Conformational Flexibility of the C-terminal Region Influences Distal Active Site Residues Across the Tautomerase Superfamily

Christopher Argueta , Andrew Parkins , Georgios Pantouris *

Posted Date: 4 November 2024

doi: 10.20944/preprints202411.0131.v1

Keywords: Tautomerase superfamily (TSF); macrophage migration inhibitory factor (MIF); D-dopachrome tautomerase (D-DT); 5-carboxymethyl-2-hydroxymuconate isomerase (CHMI); cis-3-chloroacrylic acid dehalogenase (cis-CaaD); 4-oxalocrotonate tautomerase (4-OT); malonate semialdehyde decarboxylase (MSAD); protein dynamics; molecular dynamics (MD) simulations

 doi:10.20944/preprints202411.0131.v1

Article

Conformational Flexibility of the C-Terminal Region Influences Distal Active Site Residues Across the Tautomerase Superfamily

Christopher Argueta †, Andrew Parkins † and Georgios Pantouris *

Department of Chemistry, University of the Pacific, Stockton, CA 95211, USA
* Correspondence: gpantouris@pacific.edu
† These authors contributed equally to this work.

Abstract: Consisting of more than 11,000 members distributed over five families, the tautomerase superfamily (TSF) is a large collection of proteins with diverse biological functions. While a lot of attention has been given to individual TSF enzymes, a majority remain structurally and functionally uncharacterized. Given its large size, studying a representative member of each family offers a viable approach for extracting mechanistic insights applicable to the entire superfamily. In this study, cis-3-chloroacrylic acid dehalogenase (cis-CaaD), 5-carboxymethyl-2-hydroxymuconate isomerase (CHMI), malonate semialdehyde decarboxylase (MSAD), and 4-oxalocrotonate tautomerase (4-OT) were referenced against the well-studied macrophage migration inhibitory factor (MIF) and D-dopachrome tautomerase (D-DT) using triplicate 1 μs molecular dynamics (MD) simulations for a total of 18 μs. Through root mean square fluctuation (RMSF) measurements and correlation analyses comparisons to previous crystallographic structures, we reveal key mechanistic insights that promote understanding of the catalytic activities in TSF. Collectively, our findings from these functionally diverse TSF proteins provide key information on allosteric coupling, long-range intra- and intersubunit communications as well as structure-activity relationships that enable new studies in the superfamily.

Keywords: tautomerase superfamily (TSF); macrophage migration inhibitory factor (MIF); D-dopachrome tautomerase (D-DT); 5-carboxymethyl-2-hydroxymuconate isomerase (CHMI); cis-3-chloroacrylic acid dehalogenase (cis-CaaD); 4-oxalocrotonate tautomerase (4-OT); malonate semialdehyde decarboxylase (MSAD); protein dynamics; molecular dynamics (MD) simulations

1. Introduction

The TSF is a ubiquitous group of proteins found across all domains of life, including Archaea, Bacteria, and Eukarya [1]. This superfamily is composed of five families, each represented by a founding protein: MIF, cis-CaaD, CHMI, MSAD, and 4-OT. Members of TSF are involved in a variety of biological processes and possess diverse catalytic activities across a range of substrates [1]. Despite their functional versatility, TSF proteins are characterized by a shared β-α-β structural motif and a conserved N-terminal proline (P1) that serves as either a catalytic base or acid [2,3]. A striking exception to this rule is ~3% of this family's proteins (346 out 11,395) lack P1, yet they are clearly part of this superfamily [1]. Oligomeric organization of TSF proteins leads to the formation of functional dimers [3], trimers [4,5], and hexamers [6] with a solvent channel at the center of the assembly. Studies have shown that ordered water molecules found within the channel play a functional role in influencing the catalytic activity of the protein carrier [3,7].

Regardless of the knowledge we have gained thus far, the majority of TSF enzymes remain structurally and functionally uncharacterized. Only select members of this superfamily, such as MIF and D-DT, have attracted a lot of attention due to their pleiotropic functionalities in human

 doi:10.20944/preprints202411.0131.v1

pathophysiology. For their structural [8–12] and functional [12–18] characterization, small molecule modulators discovered through the shared tautomerase activity [19,20] were used as experimental tools. Other TSF enzymes that have been characterized, but are not found in humans, are the founding members of the remaining four families: 4-OT, CHMI, cis-CaaD, and MSAD.

4-OT is a small protein consisting of just 62 amino acids [21] that catalyzes the isomerization of unsaturated alpha-keto acids [22]. Its homohexameric biological assembly, drastically differs from the typical trimeric structures of CHMI [23], cis-CaaD [24], and MSAD [25]. In 4-OT, P1 facilitates catalysis by abstracting a proton from the substrate (e.g., 4-oxalocrotonate) thus serving as a catalytic base [3]. Protein crystallography, nuclear magnetic resonance (NMR), and biochemical assays exposed R11, R39, and F50 as additional key active site residues [3].

With two fused β-α-β units, CHMI is approximately twice as long as monomeric 4-OT. Its commonly known substrate is 5-(carboxymethyl)-2-hydroxymuconate (CHM) [23] and the catalytic mechanism involves P1 acting as a general base [2]. Similar to 4-OT, the active site environment of CHMI includes two arginine residues (R40 and R71) that serve a key role in catalysis [2]. In *Escherichia coli C*, CHMI is expressed along with other enzymes as part of the homoprotocatechuate (hpc) pathway to generate intermediates of the citric acid cycle enabling the production of carbon and energy from aromatic amino acids catabolism [26].

The TSF enzyme cis-CaaD is known to catalyzes the hydrolytic dehalogenation of cis-3-chloroacrylic acid by converting it into malonate semialdehyde. The conserved P1 exhibits a pK_a of ~9.3, allowing it to function as a general acid [27], while R70 and R73 are involved in substrate binding and stabilization [24]. H28 and Y103 were identified in crystallographic studies as potential key players in the enzyme's activity by contributing to substrate specificity as well as catalytic mechanism.

MSAD is involved in decarboxylation reactions such as the conversion of malonate semialdehyde into acetaldehyde and carbon dioxide. Similar to cis-CaaD, the catalytic residue P1 of MSAD functions as a general acid (pK_a of ~9.2) [28,29], while residues D37, R73, and R75 aide in the active site catalysis [25]. MSAD exhibits promiscuous hydratase activity, in addition to its well-known decarboxylase one. This new functionality was confirmed through inhibition studies using 3-bromo- and 3-chloropropiolate, the first identified MSAD inhibitors [30].

Because the dynamic profiles of MIF and D-DT were successfully mapped with 1 μs MD simulations [10,31] that yielded not only structural but also functional insights to these pleiotropic proteins, we applied a similar approach on CHMI, cis-CaaD, MSAD, and 4-OT. While the dynamic profiles of these four proteins is vastly unknown, our goal was to extract mechanistic insights that would be applicable not only to these proteins but also to the entire superfamily. Our findings on the four proteins highlight the key role of long-range intra- and intersubunit communication that corroborate the dynamic correlation of the C-terminal region with active site residues. Such findings offer a deeper understanding of the catalytic mechanisms and functional activities of TSF, preparing the ground for additional studies of uncharacterized superfamily members.

