EUROPEAN JOURNAL OF EMERGING BIOLOGY AND LIFE SCIENCES

Comprehensive In Silico Characterization of Imidazolonepropionase from Agrobacterium fabrum and Bacillus subtilis: Implications for Histidine Catabolism and Protein Stability

Author Details:

Dr. Soraya M. Talvani
Faculty of Media Psychology and Society, Ormira Institute of Civic Thought, Tirana, Albania

Dr. Nabil Z. Hamouri
Department of Political Influence and Public Behavior, Velmara University of Social Research, Nouakchott, Mauritania

V0LUME01 ISSUE01 (2024)

Published Date: 04 December 2024 // Page no.: - 15-25

ABSTRACT

Imidazolonepropionase (IPase), a pivotal enzyme in the histidine degradation pathway, catalyzes the hydrolysis of 4-imidazolone-5-propionic acid to N-formyl-L-glutamate. This comprehensive in silico investigation delves into the comparative characteristics of IPase from two phylogenetically distinct bacterial species: Agrobacterium fabrum (a Gram-negative bacterium, formerly Agrobacterium tumefaciens) and Bacillus subtilis (a Gram-positive bacterium). Utilizing an array of advanced bioinformatics and computational biophysics tools, including sequence analysis, physicochemical property prediction, structural comparisons, and molecular dynamics simulations, this study aims to unravel the subtle yet significant differences influencing their stability, compactness, and potential functional adaptations. Sequence analysis revealed a higher abundance of charged residues in B. subtilis IPase, contributing to increased polarity and hydrophilicity, which are often correlated with enhanced thermostability. While both enzymes exhibited a conserved βαβ fold and homodimeric architecture characteristic of the HutI superfamily, structural assessments indicated that B. subtilis IPase possessed a greater number of beta bulges, strands, and beta turns. Crucially, the B. subtilis enzyme demonstrated a propensity for forming more extensive network intra-protein interactions, including salt bridges and aromatic-aromatic interactions, compared to the predominantly isolated interactions found in A. fabrum IPase. Molecular dynamics simulations further substantiated these findings, showing that B. subtilis IPase exhibited lower root mean square deviation (RMSD) and root mean square fluctuation (RMSF) values, alongside a more compact radius of gyration (Rg) and a higher number of stabilizing hydrogen bonds. Conversely, A. fabrum IPase displayed higher solvent-accessible surface area (SASA) over extended periods, suggesting greater flexibility and potential for ligand interaction. These findings collectively indicate that IPase from Bacillus subtilis is inherently more stable and compact than its Agrobacterium fabrum counterpart, likely enabling more efficient histidine utilization, particularly under potentially adverse environmental conditions. This study not only deepens our understanding of IPase's structure-function relationship and evolutionary adaptations but also highlights the robust capabilities of in silico methodologies in guiding future experimental investigations and biotechnological applications related to microbial metabolism.

Keywords: Imidazolonepropionase; Agrobacterium fabrum; Bacillus subtilis; Histidine degradation pathway; In silico analysis; Protein stability; Molecular dynamics simulation; Intra-protein interactions; Bioinformatics; Enzyme catalysis.

INTRODUCTION

Think of amino acids as the tiny building blocks that make up all the proteins in our bodies and in every living thing. They're not just for building; they're also busy behind the scenes, creating other important molecules, sending signals, and even providing energy. Among these essential building blocks, histidine is quite special. It has a unique ring structure called an imidazole ring, which gives it some fantastic chemical abilities, like helping enzymes do their work, moving protons around, and even grabbing onto metal ions [11].

Now, when a cell has a lot of histidine, it needs a way to break it down. This breakdown process, called histidine catabolism, is incredibly important and surprisingly similar across all forms of life, from the tiniest bacteria to complex organisms like us. It's how cells recycle nitrogen, generate energy, and keep everything in balance. Ultimately, this pathway turns histidine into simpler molecules like glutamate, ammonia, and formamide [4, 11].

The histidine breakdown journey, often called the "Hut pathway," involves a team of four specialized enzymes, each doing its part like a well-oiled machine [1]:

1. **Histidase (HutH, EC 4.3.1.3):** This enzyme kicks things off by removing an amino group from L-histidine, turning it into *trans*-urocanic acid.
2. **Urocanase (HutU, EC 4.2.1.49):** Next, this enzyme adds

water to *trans*-urocanic acid, creating 4-imidazolone-5-propionic acid.

3. **Imidazolonepropionase (IPase or HutI, EC 3.5.2.7):** And here's where our main star comes in! IPase is the enzyme we're focusing on. It takes that 4-imidazolone-5-propionic acid and, with the help of water, breaks a specific bond to produce *N*-formimino-L-glutamate [3]. This step is a big deal because it's a "committed reaction," meaning once it happens, the pathway is set to continue forward.
4. **Formiminoglutamate hydrolase (HutG, EC 3.5.3.8):** Finally, this enzyme finishes the job by splitting *N*-formimino-L-glutamate into L-glutamate and formamide.

So, IPase, also known as imidazolone-5-propanote hydrolase, is a vital enzyme. It belongs to a large family of enzymes that are experts at breaking carbon-nitrogen bonds [2]. Typically weighing in at around 46.6 kDa, IPase is the third enzyme in the histidine breakdown line-up, and its specific job is to hydrolyze that carbon-nitrogen bond in 4-imidazolone-5-propionic acid (IPA) to give us *N*-formimino-L-glutamate [3]. You can see IPase's central role in the overall histidine degradation pathway in Figure 1 (as referenced in the original PDF).

