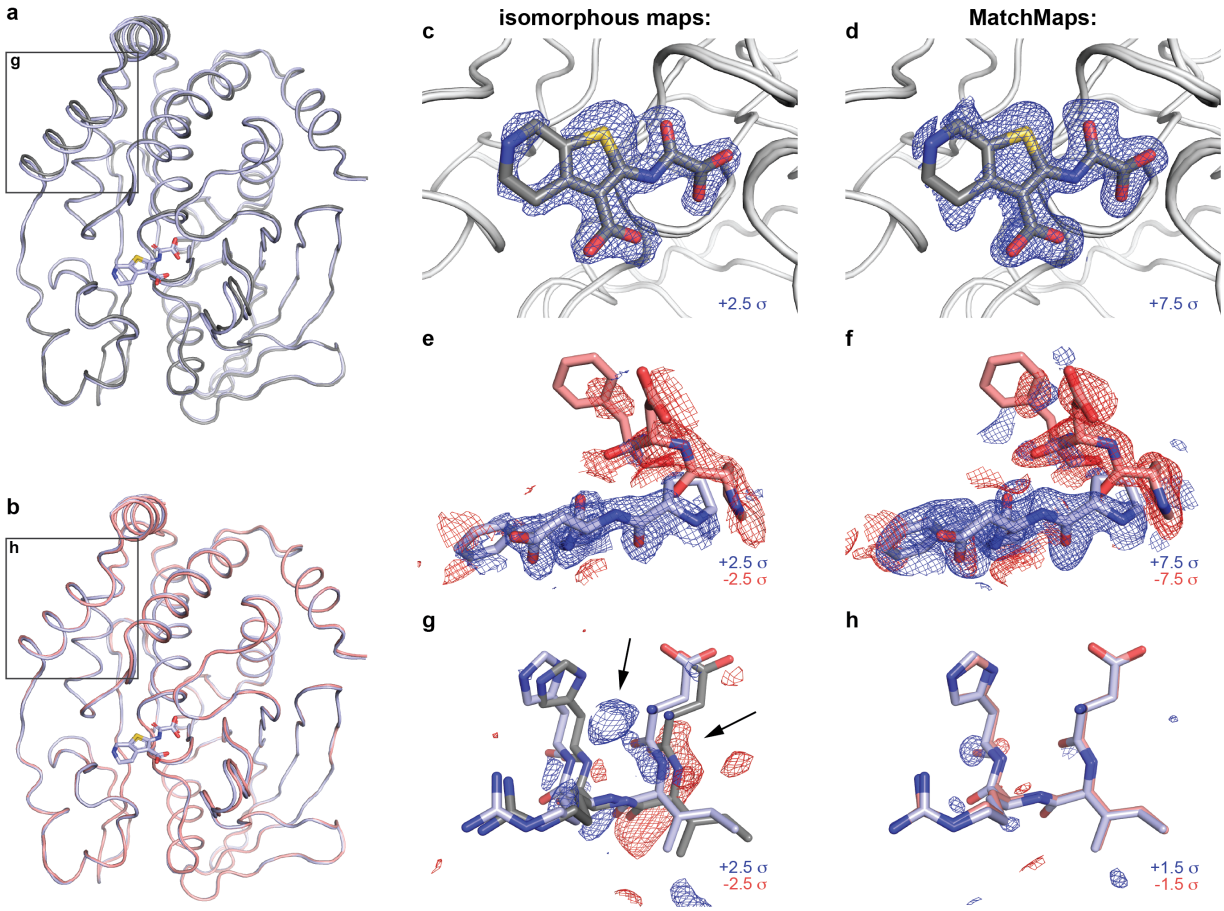

Figure 3: MatchMaps removes artifact even in isomorphous case Comparison of apo (7RIN) and TCS401-bound (7MM1) structures of protein tyrosine phosphatase 1B. (a) The apo (gray) and bound (blue) structural models overlay well, but differ by a slight rotation. The difference in the models is especially apparent in the boxed region (see g). (b) Aligning the apo model (red) to the bound model (blue) reveals that the structures overlay even better than the original coordinates suggest (see h). (c-d) Both the isomorphous map (c) and MatchMaps map (d) are able to clearly show the bound ligand. The TCS401 ligand (gray sticks) is shown for clarity but was not included in the computation of either map. (e-f) Closeup of residues 180-182. Similarly to (c-d), the change in loop equilibrium between open (red mesh, red sticks) and closed (blue mesh, blue sticks) is apparent in both maps. (g-h) Residues 22-25 are shown. Though these structures meet the requirements for isomorphism, the refined protein models still differ by a slight rotation. (g) The isomorphous difference map recognizes the artifactual difference between 7MM1 (blue) and 7RIN (gray) model locations, which manifests as strong difference signal. This artifact is comparable in magnitude to the "true" signal in panels (c) and (e). (h) MatchMaps internally aligns the models before subtraction and therefore avoids this artifact. The bound model after alignment to the apo model is shown in red. At +/- 2.5 σ, there is no signal in the MatchMaps map for this region.