2. Results

2.1. Crystallographic Analysis of TSF Representative Members Reveals Noticeable Differences in the C-Terminal Region

Previously published studies on the TSF human members, MIF and D-DT, showed that the C-terminal region has a multi-tiered role associated with the structure and function of these proteins [8,10,11,13,31–33]. Having these two proteins as our benchmark, we probed for broader insights into this superfamily. While interrogation of over 11,000 proteins is unrealistic, we selected representative members based on a previously published sequence similarity network (SSN) analysis that partitioned the TSF proteins into five families; MIF, cis-CaaD, MSAD, CHMI, and 4-OT [1].

A multiple sequence alignment was performed utilizing the amino acid sequences of human MIF, human D-DT, *Coryneform bacterium* cis-CaaD, *Pseudomonas aeruginosa* CHMI, *Pseudomonas pavonaceae* MSAD, and *Pseudomonas sp.* (strain CF600) 4-OT (**Figure S1**). The highest sequence identity (SeqID) of 34.2% between any protein pair was obtained between MIF and D-DT (**Table S1**). The

 doi:10.20944/preprints202411.0131.v1

remaining protein pairs yielded values under 30%, with MSAD exhibiting similar SeqIDs with 4-OT (25.8%), cis-CaaD (25.0%), and CHMI (24.7%). Despite the low amino acid sequence identity, the six proteins demonstrate an overall satisfactory structural homology at the quaternary level, yet with diverse C-terminal segments (**Figure 1A**). Specifically, key differences were noted in the length (**Figure S1**) and secondary structure organization of the C-terminal tail (**Figure 1A**) as well as the position of the C-terminus in relation to the active site pocket (**Figure 1B**). For MIF, D-DT, MSAD, and 4-OT, the C-terminus is proximal to the active site opening, whereas in the cases of cis-CaaD and CHMI distal from it (**Figure 1B**). Interestingly, the active site opening of cis-CaaD is completely blocked by the β8/α3 loop, which is part of the C-terminal tail.

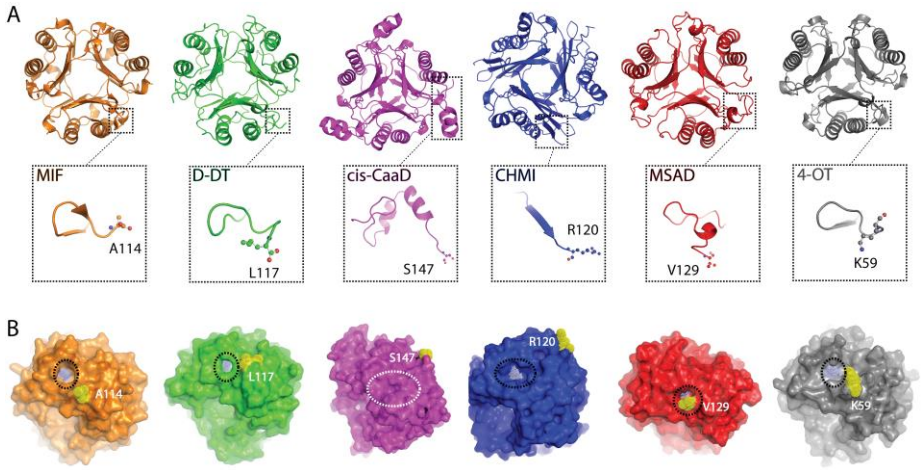

Figure 1. Structural homology between the TSF members. **(A)** The biological assemblies of MIF (orange - PDB:3DJH), D-DT (green - PDB:1DPT), cis-CaaD (magenta - PDB:2FLZ), CHMI (blue - PDB:3E6Q), MSAD (red - PDB:2AAG), and 4-OT (grey - PDB:1OTF) demonstrate an overall satisfactory structural homology, with some noticeable differences in the C-terminal region (dashed boxes). The C-terminus residue for each protein is shown in ball and sticks representation. **(B)** Surface analysis focused on the area around the active site opening (black dashed circles) shows the distance between P1 and the C-terminus residue, for each protein. P1 and the C-terminus residue are illustrated with light blue and yellow dots, respectively. In the case of cis-CaaD, the active site opening (white dashed circle) is not visible as it is blocked by the β8/α3 loop of C-terminus.

2.2. RMSF Analysis Across the Target Proteins Demonstrates Diverse Dynamic Profiles

Upon identifying these structural differences in the C-terminal region, we performed MD simulations and analyzed the RMSF profiles of the six proteins (**Figures 2 and S2**). 1 μs trajectories were considered suitable for investigating the C-terminal motions [34], while each calculation was repeated in triplicate. Globally, the six proteins demonstrated similar fluctuations with an average RMSF value of 0.71 ± 0.2 Å, 0.69 ± 0.3 Å, 0.88 ± 0.7 Å, 0.67 ± 0.5 Å, 0.79 ± 0.5 Å, and 0.87 ± 0.7 Å for MIF, D-DT, cis-CaaD, CHMI, MSAD, and 4-OT, respectively (**Table S2, Figure S2**). To confirm the accuracy of our approach, we compared our findings with previously published RMSF values obtained from 1 μs trajectories. In the absence of any data for the bacterial proteins, we considered only data derived from the two human proteins, MIF and D-DT. The previously published RMSF values of 0.90 Å [31] and 0.70 Å [10], for MIF and D-DT respectively, are in agreement with the findings of this study (0.71 ± 0.2 Å (MIF) & 0.69 ± 0.3 Å (D-DT)). Having the accuracy of our calculation being confirmed, we performed an in-depth analysis of the RMSF results focusing on the poorly studied bacterial proteins. For clarity, the secondary structure features of cis-CaaD, CHMI, MSAD monomers, and 4-OT dimer are provided (**Figure S3**).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 November 2024 doi:10.20944/preprints202411.0131.v1