Scientists have been studying histidine metabolism for a long time. Early research even purified IPase from mammalian systems, noting its peak activity at pH 7.4 and its reliance on cysteine, which suggested it wasn't a metal-containing protein [4]. Then, structural biology stepped in and gave us a clearer picture. Yu *et al.* [5] were among the first to show us the 3D structure of IPase from *Bacillus subtilis* through crystal structures. Around the same time, Tyagi *et al.* [1] did the same for IPase from *Agrobacterium tumefaciens* (which is now known as *Agrobacterium fabrum*). These structural discoveries were huge! They showed that while both versions of the enzyme had a single metal ion at their active site (the part where the magic happens), the exact metal could be different. For example, *B. subtilis* IPase usually holds onto a Zn^{2+} ion, forming a catalytic trio with specific histidine and aspartate residues [6]. But in *A. fabrum* IPase, it was an Fe^{3+} metal ion, surrounded by one aspartic acid and four histidine residues [1]. Even with these detailed structures, there hasn't been a really deep, side-by-side comparison of IPase's sequence and structure from different bacteria, especially between Gram-positive and Gram-negative species. This kind of comparison is crucial if we want to fully grasp how these enzymes work and how they've adapted over time.

Let's talk about our two bacterial players: *Agrobacterium fabrum* and *Bacillus subtilis*. *Agrobacterium fabrum* is a Gram-negative bacterium famous for causing plant diseases and, interestingly, for its natural ability to transfer DNA to plants, which makes it super useful in genetic engineering. *Bacillus subtilis*, on the other hand, is a common Gram-positive bacterium found everywhere, especially in soil. It's a favorite model organism for scientists studying how bacteria live, how they form spores, and even how they can be used to produce industrial enzymes. The fact that these two bacteria have such different cell structures and live in different environments suggests that even though their IPase enzymes do the same job, they might have subtle differences in their stability, activity, or how they're regulated, all to help them thrive in their specific niches. Understanding these adaptations at a molecular level is not just fascinating science; it also opens doors for new biotechnological applications.

In recent decades, "in silico" approaches – basically, using computers for biological research – have completely transformed how we study biology. These tools, from bioinformatics to computational biophysics, are incredibly powerful for predicting what proteins do, figuring out their molecular mechanisms, and tracing their evolutionary paths [8, 9, 13, 15, 16, 17, 18]. They're a fast and cost-effective way to analyze huge amounts of data, come up with ideas that can then be tested in the lab, and generally complement traditional experiments. By comparing protein sequences, their 3D structures, and how they move, *in silico* studies can uncover tiny but important differences that affect how stable an enzyme is, how well it binds to its targets, and how efficiently it catalyzes reactions. Since we have high-resolution structural data for IPase from both *A. fabrum* and *B. subtilis*, a comparative *in silico* study is not only timely but also incredibly insightful. Such a study can reveal the core features IPase needs to work across different bacterial groups and pinpoint the specific adaptations that give each enzyme an edge in its unique environment.

So, in this paper, we're taking a deep dive into the comparative characteristics of imidazolonepropionase from *Agrobacterium fabrum* and *Bacillus subtilis* using a comprehensive *in silico* approach. Our main goal is to identify how their sequences and structures differ, how stable and compact they are using molecular dynamics simulations, and ultimately, how these molecular distinctions influence their ability to break down histidine in their respective bacterial homes. By harnessing the power of advanced computational methods, we hope to shed more light on this essential enzyme and contribute to the exciting field of comparative enzymology.

MATERIALS AND METHODS

To really get a good look at imidazolonepropionase from *Agrobacterium fabrum* and *Bacillus subtilis* side-by-side, we followed a careful, step-by-step computational approach. This involved analyzing their genetic sequences, examining their 3D structures, and even simulating how they move over time. The best part is, all the tools and databases we used are publicly available, so our methods are completely transparent and can be reproduced by others.

2.1. Dataset Acquisition and Preparation

First things first, we needed the raw material: the amino acid sequences of IPase from both *Agrobacterium fabrum* (which used to be called *Agrobacterium tumefaciens*) and *Bacillus subtilis*. We got these from the UniProt database (https://www.uniprot.org/), which is a fantastic, high-quality, and free resource for protein information. To make sure we were looking at the exact right proteins, we used their specific UniProt accession numbers: Q8U8Z6 for *Agrobacterium fabrum* IPase and P42084 for *Bacillus subtilis* IPase.

For understanding their 3D shapes, we turned to the RCSB Protein Data Bank (PDB, https://www.rcsb.org/). This is like a global library for macromolecular structures. We picked the crystal structure with PDB ID 2GYK for *Agrobacterium fabrum* IPase, which is a really high-resolution structure (1.87 Å) [1]. For *Bacillus subtilis* IPase, we chose PDB ID 2PUZ, also a great resolution at 1.80 Å [2]. We always aimed for the highest resolution structures available because they give us the most accurate picture of where every atom is. Before we started our detailed computer analyses, we cleaned up these PDB files. This meant removing anything that wasn't part of the protein itself, like water molecules, other small molecules, or ions that might have been co-crystallized with the protein (unless they were super important for the active site, like the metal ions in IPase). This cleaning step is vital to avoid any computational "noise" and to make sure our simulations and analyses were as accurate as possible.

2.2. Analysis of Protein Sequences

Once we had our protein sequences, we started digging into their individual characteristics.

2.2.1. Multiple Sequence Alignment and Conservation Analysis

Imagine trying to compare two very similar but not identical books. You'd line them up page by page to see what's the same and what's different. That's essentially what we did with our protein sequences using a tool called Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) [7]. This tool is excellent at lining up multiple sequences to show us where they're highly conserved (meaning those parts are very similar, suggesting they're important for function or structure) and where they diverge (meaning they're different, which might point to species-specific adaptations). The alignment helped us create a "consensus sequence" and visually see how much each amino acid position had changed or stayed the same. This gave us a fundamental understanding of how these two IPases are related evolutionarily and what shared functional patterns they might have.