6.4 Figure 4

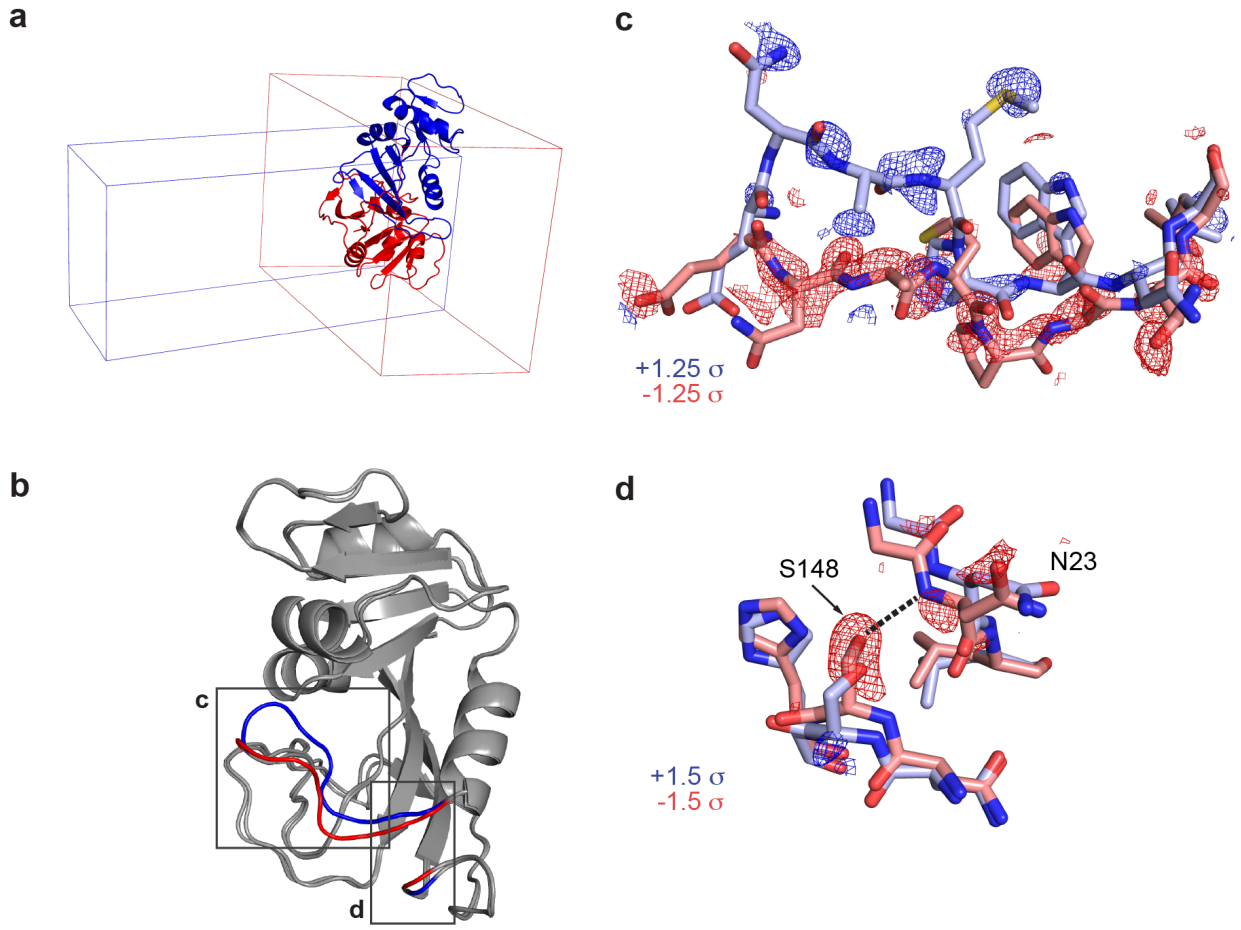

Figure 4: MatchMaps for difference maps across difference spacegroups A variant of MatchMaps (implemented in the command line as `matchmaps.mr`) can be used to compute difference maps between two crystallographic datasets in entirely different spacegroups. (a) Overlay of structural models of DHFR in spacegroup $P2_12_12_1$ (PDB ID 1RX1, blue cartoon) and spacegroup $C2$ (PDB ID 1RA1, red cartoon) along with the respective unit cells for each. (b) Alignment of structures in $P2_12_12_1$ and $C2$ shows global agreement of structures. Structural differences are localized to the active site (boxed regions, $P2_12_12_1$ structure in blue, $C2$ structure in red) and are known to result from differences in crystal packing. (c) Closeup on residues 17-24. MatchMaps positive (red) and negative (blue) difference density clearly correspond to the refined structural coordinates for the $P2_12_12_1$ (blue) and $C2$ (red) models. Remarkably, the positive difference density is strong and clearly corresponds to the $C2$ structure, despite the $C2$ structure never being used in the creation of the map. (d) Closeup on the hydrogen bond between residues Ser148 and Asn23, which is only present in the $C2$ crystal form (red sticks).

6.5 Figure 5

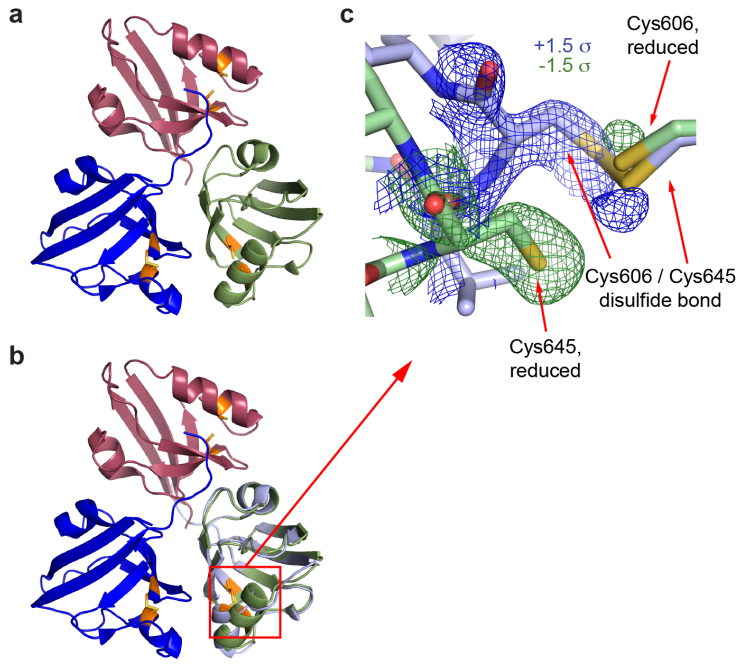

Figure 5: MatchMaps for internal difference maps across non-crystallographic symmetry operations A variant of MatchMaps (implemented in the command line as `matchmaps.ncs`) can be used to compute internal difference maps across an NCS operation. (a) Overview of the three PDZ domains related by non-crystallographic symmetry. Chain A is shown in red, chain B in green, and chain C in blue. Residues Cys606 and Cys645, which can form a disulfide bond, are shown in orange. Coloring matches figure 2C from ref [17]. (b) Same as (a), with the addition of chain C aligned and superimposed onto chain B, shown in light blue. (c) Zoom on the disulfide bond formation. Chain C (light blue) contains a disulfide bond between Cys606 and Cys645, whereas chain B (green) does not. The difference density corresponding to each chain is clearly visualized by MatchMaps.