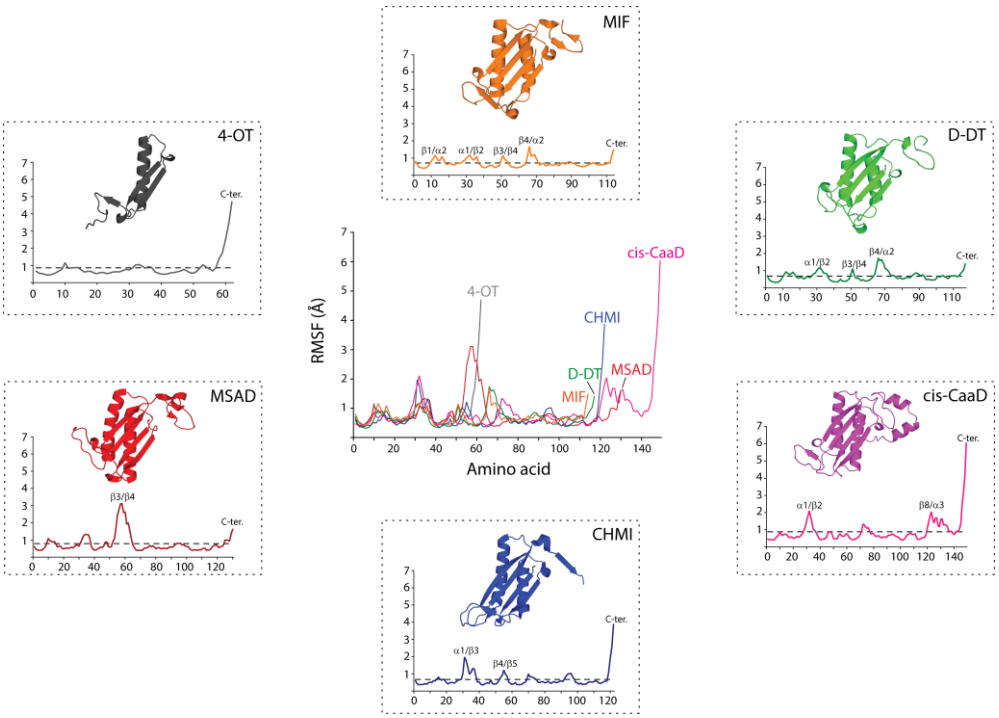

Figure 2. Average RMSF profiles of TSF proteins. The average fluctuations of protein residues (x-axis) were derived from three independent runs and monitored over the course of 1 μs MD simulations. The RMSF magnitude (y-axis) is expressed in angstroms (Å). Regions with fluctuation values greater than 1σ of the mean RMSF value (dashed lines) are marked using the secondary structure features of each protein. The MIF (orange), D-DT (green), cis-CaaD (magenta), CHMI (blue), MSAD (red), and 4-OT (grey) monomers are also shown. For comparison, the overlay of all six profiles is shown at the center. .

The high similarity between the RMSF profiles of MIF and D-DT is apparent even upon a brief examination. However, each of the four bacterial proteins exhibits characteristic RMSF patterns that suggest unique intra- and intersubunit communication pathways (**Figure 2**). Notably, for the two human proteins the highest RMSF value of any region does not surpass 2 Å. Meanwhile, cis-CaaD, CHMI, MSAD, and 4-OT enclose one highly flexible region with an RMSF value far exceeding 2 Å. For CHMI, cis-CaaD, and 4-OT, this region is located in the C-terminal, whereas in MSAD, the β3/β4 loop is highly flexible. The reduced flexibility noted in the C-terminal residues of MSAD more closely resembles what is observed in MIF and D-DT (**Figure 2**). From a structural point of view, the high flexibility of cis-CaaD in the C-terminal tail, is of a great functional interest as it influences the mobility of β8/α3 loop, which in turn blocks the opening to the active site (**Figure 1B**).

Regions with fluctuation values greater than 1σ of the mean RMSF value were marked using the secondary structure features of each protein (**Figures 2 and S3**). Our findings show that the human proteins have more regions with statistically significant RMSF values in comparison to the bacterial ones. This finding is not attributed to enhanced flexibility of MIF and D-DT, but rather, it is explained by the presence of highly flexible regions within the four bacterial proteins. For this reason, we overlayed the six profiles and examined their fluctuation features region-by-region (**Figure 2**). The varying secondary structural features and length of the six proteins account for the differences seen in the overlayed RMSF illustration. Despite this, a similar fluctuation pattern with values above the average was noted in the α1/β2 loop (corresponds to the α1/β3 loop of CHMI). This loop is located in the active site cavity of all the proteins and is specifically adjacent to the catalytic residue P1(**Figures 2 and S4**). With reference to the better studied human proteins, MIF and D-DT, this loop includes residues that are important for catalysis and ligand binding [11,12,19,35]. Excluding MSAD, fluctuation above the mean value was also noted for residues found in the β4/α2 loop of MIF/D-DT and the corresponding β5/α2 and β1/α1 loops of cis-CaaD/CHMI and 4-OT, respectively. This loop, which is also located next to the catalytic residue P1, harbors an active site residue: I64 in MIF and D-

 doi:10.20944/preprints202411.0131.v1

DT, R11 in 4-OT, R71 in CHMI, R73 in cis-CaaD, and R75 in MSAD. In all proteins, fluctuation of the C-terminal region was found to be significant, surpassing 1σ of the mean RMSF value.

2.3. Correlation Plots Expose Communication Pathways with Mechanistic Interest

Correlation analyses of the Cα atoms were performed for the six proteins included in this study (**Figure 3A–F**). Similar to the RMSF analysis, MIF and D-DT (**Figure 3A,B**) were only used as controls to validate our data against previously published findings [10,31]. Once the reproducibility of our approach was confirmed, our attention shifted to the four bacterial proteins whose communication pathways are unknown (**Figure 3C–F**).

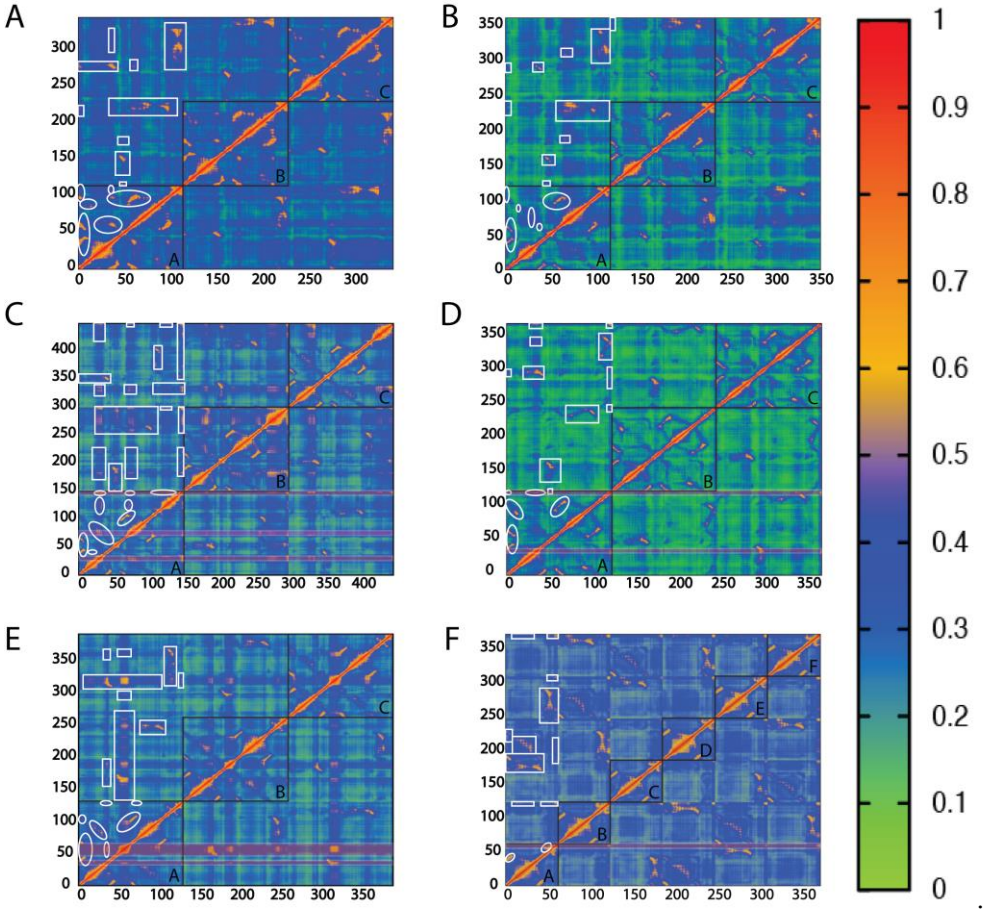

Figure 3. Correlation analysis of the TSF members. Generalized correlated Cα motions for **(A)** MIF, **(B)** D-DT, **(C)** cis-CaaD, **(D)** CHMI, **(E)** MSAD, and **(F)** 4-OT are provided side-by-side for comparison. Each correlation plot represents the average analysis derived from three 1 μs trajectories. Correlation coefficients range from 0 to 1, as shown in the heat map. Oval circles and boxes highlight intrasubunit and intersubunit correlated regions, respectively. The black boxes A-C shown in the correlation plots of MIF, D-DT, cis-CaaD, CHMI, and MSAD demonstrate the three subunits of these proteins. The corresponding six subunits of 4-OT are shown by the black boxes A-F. The Y- and X- axes show the residue numbers of each protein. Correlations possessing mechanistic interest are marked with horizontal rectangles filled with transparent red.