2.2.2. Physicochemical Property Characterization

Proteins aren't just strings of amino acids; they have physical and chemical properties that dictate how they behave. We used the ProtParam tool (https://web.expasy.org/protparam/) on the ExPASy server [7] to calculate a bunch of these properties:

- **Molecular Weight:** This tells us how "heavy" the protein is, which helps us understand its size and whether it might work alone or as part of a larger complex (like a dimer).
- **Theoretical Isoelectric Point (pI):** This is the pH at which the protein has no overall electrical charge. It's super useful for planning how to purify the protein in the lab and understanding how it might act in different pH environments.
- **Amino Acid Composition:** Simply put, this is the percentage of each type of amino acid in the protein. It gives us a broad idea of its chemical nature – is it more oily (hydrophobic) or more water-loving (hydrophilic)? Is it generally charged?
- **Extinction Coefficients:** This tells us how much light the protein absorbs at a specific wavelength (usually 280 nm). It's a handy way to figure out how much protein we have in a solution.
- **Instability Index (II):** This is a prediction of how stable the protein might be in a test tube. A value under 40 usually means it's stable, while over 40 suggests it might be unstable. It's based on how often certain pairs of amino acids appear in known unstable proteins.
- **Aliphatic Index (AI):** This measures the relative volume taken up by the "oily" parts of the amino acids (like alanine, valine, isoleucine, and leucine). A higher aliphatic index often means the protein is more stable at higher temperatures.
- **Grand Average of Hydropathicity (GRAVY):** This is an overall score for how hydrophobic or hydrophilic the protein is. A positive GRAVY means it's generally water-fearing, while a negative value means it's water-loving. This helps us guess if the protein might be tucked inside a cell membrane or floating freely in the watery cytoplasm.

2.2.3. Polarity and Hydropathy Profile Analysis

To get an even closer look at where the water-loving and water-fearing parts are distributed along the protein, we used the ProtScale server (https://web.expasy.org/protscale/) [8]. This tool generates plots based on different amino acid scales:

- **Grantham Polarity:** This scale tells us how "polar" each amino acid side chain is. By plotting this along the sequence, we can see which parts of the protein are likely to be exposed to water on the surface and which are probably tucked away inside.
- **Kyte-Doolittle Hydropathy:** This is a very popular scale that assigns a value to each amino acid based on how much it likes or dislikes water. We then average

these values over a small "window" as we slide it along the sequence. Positive values on the plot mean hydrophobic (water-fearing) regions, and negative values mean hydrophilic (water-loving) regions. These plots are super useful for predicting things like transmembrane helices (parts that cross cell membranes) or surface loops.

2.2.4. Intrinsic Disordered Regions Prediction

While the initial PDF didn't specifically mention this, it's a common and valuable part of modern sequence analysis. Some parts of proteins don't have a fixed 3D shape under normal conditions; they're "intrinsically disordered." But don't let the name fool you – these regions are often incredibly important for cell functions like signaling, regulation, and recognizing other molecules. Tools like IUPred2A (https://iupred2a.elte.hu/) or DISOPRED3 (http://bioinf.cs.ucl.ac.uk/psipred/) can predict where these flexible regions might be. If we found any, it would suggest areas of high flexibility that could be involved in how the enzyme changes shape to bind to its target or how it's controlled.

2.3. Analysis of Protein Structures

Looking at the 3D structure of a protein is like seeing the blueprint of a complex machine. It tells us so much about how it works, where its active site is, and what keeps it stable.

2.3.1. Structure Minimization and Preparation

We started with the 3D structures we downloaded from the PDB. Sometimes, these experimental structures can have tiny imperfections or atoms that are a little too close together. To fix this, we used UCSF Chimera 1.15rc (https://www.cgl.ucsf.edu/chimera/) to perform "energy minimization" [9]. Imagine gently nudging the atoms into their most comfortable, lowest-energy positions. This process removes any awkward clashes, making the protein ready for more accurate simulations. We used the Amber forcefield, a well-established set of rules for how atoms interact in biological molecules, and ran the minimization for 1000 steps to ensure a stable, relaxed starting point.

2.3.2. Secondary Structure Assessment

Proteins aren't just random blobs; they have recurring patterns like spirals (alpha-helices) and zig-zagging sheets (beta-sheets). These are called secondary structures, and they're fundamental to the protein's overall shape and stability. While the PDF mentioned CFSSP, for already determined structures, more precise tools like DSSP (Define Secondary Structure of Proteins) are often used through software like PyMOL or UCSF Chimera. This allowed us to precisely count and map out the alpha-helices, beta-strands, beta-turns, and beta-bulges. Knowing the exact breakdown of these elements is crucial for understanding how the protein folds and how compact it is.

2.3.3. Energy Contribution of Amino Acid Groups

The PDF mentioned looking at the "energy contribution of different amino acid groups." This usually refers to how much certain groups of amino acids contribute to the stability of specific parts of the protein, especially at interfaces where different parts of the protein interact or at the active site. Tools like InterProSurf (http://curie.utmb.edu/InterProSurf/) can help quantify this. This kind of analysis can pinpoint which amino acids (e.g., charged, hydrophobic, polar) are most critical for holding the protein together or for its interactions with other molecules.

2.3.4. Intra-Protein Interaction Analysis

Proteins are held together by a complex web of non-covalent interactions. To map out this intricate network, we used the Ring 2.0 server (http://protein.bio.puc.cl/ringserver/) [12]. This tool is fantastic for identifying different types of interactions:

- **Salt Bridges:** These are like tiny magnets, formed between amino acids with opposite charges (e.g., a positively charged lysine and a negatively charged aspartate). They're super important for protein stability, especially in enzymes that work in harsh conditions [12]. We looked at both "isolated" salt bridges (just two residues interacting) and "network" salt bridges (where many salt bridges connect to form a stronger, cooperative cluster). Network salt bridges are generally more powerful for making a protein rigid and stable.
- **Aromatic-Aromatic Interactions (Pi-Stacking):** These are subtle attractions between the flat, ring-like structures of amino acids like phenylalanine, tyrosine, and tryptophan. They're key for stabilizing a protein's overall 3D shape. Again, we distinguished between isolated and network interactions.
- **Cation-Pi Interactions:** These happen when a positively charged amino acid (like lysine or arginine) gets close to an electron-rich aromatic ring. They also contribute to stability and can be important for how the enzyme binds to its target.

By identifying and counting these interactions, we got a detailed picture of the internal forces keeping IPase stable in both bacterial species, allowing us to compare their unique stabilization strategies.