In all four proteins, β1 strand is strongly correlated with the two adjacent strands of the monomeric β sheet, dimeric in the case of 4-OT. This finding is consistent with previously published observations of the two human proteins [10,31] and it reflects the fundamental role of β sheet in correlated motions [36,37]. With cis-CaaD being a notable exception, we found another shared correlation between the first residue of β1 and the strand defining the solvent channel opening; β5 for MSAD, β7 for CHMI, and β2 of the adjacent monomer for 4-OT (**Figure S3**). This correlation is of

 doi:10.20944/preprints202411.0131.v1

a great importance for the two human proteins as it was shown to modulate their catalytic activities via allosteric coupling [7,38].

With the objective to discover mechanistic insights into cis-CaaD, CHMI, MSAD, and 4-OT, we probed characteristic communications for each protein. In cis-CaaD, T34 stands out as a key residue situated on the flexible α1/β2 loop (**Figure 4A**). Our analysis showed that this residue is strongly correlated with multiple domains across the biological assembly of cis-CaaD (**Figure 3C**). Within the same subunit, T34 communicates with the β8/α3 loop, the α3 helix, and the C-terminal tail, which are all found proximal to this residue (**Figure 4A**). Noteworthy, β8/α3 is the loop blocking accessibility to the active site pocket (**Figure 1B**). T34 of monomer A also forms intersubunit communications with the α1/β2 loops of monomers B and C, including their T34 residues which are found ~34 Å apart (**Figure 4A**). In addition, T34 of monomer A is correlated with the β8/α3 loop and the α3 helix of monomer B as well as the three C-termini of cis-CaaD. These communications take place via long-range intersubunit crosstalk with the participation of β5/α2 and β6/β7 loops as well as segments of α2 helix and β6 strand (**Figure S3**). With exception of β5/α2 loop, the remaining regions are located in the subunit-subunit interface and enable communications with the adjacent monomer. These findings led to the conclusion that catalysis in cis-CaaD is a highly coordinated process, where T34 plays a key role via major conformational changes of the C-terminal region β8/α3 loop, α3 helix, and the C-terminal tail.

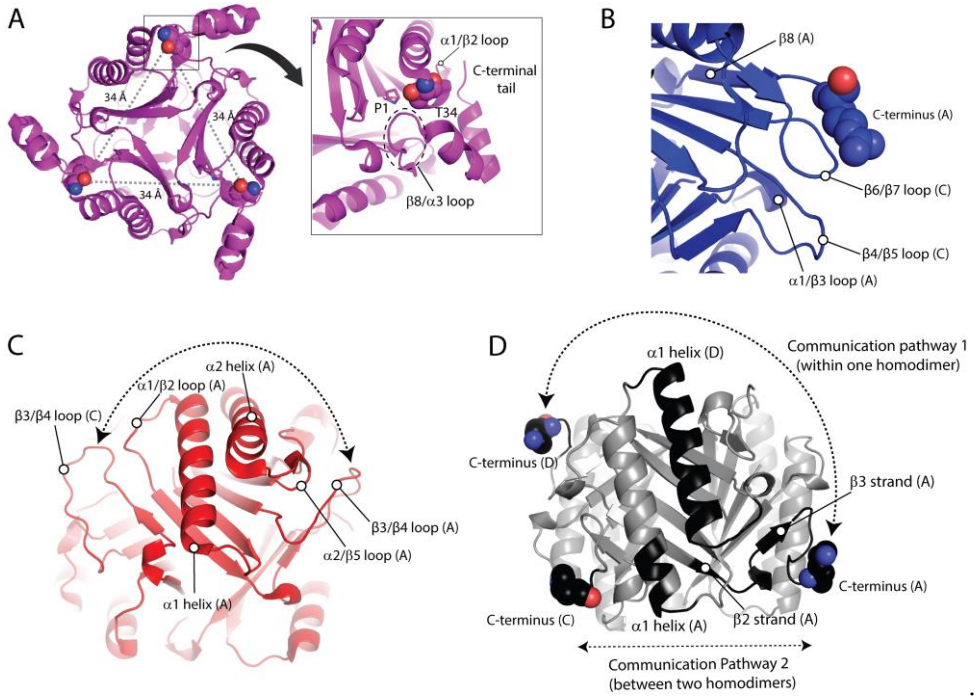

Figure 4. Correlations with mechanistic interest in cis-CaaD, CHMI, MSAD, and 4-OT. **(A)** T34 demonstrates high correlations with itself in all the subunits of cis-CaaD, despite the large distance of 34 Å between them. Intrasubunit correlations of the α1/β2 loop, on which T34 is situated, involve interactions with the C-terminal tail and the β8/α3 loop (dashed circles). P1 and T34 are shown as sticks and space filling representation, respectively. **(B)** The C-terminus (in space filling representation) of CHMI communicates with the α1/β3 loop through intersubunit correlations that involve the β4/β5 and β6/β7 loops of the adjacent monomer. **(C)** The highly flexible β3/β4 loop of MSAD from subunit A is highly correlated with the β3/β4 loop of the adjacent subunit C. These correlated motions are mediated through intermediate communications that involve the α1/β2 loop, α1 helix, α2 helix, and α2/β5 loop of subunit A. **(D)** Within the same pseudo-monomer of 4-OT, the C-terminus (in space filling representation) of monomer A communicates with the C-terminus of monomer D through the α1/β2 loop, α1 and β1/α1 loops. Communication between the C-terminus of monomer A and monomer C, which is the subunit of the adjacent homodimer, occur in the order described via the β3 strand, β2/β3 loop, β2 strand, and α1 helix.

 doi:10.20944/preprints202411.0131.v1

For CHMI, an interesting observation is that all the C-termini are correlated with one another, despite being situated ~47 Å apart (**Figure 3D**). These correlations are enabled through the participation of multiple regions, including the α1/β3 loop that is strongly linked with the C-terminus (**Figure 4B**). Interestingly, the β8 strand that is found just before the C-terminus packs against the adjacent subunit. When examining subunit A, we observe that the β8 strand of this subunit interacts with the β7 strand of subunit C, bringing the C-terminus of subunit A proximal to the β6/β7 loop of subunit C (**Figure 4B**). The β6/β7 loop of subunit C is highly correlated with the β4/β5 loop of the same subunit, as well as the α1/β3 loop of subunit A. These correlated motions initiate from the highly flexible C-terminal region of subunit A, propagate through the β6/β7 loop of subunit C, and continue into the β4/β5 loop of the same subunit, ultimately reaching the α1/β3 loop of subunit A (**Figure 4B**). The α1/β3 loop of CHMI is characterized as the second highest RMSF region, surpassed only by the C-terminus. The correlated motions travel up towards the ends of β5 and β7 and ultimately reach the C-terminal solvent channel opening. Residues located at the end of β7 and within the β7/β8 loop of subunit A are correlated with the same residues of subunit C due to proximity, effectively bridging the C-terminus of subunit A and the C-terminus of subunit C.

From the RMSF analysis (**Figure 2**), it was clearly shown that MSAD exhibits distinct dynamic characteristics from cis-CaaD, CHMI, and 4-OT. A unique aspect of MSAD is that the region with the highest RMSF value is not the C-terminus, as is typically observed in the other three bacterial proteins, but rather the loop between β3 and β4 strands (**Figure 4C**). The correlation analysis also highlights that the β3/β4 loop is a region of high mechanistic value forming multiple strong intra- and intersubunit communications across the biological assembly of MSAD (**Figure 3E**). The α1/β2 loop also exhibits a significant number of correlations, second only to the β3/β4 loop. The communication network formed through the β3/β4 loop as well as the α1/β2 loop facilitate the correlated motions within MSAD. Further analysis showed that the β3/β4 loop from different subunits form strong correlations with each other through intermediate communications that involve α1/β2 loop, α1, α2 helices, and α2/β5 loop (**Figure 4C**).

In the case of 4-OT, the communication pathways found in the biological assembly are more complicated in comparison to the other three proteins due to the homohexameric structure. For comparative analysis with the other TSF proteins, a dimer of 4-OT effectively acts as a pseudo-monomer, thereby incorporating two C-terminal regions within one monomer (**Figure S3**). Monomers A&D, B&E, and C&F form three homodimers, each of which corresponds to a monomer of cis-CaaD, CHMI, and MSAD. Analysis of the correlation data showed that the six C-termini communicated to each other despite the long distances (the shortest determined at ~26 Å). The flexibility observed in the C-terminal residues of the protein (**Figure 2**) plays a key role in the intra- and intersubunit communication pathways enabling cross-talking between the C-termini. Our findings demonstrate that the communication pathway between the two C-termini of a given homodimer (pseudo-monomer) differs from the corresponding pathways formed across two homodimers (**Figure 4D**). Using the homodimer A/D as an example and in the order described, the C-terminus of monomer A communicates with the C-terminus of monomer D through the α1/β2 loop, α1 and β1/α1 loops (**Figure 4D**). Communication between the C-terminus of monomer A and monomer C, which is a subunit of the adjacent homodimer, occur through the β3 strand, β2/β3 loop, β2 strand, and α1 helix (**Figure 4D**).

2.4. Correlation Analyses Reveals Communications Between the C-Terminal Region and Active Site Residues

As previously described communications between the C-terminal region and active site residues modulate the enzymatic activities of MIF and D-DT. Having this in mind, we utilized our correlation analyses to detect distinct communication pathways in cis-CaaD, CHMI, MSAD, and 4-OT. Among the six proteins of this study, cis-CaaD is the largest, featuring a highly flexible C-terminal region (**Figure 2**) and exhibiting high intramolecular correlation with the α1 helix (**Figure 3C**). The active site residue H28 lies on this helix, but the correlated motions continue through the α1 helix towards R70 and R73 (**Figure 5A**), which are also active site residues.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 November 2024 doi:10.20944/preprints202411.0131.v1

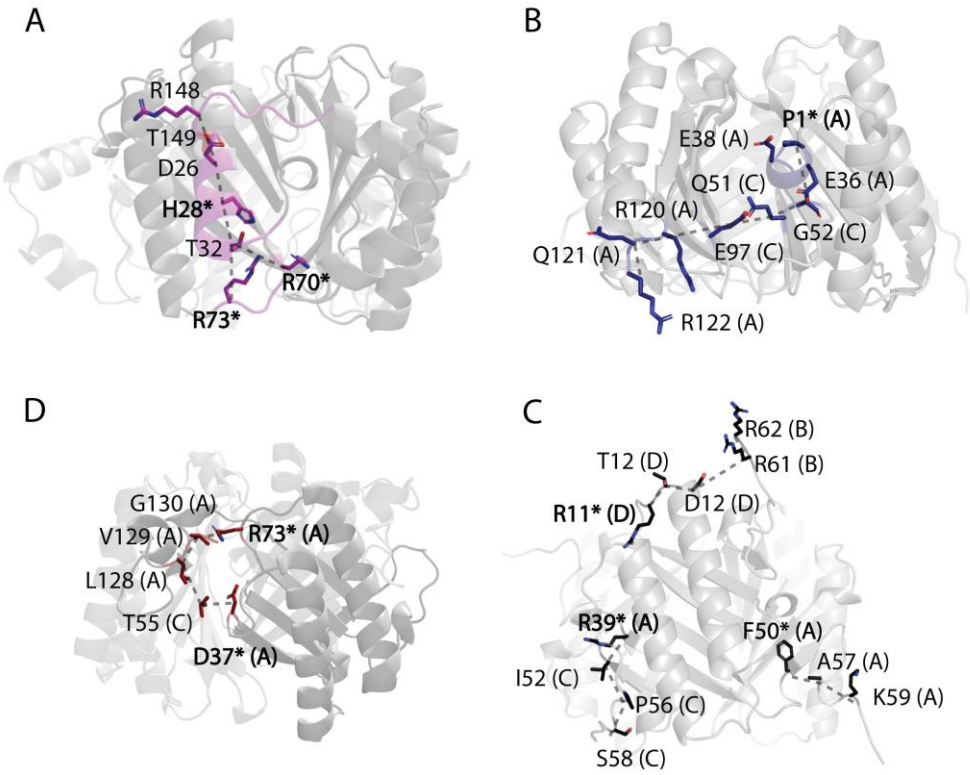

Figure 5. Communication pathways between the C-terminal region and active site residues. Correlation analyses of **(A)** cis-CaaD, **(B)** CHMI, **(C)** MSAD, and **(D)** 4-OT demonstrate that the C-terminal region and active site residues for each protein form distinct communication pathways (dashed grey lines). The residues involved in each pathway are shown as sticks. Bold residues with an asterisk indicate the active site residues. When multiple subunits (subunit A, B, etc.) are involved in the communication pathway, the letter of each subunit is indicated in parenthesis.

For CHMI, our findings demonstrate cross-talk between residues situated at the interface of two monomers that bridge the C-terminal region with the catalytic residue P1 (**Figure 5B**). Specifically, the C-terminal region of subunit A is correlated to P1 of the same subunit via the β6/β7 loop derived from subunit C. From the β6/β7 loop, the correlated motions continue towards the β4/β5 loop of subunit C, then onto α1/β3 loop of subunit A to eventually reach P1 (**Figure 5B**).

In MSAD, the C-terminal region exhibits greater restraint; however, a sizable and profoundly dynamic region of the β3/β4 loop within one subunit is positioned between the C-terminal segment and a loop adjacent to the active site residues of another monomer. This β3/β4 loop interfaces with a distinct dynamic α1/β2 loop of another monomer at the interphase between monomers. The C-terminal region of subunit A is highly correlated to the active site residue D37 via the highly fluctuating β3/β4 loop region of subunit C that lies between the two areas. The active site residue R73, located just after β4, is also highly correlated to the C-terminal region because of its proximity (**Figure 5C**).