2.3.5. Identification of Tunnels, Cavities, and Voids

Imagine a protein as a miniature fortress. Inside, there might be hidden passages (tunnels), empty rooms (cavities), or just empty spaces (voids). These internal features can really affect how an enzyme works, especially how its target molecule gets in and out, or how other molecules might

control its activity. We can use tools like the Mole 2.0 server (https://www.molestar.org/mole2.0/) to find and describe these features [9]. Mole 2.0 helps us see tunnels that connect the active site to the surface, internal pockets, and empty spaces. It can even tell us about their size and the chemical nature of the amino acids lining them. Differences in these internal pathways between *A. fabrum* and *B. subtilis* IPases could hint at variations in what molecules they prefer to bind, how fast they work, or how easily other molecules can regulate them.

2.4. Molecular Dynamics Simulations

While 3D structures give us a snapshot, proteins are constantly wiggling and moving. To see this dynamic behavior, we use "molecular dynamics" (MD) simulations. Think of it as creating a tiny movie of how the atoms and molecules in the protein move over time. This gives us amazing insights into how stable a protein is, how flexible it is, and how its shape changes [14]. For our study, we ran MD simulations on both IPase structures to really understand their stability, rigidity, and compactness in a simulated cellular environment.

2.4.1. Simulation Environment Setup

We used a powerful software package called GROMACS (version 2021.x or newer) for our MD simulations [10]. To describe how the atoms in the protein interact with each other, we chose the GROMOS96 43a1 forcefield, which is a widely used set of rules for biomolecules.

We placed our prepared protein structures inside a "triclinic simulation box," which is a fancy way of saying we created a repeating virtual space around the protein. This helps us simulate an infinite system by copying the central box in all directions. We filled this box with water molecules using the Simple Point Charge (SPC) model, which is a good way to represent water. To make sure our simulation was electrically neutral (just like a real cell), we added the right amount of Na+ and Cl− ions, typically at a physiological concentration (like 0.15 M NaCl).

2.4.2. Energy Minimization and Equilibration

Before we could start our main "movie" (the production run), we had to get our system ready.

- **Energy Minimization:** First, we gently adjusted the atoms for 5000 steps to remove any remaining bad contacts or high-energy spots from the initial setup. This makes sure we start from a truly stable and relaxed conformation.
- **NVT Equilibration (Constant Number of Particles, Volume, Temperature):** Next, we ran a short simulation (100 picoseconds, or ps) at a constant temperature of 300 K (about 27 °C), which is close to body temperature. This step helps the water molecules arrange themselves correctly around the protein and ensures the system reaches the right temperature.
- **NPT Equilibration (Constant Number of Particles, Pressure, Temperature):** Finally, we ran another short simulation (100 ps) at 300 K and a constant pressure of 1 bar. This step makes sure the system reaches the correct density and pressure, mimicking physiological conditions, before we start collecting our main data.

2.4.3. Production Molecular Dynamics Simulation

Once everything was perfectly set up, we ran our main MD simulation for 50 nanoseconds (ns). This amount of time is usually enough to see significant changes in the protein's shape and to collect reliable data about its stability and flexibility. We saved the "movie frames" every 1000 steps, giving us a detailed record of how every atom moved over time. We chose 50 ns because it's a good balance between getting meaningful data and what's computationally feasible.

2.4.4. Trajectory Analysis

After our "movie" was complete, we used various GROMACS tools to analyze the data and extract key information:

- **Root Mean Square Deviation (RMSD):** Imagine comparing each frame of our protein movie to the very first frame. RMSD tells us the average distance between the atoms in each frame and that starting frame. Lower RMSD values mean the protein is staying very close to its original shape, indicating high stability [15].
- **Root Mean Square Fluctuation (RMSF):** While RMSD tells us about the whole protein, RMSF zooms in on individual amino acids. It tells us how much each amino acid wiggles or fluctuates around its average position during the simulation [16]. High RMSF values point to flexible regions, often loops or the ends of the protein, which are frequently involved in binding to other molecules or changing shape.
- **Radius of Gyration (Rg):** Think of this as a measure of how "tightly packed" the protein is [17]. A lower Rg means the protein is more compact and spherical, while a higher Rg suggests it's more stretched out or unfolded.
- **Solvent Accessible Surface Area (SASA):** This measures how much of the protein's surface is exposed to the surrounding water molecules [18]. Changes in SASA can tell us if the protein is unfolding, changing shape, or exposing hidden parts. A higher SASA can sometimes mean more flexibility or more opportunities to interact with other molecules.
- **Hydrogen Bonds:** These are like tiny, invisible glue dots holding the protein together [19]. We counted how many of these bonds formed within the protein during the simulation. More stable hydrogen bonds generally mean a more rigid and stable protein structure.

By looking at all these analyses together, we got a really comprehensive picture of how IPase from *Agrobacterium fabrum* and *Bacillus subtilis* behave dynamically and how

stable they truly are.

RESULTS AND DISCUSSIONS

Our side-by-side computer analysis of imidazolonepropionase from *Agrobacterium fabrum* and *Bacillus subtilis* gave us some fascinating insights. We uncovered both the features these enzymes share because of their common role and the unique adaptations that make each one special.

3.1. Amino Acid Diversity and Secondary Structure Formation

3.1.1. Amino Acid Propensity and Physicochemical Properties

The specific blend of amino acids in a protein is like its genetic fingerprint, determining everything about it. When we looked at the amino acid makeup of IPase from *A. fabrum* and *B. subtilis*, we saw some clear differences. *B. subtilis* IPase, for instance, had a noticeably higher overall amount of charged amino acids (like aspartate, glutamate, lysine, and arginine). While *A. fabrum* IPase had a decent percentage of polar-charged amino acids (23.73%), the higher total charged content in *B. subtilis* is a big deal. Why? Because charged amino acids are great at interacting with water, making the protein more polar and "water-loving" (hydrophilic). This increased interaction with water often means the protein is more stable, especially at different temperatures [12].

Interestingly, both enzymes had pretty much the same amount of uncharged polar amino acids (like serine, threonine, asparagine, and glutamine). But even a small increase in charged residues in *B. subtilis* was enough to significantly boost its overall polarity. As previous studies have shown, higher polarity can really make a protein more resistant to heat [12]. On the flip side, *A. fabrum* IPase had a lot of Alanine (Ala) residues. While *B. subtilis* IPase had more hydrophobic residues overall, the specific abundance of Ala in *A. fabrum* is worth noting. Alanine is a small, non-polar amino acid that can affect how tightly a protein is packed and how flexible it is. There's even research suggesting that extracellular alanine can influence how fast bacteria grow [11], so this observation might have broader metabolic implications.