4-OT stands out due to its deviation from being a homotrimer. For comparative analysis with the other TSF proteins, a dimer of 4-OT is regarded as a pseudo-monomer and thus, contains two C-terminal segments. Focusing on the catalytically relevant residues located on monomer A, we observe that the C-terminal region of monomer A at K59 demonstrates intrasubunit correlations to the active site residue F50, which is located at the beginning of β3 strand (**Figure 5D**). Besides the intrasubunit correlations, the C-terminal region of a given monomer was found to form intersubunit correlations with active site residues from adjacent monomers. For example, S58 derived from the C-terminal region of monomer C communicates with the active site residue R39 of monomer A, via P56 and I52 of monomer C (**Figure 5D**).

 doi:10.20944/preprints202411.0131.v1

These correlation analyses have revealed robust communication between the C-terminal residues and crucial active site residues in every TSF protein included in this study. Through these correlated motions the C-terminal region dynamically influences active site residues and strongly suggests that their impact in catalysis should be further explored with kinetic assays, utilizing C-terminal variants.

3. Discussion and Conclusions

The correlation between protein dynamics and functionality in the TSF is largely unknown with the exception of the two human proteins MIF and D-DT. Previous studies on these proteins highlighted the critical role of correlated motions in the regulation of their functionality, despite the low conformational flexibility of their biological assemblies [7,10,31,39]. In an effort to enhance understanding of structure-activity relationships in the TSF, we analyzed triplicate 1 μs MD simulations of the four bacterial proteins cis-CaaD, CHMI, MSAD, and 4-OT in comparison to MD simulations of the well-studied MIF and D-DT as points of reference.

Despite being members of the same superfamily, our findings demonstrate striking distinctions between the human and bacterial proteins. First, the four bacterial proteins have one region of high flexibility that is absent from MIF and D-DT. In cis-CaaD, CHMI, and 4-OT, this region is the C-terminus, while in MSAD it is the β3/β4 loop. Another shared feature of the bacterial proteins that is not present in the two human proteins is the formation of long-range intersubunit communication expanding across their biological assembly. A representative example includes the three T34 residues of cis-CaaD, which are located ~34 Å apart in the biological assembly. Such regions, which are clearly marked on the correlation plots (3x for cis-CaaD, 2x for CHMI, 2x for MSAD, and 1x for 4-OT), possess a mechanistic value aiding efforts to understand the functionality of these proteins. In cis-CaaD, for example, it suggests that conformational changes of T34 accompanied by movements of the C-terminal region are required for substrate binding.

A consistent long-range intersubunit correlation between the C-termini of cis-CaaD, CHMI, MSAD, and 4-OT led to further investigation of its potential mechanistic value. Our analysis shows that active site residues and the C-terminal region are highly correlated implying a functional role of the latter in catalysis. While our findings provide insights into the correlation between protein dynamics and functionality, the study is limited by the constraints of computational analyses. Future studies including experimental validation, would further clarify the role of C-terminal flexibility in the functionality of these proteins. Notably, experimental analyses of MIF [33] and D-DT [11] showed that the C-terminal region is linked with the catalytic activities of these proteins.

In summary, this study reveals the dynamic maps of cis-CaaD, CHMI, MSAD, and 4-OT via of which novel findings with potential mechanistic value were extracted. Such findings promote understanding of TSF functionalities and provide the foundation for further studies focusing on structure-activity relationships of this intriguing superfamily.

4. Materials and Methods

4.1. Multiple Sequence Alignments

The FASTA sequences of MIF (PDB:3DJH), D-DT (PDB:1DPT), cis-CaaD (PDB:2FLZ), CHMI (PDB:3E6Q), MSAD (PDB:2AAG), and 4-OT (PDB:1OTF) were obtained from the corresponding deposited crystal structures. In the case of CHMI, the N-terminus tag was removed prior to the alignment. MSA alignment of the six proteins was generated using the standalone program ESPript_3.0 [40]. The core input files were pre-aligned in Clustal, FASTA, MultAlin, NPS@ or ProDom format. The query file for the EsPript run was obtained from Clustal Omega [41]. EsPript then calculates a similarity score for each residue in the aligned sequences. A red box and a white character designate strict identity. The red characters designate similarity in a group. The blue frame designates similarity across groups. The default global score of 0.7 was changed to 0.6. If the similarity score assigned to a column was greater than 0.6, then the corresponding residues in the column are marked as highly similar.

 doi:10.20944/preprints202411.0131.v1

4.2. MD Simulations

Generation of simulation data was carried out on Linux machines utilizing, where applicable, CUDA accelerated or parallelized versions of simulation, analysis, and modeling programs. Starting structures for the 1 μs simulations were obtained from the Protein Data Bank for the corresponding proteins: 3DJH MIF, 1DPT D-DT, 1OTF 4-OT, 2FLZ cis-CaaD, 2AAG MSAD, and 3E6Q CHMI. Where necessary, symmetry pairs of the deposited structure were generated to obtain the biological assembly using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC [42]. Crystallographic waters, ions, and small molecules were removed. Missing residues were added using UCSF Chimera [43]. Protein structure files (PSF) with hydrogens were generated using the Visual Molecular Dynamics (VMD 1.9.3) [44] plugins psfgen 1.6.4. and CHARMM36 topology [45]. The resulting PSF and PDB pair were then solvated with TIP3P waters, and the net-charge of the system was checked using the VMD solvate (1.7) and autoionize (1.4) plugins, respectively [44]. Waters were added based upon the dimensions of the protein of interest with a minimum of 7.5 Å of waters bordering the perimeter of the structure. Where necessary, the charge of the system was balanced with Na^+ or Cl^- ions. The simulation process was carried out in a similar manner as previously described [9] using NAMD 2.12 [46], with a 2 fs timestep. Throughout the entire simulation process, the non-bonded interaction distance cutoff was set to 12 Å. Periodic boundary conditions were also used and set to the size of the system, while Langevin dynamics were used to simulate the system's temperature. Minimization of the system was conducted utilizing a multistep process. First, the protein was fixed in the system and only the solvent was minimized. Next, minimization of the system was done with the backbone of the protein fixed. Finally, the entire system was minimized, before a gradual heating to 300K. The system was then equilibrated for 1ns at 300K before a production run of 1ms was performed with all outputs being generated every 10 ps. All minimization, equilibration, and production runs were completed in triplicate from the solvated and charge balanced structures.

4.3. Correlation Analysis

Generalized cross-correlation analysis was performed using the Cα over the 1 μs trajectory utilizing the GROMACS plug-in g_correlation [47]. The correlation values produced from this analysis ranged between 0 and 1 (where 0 means no correlation and 1 means absolute correlation) and relate the movements of different Cα with each other. A correlation value of 1 is expected when comparing a group against itself and high values nearing 1 are typical when comparing neighboring groups. High correlation values that appear between non-neighboring groups can provide further understanding of important intra and inter-subunit interactions and elucidate changes in dynamics and communication that occur when an inhibitor is bound.

4.4. RMSF Analysis

RMSF analysis of Cα was performed in GROMACS by using the coordinates of the first frame of the production MD run and the trajectory in GROMACS TRR format. RMSF analysis was performed for each triplicate and RMSF values were averaged.

Supplementary Materials: The following supporting information can be downloaded at: Preprints.org, Figure S1: Multiple sequence alignment of TSF representative members; Figure S2: Dynamic profiles of the six TSF proteins; Figure S3: The secondary structure features of the four bacterial proteins; Figure S4: Illustration of the flexible loop, found in the active site environment of the six TSF proteins; Table S1: Sequence identity between TSF members; Table S2: Average Root Mean Square fluctuation (RMSF) profiles of TSF proteins.

Author Contributions: Conceptualization, G.P.; software, A.P.; validation, C.A., A.P. and G.P.; formal analysis, C.A. and G.P.; investigation, C.A. and G.P.; resources, G.P.; data curation, G.P.; writing—original draft preparation, C.A.; writing—review and editing, C.A., A.P. and G.P.; supervision, G.P.; funding acquisition, G.P. All authors have read and agreed to the published version of the manuscript

Funding: This research received no external funding.

 doi:10.20944/preprints202411.0131.v1

Acknowledgments: This work was supported by funds from the Chemistry Department of the University of the Pacific and Scholarly/Artistic Activities Grant (G.P.). We would like to thank Prof. Jerry Tsai for careful reading of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Davidson, R.; Baas, B. J.; Akiva, E.; Holliday, G. L.; Polacco, B. J.; LeVieux, J. A.; Pullara, C. R.; Zhang, Y. J.; Whitman, C. P.; Babbitt, P. C., A global view of structure-function relationships in the tautomerase superfamily. *J Biol Chem* **2018,** 293, (7), 2342-2357.
2. Poelarends, G. J.; Veetil, V. P.; Whitman, C. P., The chemical versatility of the beta-alpha-beta fold: catalytic promiscuity and divergent evolution in the tautomerase superfamily. *Cell Mol Life Sci* **2008,** 65, (22), 3606-18.
3. Whitman, C. P., The 4-oxalocrotonate tautomerase family of enzymes: how nature makes new enzymes using a beta-alpha-beta structural motif. *Arch Biochem Biophys* **2002,** 402, (1), 1-13.
4. Sun, H. W.; Bernhagen, J.; Bucala, R.; Lolis, E., Crystal structure at 2.6-A resolution of human macrophage migration inhibitory factor. *Proc Natl Acad Sci U S A* **1996,** 93, (11), 5191-6.
5. Sugimoto, H.; Taniguchi, M.; Nakagawa, A.; Tanaka, I.; Suzuki, M.; Nishihira, J., Crystal structure of human D-dopachrome tautomerase, a homologue of macrophage migration inhibitory factor, at 1.54 A resolution. *Biochemistry* **1999,** 38, (11), 3268-79.
6. Taylor, A. B.; Czerwinski, R. M.; Johnson, W. H., Jr.; Whitman, C. P.; Hackert, M. L., Crystal structure of 4-oxalocrotonate tautomerase inactivated by 2-oxo-3-pentynoate at 2.4 A resolution: analysis and implications for the mechanism of inactivation and catalysis. *Biochemistry* **1998,** 37, (42), 14692-700.
7. Pantouris, G.; Khurana, L.; Ma, A.; Skeens, E.; Reiss, K.; Batista, V. S.; Lisi, G. P.; Lolis, E. J., Regulation of MIF Enzymatic Activity by an Allosteric Site at the Central Solvent Channel. *Cell Chem Biol* **2020,** 27, (6), 740-750 e5.
8. Pantouris, G.; Bucala, R.; Lolis, E. J., Structural Plasticity in the C-Terminal Region of Macrophage Migration Inhibitory Factor-2 Is Associated with an Induced Fit Mechanism for a Selective Inhibitor. *Biochemistry* **2018,** 57, (26), 3599-3605.
9. Parkins, A.; Skeens, E.; McCallum, C. M.; Lisi, G. P.; Pantouris, G., The N-terminus of MIF regulates the dynamic profile of residues involved in CD74 activation. *Biophys J* **2021,** 120, (18), 3893-3900.
10. Parkins, A.; Chen, E.; Rangel, V. M.; Singh, M.; Xue, L.; Lisi, G. P.; Pantouris, G., Ligand-induced conformational changes enable intersubunit communications in D-dopachrome tautomerase. *Biophys J* **2023**.
11. Parkins, A.; Pilien, A. V. R.; Wolff, A. M.; Argueta, C.; Vargas, J.; Sadeghi, S.; Franz, A. H.; Thompson, M. C.; Pantouris, G., The C-terminal Region of D-DT Regulates Molecular Recognition for Protein-Ligand Complexes. *J Med Chem* **2024,** 67, (9), 7359-7372.
12. Parkins, A.; Das, P.; Prahaladan, V.; Rangel, V. M.; Xue, L.; Sankaran, B.; Bhandari, V.; Pantouris, G., 2,5-Pyridinedicarboxylic acid is a bioactive and highly selective inhibitor of D-dopachrome tautomerase. *Structure* **2023,** 31, (3), 355-367 e4.
13. Pantouris, G.; Syed, M. A.; Fan, C.; Rajasekaran, D.; Cho, T. Y.; Rosenberg, E. M., Jr.; Bucala, R.; Bhandari, V.; Lolis, E. J., An Analysis of MIF Structural Features that Control Functional Activation of CD74. *Chem Biol* **2015,** 22, (9), 1197-205.
14. Singh, A. K.; Pantouris, G.; Borosch, S.; Rojanasthien, S.; Cho, T. Y., Structural basis for decreased induction of class IB PI3-kinases expression by MIF inhibitors. *J Cell Mol Med* **2017,** 21, (1), 142-153.
15. Tilstam, P. V.; Pantouris, G.; Corman, M.; Andreoli, M.; Mahboubi, K.; Davis, G.; Du, X.; Leng, L.; Lolis, E.; Bucala, R., A selective small-molecule inhibitor of macrophage migration inhibitory factor-2 (MIF-2), a MIF cytokine superfamily member, inhibits MIF-2 biological activity. *J Biol Chem* **2019,** 294, (49), 18522-18531.
16. Parkins, A.; Sandin, S. I.; Knittel, J.; Franz, A. H.; Ren, J.; de Alba, E.; Pantouris, G., Underrepresented Impurities in 4-Hydroxyphenylpyruvate Affect the Catalytic Activity of Multiple Enzymes. *Anal Chem* **2023,** 95, (11), 4957-4965.
17. Bloom, J.; Pantouris, G.; He, M.; Aljabari, B.; Mishra, L.; Manjula, R.; Parkins, A.; Lolis, E. J.; Al-Abed, Y., Iguratimod, an allosteric inhibitor of macrophage migration inhibitory factor (MIF), prevents mortality and oxidative stress in a murine model of acetaminophen overdose. *Mol Med* **2024,** 30, (1), 43.
18. Li, J.; Leng, L.; Pantouris, G.; Manjula, R.; Piecychna, M.; Abriola, L.; Hu, B.; Lolis, E.; Armstrong, M. E.; Donnelly, S. C.; Bucala, R., A small-molecule allele-selective transcriptional inhibitor of the MIF immune susceptibility locus. *J Biol Chem* **2024,** 300, (7), 107443.
19. Lubetsky, J. B.; Swope, M.; Dealwis, C.; Blake, P.; Lolis, E., Pro-1 of macrophage migration inhibitory factor functions as a catalytic base in the phenylpyruvate tautomerase activity. *Biochemistry* **1999,** 38, (22), 7346-54.