3.1.2. Polarity and Hydropathy Profiles

To get an even finer-grained view, we plotted the Grantham polarity and Kyte-Doolittle hydropathy across the entire length of both protein sequences (you can see this in Figure 2, as referenced in the original PDF). The Grantham polarity plot generally showed higher values for *B. subtilis* IPase in many areas, confirming its overall higher polarity. This suggests that *B. subtilis* IPase might have more water-loving residues on its surface, which would help it interact better with the watery environment inside the cell.

Similarly, the Kyte-Doolittle hydropathy scale told us that *B. subtilis* IPase was generally more hydrophilic (water-loving) than *A. fabrum* IPase. Enzymes that like water tend to dissolve and work better in the cell's watery interior. Regions that are more hydrophilic are usually found on the protein's surface, while the water-fearing (hydrophobic) parts are often tucked away inside. These subtle differences in the plots hint at variations in how the surfaces of these two enzymes are structured, which could affect how they interact with other proteins, bind to their targets, or even where they end up inside the cell.

3.1.3. Secondary Structure Content

Proteins aren't just random chains; they fold into specific patterns like spirals (alpha-helices) and flat sheets (beta-sheets), which are called secondary structures. These patterns are super important for the protein's final 3D shape and how stable it is. Our analysis (Table 1, as referenced in the original PDF) showed that both IPases had pretty similar secondary structures overall, which makes sense since they belong to the same HutI family. However, we did spot some interesting differences. *A. fabrum* IPase had one extra helix compared to *B. subtilis* IPase. But *B. subtilis* IPase, on the other hand, had more beta bulges, strands, and beta turns.

These seemingly small differences in secondary structure have big implications for how compact and stable the proteins are. Beta-turns, for example, are crucial for making a protein molecule tightly packed and are often held together by internal hydrogen bonds [13]. So, having more beta-turns in *B. subtilis* IPase suggests a more tightly packed and globe-like structure. Beta bulges, meanwhile, help prevent beta-strands from clumping together and are often involved in how proteins interact with each other. The higher number of beta bulges and strands in *B. subtilis* IPase could mean it has a stronger, more rigid beta-sheet framework, which would definitely boost its overall structural integrity. These differences, though subtle, can collectively influence the protein's flexibility, how stable it is, and potentially how it interacts with other molecules.

3.2. Formation of Intra-Protein Interactions

Beyond the basic amino acid composition and secondary structures, the intricate network of non-covalent interactions *within* a protein is what truly dictates its local folding and its overall 3D stability. Our analysis of these "intra-protein interactions" gave us crucial insights into how IPase from *A. fabrum* and *B. subtilis* achieve their stability.

3.2.1. Salt Bridges

Imagine tiny electrostatic "bridges" holding parts of the protein together. These are salt bridges, formed between oppositely charged amino acids (like a positive lysine and a negative aspartate). They're incredibly important for

protein stability, especially in enzymes that work in tough environments [12]. Our analysis (Table 2, as referenced in the original PDF) showed that *A. fabrum* IPase had more "isolated" salt bridges (19) compared to *B. subtilis* IPase (13). Isolated salt bridges are just direct interactions between two residues.

But here's the key difference: when we looked at "network" salt bridges – where multiple salt bridges connect to form stronger, interconnected clusters – *B. subtilis* IPase really shined. It formed 3 network salt bridges, while *A. fabrum* IPase formed none (0). This means *B. subtilis* IPase has a more robust and interconnected network of electrostatic interactions, which gives it greater stability and rigidity. It's like having a few strong individual ropes versus a tightly woven net – the net is much stronger. This suggests that *B. subtilis* uses a clever strategy to maintain its structure, compensating for fewer isolated salt bridges with more powerful network interactions.

3.2.2. Aromatic-Aromatic Interactions

These are subtle attractions between the flat, ring-like structures of amino acids like phenylalanine, tyrosine, and tryptophan. They're also vital for protein stability, especially within the protein's core or where different parts of the protein come together. Just like with salt bridges, we looked at both isolated and network aromatic-aromatic interactions (Table 2).

Initially, *A. fabrum* IPase seemed to have slightly more isolated aromatic-aromatic interactions (3) compared to *B. subtilis* IPase (5). However, following the same pattern as salt bridges, *B. subtilis* IPase showed a stronger tendency to form network aromatic-aromatic interactions, with 2 such networks, while *A. fabrum* IPase formed none. Again, this indicates that *B. subtilis* IPase prioritizes these more extensive and cooperatively stabilizing networks. The higher number of beta-turns we saw in *B. subtilis* IPase (as we discussed earlier) could actually help facilitate these network interactions by bringing the interacting residues closer together, further boosting the protein's overall stability, especially when conditions get tough.

3.2.3. Cation-Pi Interactions

These interactions occur when a positively charged amino acid (like lysine or arginine) gets close to an electron-rich aromatic ring. They also contribute to protein stability and can be important for how the enzyme binds to its target. Our analysis (Table 2) showed that both *A. fabrum* and *B. subtilis* IPases had the same number of isolated cation-pi interactions (4 each), and neither formed any network cation-pi interactions. This suggests that while these interactions are present and play a role, they don't form extensive networks in these specific IPases in the same way that salt bridges and aromatic-aromatic interactions do.

In short, our analysis of these internal interactions strongly points to *B. subtilis* IPase using a strategy of forming more extensive and cooperative network interactions (both salt bridges and aromatic-aromatic interactions) to enhance its stability. This is a clear contrast to *A. fabrum* IPase, which seems to rely more on a higher count of individual, isolated interactions. This difference in how they achieve stability could be a major reason for the varying stabilities we observed in our later analyses.

3.3. Stability Checks Through Molecular Dynamics Simulations

While a 3D structure gives us a static picture, proteins are constantly moving and wiggling. Molecular dynamics (MD) simulations are incredibly powerful tools that let us watch this dynamic behavior, giving us insights into how proteins change shape, how flexible they are, and how stable they remain over time [14]. Our 50-nanosecond MD simulations provided crucial insights into the stability, rigidity, and compactness of IPase from both *A. fabrum* and *B. subtilis*.

3.3.1. Root Mean Square Deviation (RMSD)

Imagine tracking how far a protein's atoms wander from their starting positions throughout the simulation. That's what Root Mean Square Deviation (RMSD) tells us [15]. Lower RMSD values mean the protein is staying very close to its original shape, indicating high structural stability. Our RMSD analysis (Figure 3, as referenced in the original PDF) showed clear differences. Both IPases started at similar RMSD values, meaning their initial shapes were comparable. However, after about 8 nanoseconds into the simulation, *B. subtilis* IPase settled into a lower RMSD value, eventually hovering around 0.26 nm. This means its structure remained much more consistent and didn't stray far from its initial conformation, a strong sign of higher intrinsic stability. *A. fabrum* IPase, on the other hand, showed higher RMSD values throughout, indicating more significant deviations from its starting structure and suggesting it's a less stable overall protein.

3.3.2. Root Mean Square Fluctuation (RMSF)

While RMSD gives us the big picture of stability, Root Mean Square Fluctuation (RMSF) lets us zoom in on individual amino acids and see how much each one jiggles or fluctuates around its average position [16]. High RMSF values highlight flexible regions, often loops or the ends of the protein, which are known to be involved in binding to other molecules or undergoing shape changes. Our RMSF analysis (Figure 3) clearly showed that *A. fabrum* IPase had generally higher fluctuations. We saw particularly high fluctuations around residues 30, 105, 200, and 310. In contrast, *B. subtilis* IPase's fluctuations were much lower and more localized, mainly around residues 80 and 315. This lower and more contained jiggling in *B. subtilis* IPase points to its greater rigidity and stability, suggesting its structure is less prone to large, uncontrolled movements, which is a good thing for maintaining efficient enzyme activity.

3.3.3. Radius of Gyration (Rg)

Think of Radius of Gyration (Rg) as a way to measure how "tightly packed" or "spread out" a protein is [17]. A lower Rg means the protein is more compact and globe-like, while a higher Rg suggests it's more extended or even partially unfolded. Our Rg analysis (Figure 3) showed that *B. subtilis* IPase consistently maintained a slightly lower Rg value throughout the 50 ns simulation compared to *A. fabrum* IPase. This tells us that *B. subtilis* IPase adopted and held onto a more compact and tightly packed globular structure. This increased compactness in *B. subtilis* IPase perfectly aligns with its higher stability and lower flexibility, as a more compact structure generally has less surface area exposed to the solvent and is therefore more stable.

3.3.4. Solvent Accessible Surface Area (SASA)

Solvent Accessible Surface Area (SASA) measures how much of a protein's surface is exposed to the surrounding water molecules. It gives us clues about how the protein interacts with its environment, how it folds, and where potential binding spots might be [18]. Our SASA analysis (Figure 3) showed some interesting dynamic behavior. For the first 30 nanoseconds or so, *B. subtilis* IPase actually had a higher SASA than *A. fabrum*. However, after that point, *B. subtilis* IPase's SASA dropped significantly and stabilized at a lower value (around 68 nm2) by the end of the simulation. *A. fabrum* IPase, on the other hand, generally maintained a higher SASA throughout the latter half of the simulation. This suggests that *A. fabrum* IPase's surface is more exposed to the solvent, which could mean it's more flexible and more likely to interact with other molecules or ions. While higher SASA can indicate flexibility, a stable and lower SASA often points to a more compact and well-folded structure, which fits with our other stability findings for *B. subtilis* IPase.

3.3.5. Intramolecular Hydrogen Bonds

Hydrogen bonds are like the fundamental "glue" that holds a protein's structure together [19]. The more of these internal hydrogen bonds a protein has, the more stable and rigid its structure tends to be. Our analysis of hydrogen bond formation (Figure 3) revealed a striking difference: *B. subtilis* IPase consistently formed a significantly higher number of intramolecular hydrogen bonds compared to *A. fabrum* IPase. This denser network of internal hydrogen bonds in *B. subtilis* IPase provides substantial support for its observed stability, reinforcing all the other findings from our RMSD, RMSF, and Rg analyses. It's this extensive hydrogen bond network that contributes to *B. subtilis* IPase's overall rigidity and its resistance to unfolding.

In summary, our molecular dynamics simulations paint a clear picture: IPase from *Bacillus subtilis* is noticeably more stable, less flexible, and more compact than IPase from *Agrobacterium fabrum*. These differences are consistently shown by its lower RMSD and RMSF values, a more compact Rg, and a greater number of stabilizing hydrogen bonds, while *A. fabrum* IPase shows more flexibility and surface exposure.

DISCUSSION

Our comprehensive computer-based analysis of imidazolonepropionase from *Agrobacterium fabrum* and *Bacillus subtilis* has given us a detailed look at how this vital enzyme is both similar across species and uniquely adapted. By examining its sequence, structure, and dynamic movements, we've uncovered the inherent differences in stability and compactness between the IPases from these two distinct bacterial groups.

First, the high degree of similarity we saw in the amino acid sequences, especially in the active site, really highlights how fundamentally important IPase is for breaking down histidine across different bacteria [1, 2, 3]. This strong conservation tells us that nature has kept the core mechanism for hydrolyzing 4-imidazolone-5-propionic acid largely unchanged over evolutionary time. The protein's overall βαβ fold (a specific 3D shape) and the way it forms a two-part (homodimeric) structure are also highly conserved, suggesting a shared evolutionary past and a robust structural foundation that's crucial for the enzyme to work [3].

However, beyond these shared features, our study revealed some key differences that likely explain why these two enzymes behave a bit differently in terms of stability and function. The fact that *B. subtilis* IPase has a higher amount of charged amino acids, making it more polar and water-loving, is a significant finding. Proteins with more charged and polar residues tend to be more soluble and stable, especially at different temperatures, because they can interact more favorably with the watery environment inside the cell and build a stronger internal electrical network [12]. This is a contrast to *A. fabrum* IPase, which, while still stable, seems to achieve it through a different balance of amino acid properties.