 doi:10.20944/preprints202411.0131.v1

20. Odh, G.; Hindemith, A.; Rosengren, A. M.; Rosengren, E.; Rorsman, H., Isolation of a new tautomerase monitored by the conversion of D-dopachrome to 5,6-dihydroxyindole. *Biochem Biophys Res Commun* **1993,** 197, (2), 619-24.
21. Chen, L. H.; Kenyon, G. L.; Curtin, F.; Harayama, S.; Bembenek, M. E.; Hajipour, G.; Whitman, C. P., 4-Oxalocrotonate tautomerase, an enzyme composed of 62 amino acid residues per monomer. *J Biol Chem* **1992,** 267, (25), 17716-21.
22. Stivers, J. T.; Abeygunawardana, C.; Whitman, C. P.; Mildvan, A. S., 4-Oxalocrotonate tautomerase, a 41-kDa homohexamer: backbone and side-chain resonance assignments, solution secondary structure, and location of active site residues by heteronuclear NMR spectroscopy. *Protein Sci* **1996,** 5, (4), 729-41.
23. Subramanya, H. S.; Roper, D. I.; Dauter, Z.; Dodson, E. J.; Davies, G. J.; Wilson, K. S.; Wigley, D. B., Enzymatic ketonization of 2-hydroxymuconate: specificity and mechanism investigated by the crystal structures of two isomerases. *Biochemistry* **1996,** 35, (3), 792-802.
24. de Jong, R. M.; Bazzacco, P.; Poelarends, G. J.; Johnson, W. H., Jr.; Kim, Y. J.; Burks, E. A.; Serrano, H.; Thunnissen, A. M.; Whitman, C. P.; Dijkstra, B. W., Crystal structures of native and inactivated cis-3-chloroacrylic acid dehalogenase. Structural basis for substrate specificity and inactivation by (R)-oxirane-2-carboxylate. *J Biol Chem* **2007,** 282, (4), 2440-9.
25. Almrud, J. J.; Poelarends, G. J.; Johnson, W. H., Jr.; Serrano, H.; Hackert, M. L.; Whitman, C. P., Crystal structures of the wild-type, P1A mutant, and inactivated malonate semialdehyde decarboxylase: a structural basis for the decarboxylase and hydratase activities. *Biochemistry* **2005,** 44, (45), 14818-27.
26. Johnson, W. H., Jr.; Hajipour, G.; Whitman, C. P., Stereochemical Studies of 5-(Carboxymethyl)-2-hydroxymuconate Isomerase and 5-(Carboxymethyl)-2-oxo-3-hexene-1,6-dioate Decarboxylase from Escherichia coli C: Mechanistic and Evolutionary Implications. *Journal of the American Chemical Society* **1995,** 117, (34), 8719-8726.
27. Poelarends, G. J.; Whitman, C. P., Evolution of enzymatic activity in the tautomerase superfamily: mechanistic and structural studies of the 1,3-dichloropropene catabolic enzymes. *Bioorg Chem* **2004,** 32, (5), 376-92.
28. Poelarends, G. J.; Johnson, W. H., Jr.; Murzin, A. G.; Whitman, C. P., Mechanistic characterization of a bacterial malonate semialdehyde decarboxylase: identification of a new activity on the tautomerase superfamily. *J Biol Chem* **2003,** 278, (49), 48674-83.
29. Poelarends, G. J.; Serrano, H.; Johnson, W. H., Jr.; Hoffman, D. W.; Whitman, C. P., The hydratase activity of malonate semialdehyde decarboxylase: mechanistic and evolutionary implications. *J Am Chem Soc* **2004,** 126, (48), 15658-9.
30. Poelarends, G. J.; Serrano, H.; Johnson, W. H., Jr.; Whitman, C. P., Inactivation of malonate semialdehyde decarboxylase by 3-halopropiolates: evidence for hydratase activity. *Biochemistry* **2005,** 44, (26), 9375-81.
31. Pantouris, G.; Ho, J.; Shah, D.; Syed, M. A.; Leng, L.; Bhandari, V.; Bucala, R.; Batista, V. S.; Loria, J. P.; Lolis, E. J., Nanosecond Dynamics Regulate the MIF-Induced Activity of CD74. *Angew Chem Int Ed Engl* **2018,** 57, (24), 7116-7119.
32. Mischke, R.; Gessner, A.; Kapurniotu, A.; Juttner, S.; Kleemann, R.; Brunner, H.; Bernhagen, J., Structure activity studies of the cytokine macrophage migration inhibitory factor (MIF) reveal a critical role for its carboxy terminus. *FEBS Lett* **1997,** 414, (2), 226-32.
33. El-Turk, F.; Cascella, M.; Ouertatani-Sakouhi, H.; Narayanan, R. L.; Leng, L.; Bucala, R.; Zweckstetter, M.; Rothlisberger, U.; Lashuel, H. A., The conformational flexibility of the carboxy terminal residues 105-114 is a key modulator of the catalytic activity and stability of macrophage migration inhibitory factor. *Biochemistry* **2008,** 47, (40), 10740-56.
34. Henzler-Wildman, K.; Kern, D., Dynamic personalities of proteins. *Nature* **2007,** 450, (7172), 964-72.
35. McLean, L. R.; Zhang, Y.; Li, H.; Choi, Y. M.; Han, Z.; Vaz, R. J.; Li, Y., Fragment screening of inhibitors for MIF tautomerase reveals a cryptic surface binding site. *Bioorg Med Chem Lett* **2010,** 20, (6), 1821-4.
36. Fenwick, R. B.; Orellana, L.; Esteban-Martin, S.; Orozco, M.; Salvatella, X., Correlated motions are a fundamental property of beta-sheets. *Nat Commun* **2014,** 5, 4070.
37. Bouvignies, G.; Bernadó, P.; Meier, S.; Cho, K.; Grzesiek, S.; Brüschweiler, R.; Blackledge, M., Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. *Proc Natl Acad Sci U S A* **2005,** 102, (39), 13885-90.
38. Chen, E.; Widjaja, V.; Kyro, G.; Allen, B.; Das, P.; Prahaladan, V. M.; Bhandari, V.; Lolis, E. J.; Batista, V. S.; Lisi, G. P., Mapping N- to C-terminal allosteric coupling through disruption of a putative CD74 activation site in D-dopachrome tautomerase. *J Biol Chem* **2023,** 299, (6), 104729.
39. Skeens, E.; Pantouris, G.; Shah, D.; Manjula, R.; Ombrello, M. J.; Maluf, N. K.; Bhandari, V.; Lisi, G. P.; Lolis, E. J., A Cysteine Variant at an Allosteric Site Alters MIF Dynamics and Biological Function in Homo- and Heterotrimeric Assemblies. *Front Mol Biosci* **2022,** 9, 783669.
40. Robert, X.; Gouet, P., Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **2014,** 42, (Web Server issue), W320-4.

 doi:10.20944/preprints202411.0131.v1

41. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **2011,** 7, 539.
42. DeLano, W. L., The PyMOL Molecular Graphics System. Delano Scientific, San Carlos. **2002**.
43. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004,** 25, (13), 1605-12.
44. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** 14, (1), 33-8, 27-8.
45. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **2017,** 14, (1), 71-73.
46. Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Henin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kale, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E., Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* **2020,** 153, (4), 044130.
47. Lange, O. F.; Grubmuller, H., Generalized correlation for biomolecular dynamics. *Proteins* **2006,** 62, (4), 1053-61.