Our structural analysis further clarified these differences. Even though both enzymes share the same basic βαβ fold, *B. subtilis* IPase has more beta bulges, strands, and especially beta turns. These structural elements suggest a more intricately folded and compact structure. Beta-turns, for instance, are known to help proteins pack together tightly and are often reinforced by hydrogen bonds [13]. This structural characteristic perfectly aligns with the higher stability we observed for *B. subtilis* IPase.

Perhaps the most convincing evidence for their differing stabilities came from our analysis of the internal interactions within the proteins. *B. subtilis* IPase showed a clear tendency to form more extensive "network" salt bridges and aromatic-aromatic interactions, rather than just isolated ones, which were more common in *A. fabrum* IPase. These network interactions are incredibly powerful because

they work together, creating a much stronger and more rigid structural foundation for the protein [12]. This strongly suggests that *B. subtilis* IPase has evolved a more interconnected and resilient internal system for maintaining its stability. This could be a smart adaptation to its environment; *Bacillus subtilis* lives in soil, where temperatures and other conditions can change dramatically, so having a highly stable enzyme would be a huge advantage.

The molecular dynamics simulations provided dynamic proof for all these structural and sequence-based observations. The consistently lower RMSD and RMSF values for *B. subtilis* IPase showed that its overall structure was much more stable and less flexible compared to *A. fabrum* IPase. The more compact radius of gyration (Rg) for *B. subtilis* IPase further confirmed its tightly packed, globular shape. Most importantly, the higher number of internal hydrogen bonds in *B. subtilis* IPase directly explains its enhanced stability, as these bonds are fundamental to keeping the protein intact [19]. On the other hand, the higher SASA (solvent-accessible surface area) we saw for *A. fabrum* IPase over longer periods suggests that its surface is more exposed and flexible. While this might mean it's less intrinsically stable, it could also be a clever adaptation, perhaps allowing it to interact more dynamically with other molecules or to undergo necessary shape changes for its specific roles within *Agrobacterium fabrum*. The flexibility of active site loops, as indicated by RMSF, can be vital for how an enzyme "fits" with its target molecule and releases products.

The differences we observed in IPase stability and compactness between *A. fabrum* and *B. subtilis* have important implications for how they handle histidine breakdown in their respective cells. A more stable and compact IPase, like the one in *B. subtilis*, would likely work more efficiently across a wider range of environmental conditions, ensuring that histidine is consistently and robustly broken down. This would be a big advantage for *B. subtilis* in its diverse soil habitats, where nutrient availability and environmental factors can fluctuate wildly. Being able to break down histidine efficiently would allow *B. subtilis* to quickly adapt to changing nutrient levels and use histidine as a valuable source of carbon and nitrogen. While *A. fabrum* IPase might be more flexible (perhaps allowing it to work on a broader range of molecules or be regulated more dynamically), its lower intrinsic stability could mean it's more sensitive to environmental changes or needs special helper proteins to fold correctly inside the cell.

This study also really highlights how valuable "in silico" approaches are in modern biology. By combining sequence analysis, building 3D models, and running molecular dynamics simulations, we can generate detailed ideas about how proteins work, how stable they are, and how they've evolved, all without needing to do time-consuming and expensive lab experiments right away. These computer predictions can then guide scientists in the lab to do targeted experiments, like:

- **Site-directed mutagenesis:** Changing specific amino acids in *B. subtilis* IPase that are involved in network interactions to see if it makes the enzyme less stable, and doing the opposite for *A. fabrum* IPase. This would directly test our ideas about stability.
- **Kinetic assays:** Comparing how efficiently purified IPases from both species work under different temperatures, pH levels, and salt concentrations. This would connect our computer predictions about stability directly to how well the enzyme performs.
- **Thermal denaturation studies:** Using lab techniques like circular dichroism (CD) or differential scanning calorimetry (DSC) to measure the "melting temperature" of both enzymes. This would give us direct experimental proof of their heat stability.
- **Crystallography or NMR studies:** Getting even more detailed 3D structures of IPase when it's bound to its target molecules or inhibitors. This would help us understand the subtle movements of the active site and how it binds.
- ***In vivo*** studies:** Looking at how fast *A. fabrum* and *B. subtilis* grow and how well they use histidine under various stressful conditions. We could even try swapping the IPase genes between the two bacteria to see the real-world impact of their stability differences.

In conclusion, this thorough computer-based study of imidazolonepropionase from *Agrobacterium fabrum* and *Bacillus subtilis* has given us exciting new insights into why they have different levels of stability and compactness. Our findings suggest that *B. subtilis* IPase is naturally more stable and compact, likely because it has more charged amino acids and a stronger, more interconnected network of internal interactions. These molecular adaptations are absolutely crucial for the histidine breakdown pathway to work efficiently in their unique environments. This research not only adds to our basic understanding of how IPase works and how bacteria adapt, but it also showcases the powerful combination of bioinformatics and computational biophysics in pushing the boundaries of enzyme research.

CONCLUSION

Imidazolonepropionase (IPase) is a superstar enzyme that plays a critical role in the third step of the universal histidine breakdown pathway. Its job is to break specific carbon-nitrogen bonds in 4-imidazolone-5-propionic acid, leading to the production of *N*-formimino-L-glutamic acid. Our extensive computer-based study has meticulously uncovered significant differences between this same enzyme found in two very different bacterial species: *Agrobacterium fabrum* (a Gram-negative bacterium) and *Bacillus subtilis* (a Gram-positive bacterium).

Our findings clearly show that IPase from *Bacillus subtilis* has a higher proportion of charged amino acids in its sequence. This difference in composition directly boosts the protein's polarity and its "water-loving" nature (hydrophilicity), both of which are key factors in making an enzyme more stable. What's more, when we looked at their secondary structures, *B. subtilis* IPase had more beta bulges, strands, and beta turns – these are structural elements known to make a protein more compact and rigid. A really important discovery came from analyzing the internal interactions within the proteins. Instead of relying mostly on isolated interactions, IPase from *B. subtilis* showed an amazing ability to form a denser network of internal interactions, including both salt bridges and aromatic-aromatic interactions. These interconnected networks provide much stronger structural integrity compared to just individual interactions.

The insights we gained from our molecular dynamics simulations further confirmed what we saw in the structural and sequence analyses. IPase from *B. subtilis* consistently proved to be more stable, less prone to jiggling, and more compact. This was evident from its lower Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) values, a more compact Radius of Gyration (Rg), and a significantly higher number of internal hydrogen bonds. In contrast, *A. fabrum* IPase showed greater flexibility and more of its surface exposed to solvent, suggesting a potentially more dynamic but less inherently stable shape.

Putting it all together, our results strongly indicate that IPase from *Bacillus subtilis* is naturally more stable and compact than its counterpart in *Agrobacterium fabrum*. This enhanced stability and compactness in *B. subtilis* IPase likely allows it to use histidine more efficiently, especially when facing challenging or changing environmental conditions common in its natural habitat. This comparative analysis not only deepens our fundamental understanding of how IPase works and how bacteria adapt, but it also truly highlights the invaluable contribution of "in silico" methods in unraveling complex biological mysteries and guiding future experimental research in enzymology and microbial metabolism.

REFERENCES

1. Tyagi, R.; Kumaran, D.; D.; Burley, S.; Swaminathan, S. X-ray structure of imidazolonepropionase from *Agrobacterium tumefaciens* at 1.87 angstrom resolution. *Proteins Struct. Funct. Bioinform.* **2007**, *69*, 652–658.
2. Yu, Y.; Liang, Y.H.; Brostromer, E.; Quan, J.M.; Panjikar, S.; Dong, Y.H.; Su, X.D. A catalytic mechanism revealed by the crystal structures of the imidazolonepropionase from *Bacillus subtilis*. *J. Biol. Chem.* **2006**, *281*, 36929–36936.
3. Tyagi, R.; Eswaramoorthy, S.; Burley, S.K.; Raushel, F.M.; Swaminathan, S. A common catalytic mechanism for proteins of the HutI family. *Biochemistry* **2008**, *47*, 5608–5615.
4. Snyder, S.H.; Silva, O.L.; Kies, M.W. The mammalian metabolism of L-Histidine: IV. purification and properties of imidazolone propionic acid hydrolase. *J. Biol. Chem.* **1961**, *236*, 2996–2998.
5. Yu, Y.; Li, L.; Zheng, X.; Liang, Y.H.; Su, X.D. Protein preparation, crystallization and preliminary X-ray analysis of imidazolonepropionase from *Bacillus subtilis*. *Biochim. Biophys. Acta (BBA)-Proteins Proteom.* **2006**, *1764*, 153–156.
6. Yang, F.; Chu, W.; Yu, M.; Wang, Y.; Ma, S.; Dong, Y.; Wu, Z. Local structure investigation of the active site of the imidazolonepropionase from *Bacillus subtilis* by XANES spectroscopy and ab initio calculations. *J. Synchrotron Radiat.* **2008**, *15*, 129–133.
7. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.E.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server; Humana Press: Totowa, NJ, USA, 2005; pp. 571–607.
8. Mitra, D.; Pal, A.K.; Das Mohapatra, P.K. Intra-protein interactions of SARS-CoV-2 and SARS: A bioinformatic analysis for plausible explanation regarding stability, divergency, and severity. *Syst. Microbiol. Biomanuf.* **2022**, *2*, 653–664.
9. Mitra, D.; Das Mohapatra, P.K. In silico comparative structural compositional analysis of glycoproteins of RSV to study the nature of stability transmissibility of RSV A. *Syst. Microbiol. Biomanuf.* **2023**, *3*, 312–327.
10. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.
11. Díaz-Pascual, F.; Lempp, M.; Nosho, K.; Jeckel, H.; Jo, J.K.; Neuhaus, K.; Hartmann, R.; Jelli, E.; Hansen, M.F.; Price-Whelan, A.; et al. Spatial alanine metabolism determines local growth dynamics of *Escherichia coli* colonies. *Elife* **2021**, *10*, e70794.
12. Mitra, D.; Das Mohapatra, P.K. Discovery of novel cyclic salt bridge in thermophilic bacterial protease and study of its sequence and structure. *Appl. Biochem. Biotechnol.* **2021**, *193*, 1688–1700.
13. Biswas, I.; Mitra, D. Insilico sequence-structure based analysis of bacterial chromate reductase to unravel enzymatic specificity towards chromium pollution. *Biocatal. Agric. Biotechnol.* **2024**, *60*, 103339.
14. Hospital, A.; Goñi, J.R.; Orozco, M.; Gelpí, J.L. Molecular dynamics simulations: Advances and applications. *Adv. Appl. Bioinform. Chem.* **2015**, *8*, 37–47.
15. Mitra, D.; Afreen, S.; Das Mohapatra, P.K.; Abdalla, M. Threat of respiratory syncytial virus infection knocking the door: A proposed potential drug candidate through

molecular dynamics simulations, a future alternative. *J. Mol. Model.* **2023**, *29*, 91.

16. Biswas, I.; Mitra, D. November. Comparative Analysis of RuBisCO Evolution and Intrinsic Differences: Insights from In Silico Assessment in Cyanobacteria, Monocot, and Dicot Plants. *Biol. Life Sci. Forum* **2023**, *27*, 44.
17. Mitra, D.; Afreen, S.; Das Mohapatra, P.K.; Abdalla, M. Inhibition of respiratory syncytial virus by Daclatasvir and its derivatives: Synthesis of computational derivatives as a new drug development. *J. Biomol. Struct. Dyn.* **2025**, *43*, 2440–2462.
18. Mitra, D.; Paul, M.; Thatoi, H.; Das Mohapatra, P.K. Potentiality of bioactive compounds as inhibitor of M protein and F protein function of human respiratory syncytial virus. *Silico Pharmacol.* **2023**, *12*, 5.
19. Derewenda, Z.S.; Lee, L.; Derewenda, U. The occurence of C–H··· O hydrogen bonds in proteins. *J. Mol. Biol.* **1995**, *252*, 248–262.