Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

Insight into gene fusion from molecular dynamics simulation of fused and un-fused IGPS (Imidazole Glycerol Phosphate Synthetase)

Yu Yiting, Li Lei, Meena Kishore Sakharkar, Pandjassarame Kangueane*

School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50, Nanyang Avenue, Singapore 639798; Pandjassarame Kangueane* - E-mail: mpandjassarame@ntu.edu.sg; Phone: +65 6790 4957; Fax: +65 6774 4340;

* Corresponding author

received February 4, 2006; accepted February 27, 2006; published online February 28, 2006

Abstract:

Gene fusion produces proteins with novel structural architectures during evolution. Recent comparative genome analysis shows several cases of fusion/fission across distant phylogeny. However, the selection forces driving gene fusion are not fully understood due to the lack of structural, dynamics and kinetics data. Available structural data at PDB (protein databank) contains limited cases of structural pairs describing fused and un-fused structures. Nonetheless, we identified a pair of IGPS (imidazole glycerol phosphate synthetase) structures (comprising of HisF - glutaminase unit and HisH – cyclase unit) from *S. cerevisiae* (SC) and *T. thermophilus* (TT). The HisF-HisH structural units are domains in SC and subunits in TT. Hence, they are fused in SC and un-fused in TT. Subsequently, a domain-domain interface is formed in SC and a subunit-subunit interface in TT between HisF and HisH. Our interest is to document the structure and dynamics differences between fused and un-fused IGPS. Therefore, we probed into the structures of fused IGPS in SC and un-fused IGPS in TT using molecular dynamics simulation for 5ns. Simulation shows that fused IGPS in SC has larger interface area between HisF-HisH and greater radius of gyration compared to un-fused IGPS in TT. These structural features for the first time demonstrate the evolutionary advantage in generating proteins with novel structural architecture through gene fusion.

Keywords: gene fusion; fused proteins; evolution; molecular dynamics; interface; domains; subunits

Background:

Proteins with novel structural architectures are generated by gene fusion in one species' compared to another species. **[1, 2]** Proteome wide comparative analyses within and across kingdoms showed a large number of fused structures. **[3]** Proteins created by gene fusion are shown to have enhanced role in pathways by Yanai *et al.,* **[4]**, simulate protein subunit interaction by Marcotte *et al.,* **[5]**, novel function by Long **[6]**, enhanced substrate specificity by Katzen *et al.,* **[7]** and enzyme multi-functionality by Berthonneau and Mirande. **[8]** These reports indicate the existence of several isolated cases of fused protein as a result of gene fusion in evolutionary history. However, the advantage (structure, dynamics and kinetics) of producing fused proteins in one species compared to the un-fused protein orthologs in another species is not fully understood.

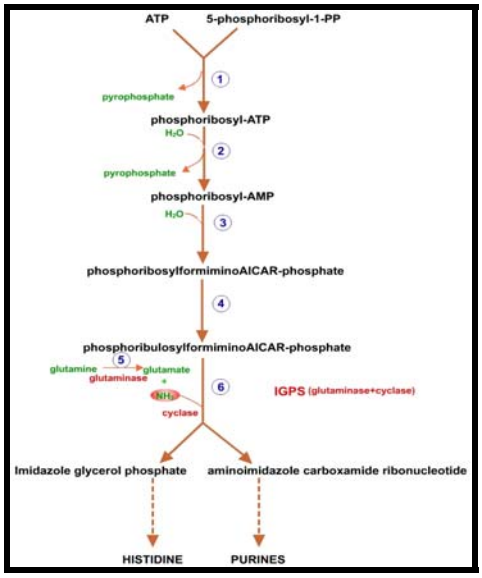

Figure 1: A schematic representation of histidine biosynthetic pathway is given. IGPS (imidazole glycerol phosphate synthase) catalyzes the fifth and sixth step of the histitine biosynthetic pathway in microbes, fungi, and plants. IGPS catalyzes the bifurcation step of the histidine and *de novo* purine biosynthesis pathways.

Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

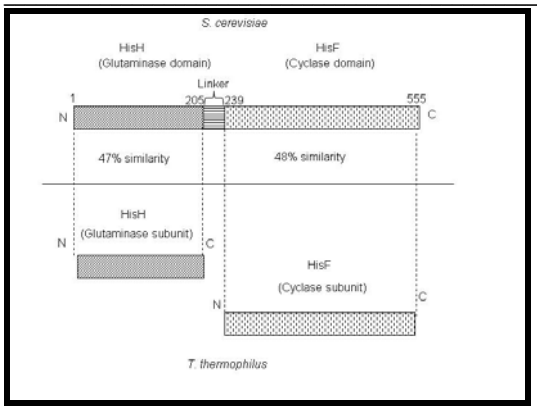

Figure 2: A schematic representation of the 555 residue long IGPS from SC and TT is shown. The HisH and HisF domains in SC are fused by a 33 residue long linker (206-238). However, the HisH and HisF subunits are un-fused in TT and the linker is absent.

Despite the availability of several protein structures at the PDB (protein databank), cases of structural pairs describing fused and un-fused protein structures are limited. Nevertheless, we identified a pair of IGPS (imidazole glycerol phosphate synthetase) structures (comprising of HisF - glutaminase unit and HisH – cyclase unit) from *S. cerevisiae* (SC) and *T. thermophilus* (TT). IGPS catalyzes the fifth and sixth steps of the histidine biosynthetic pathway in microbes, fungi, and plants. It forms the imidazole ring of the histidine precursor imidazole glycerol phosphate. **[9,10]** IGPS is a glutamine amido-transferase that catalyzes the formation of IGP (Imidazole glycerol phosphate) and AICAR (5-aminoimidazole-4-carboxamide ribonucleotide) from PRFAR (N^1- ((5′-phosphoribulosyl) formimino) - 5-aminoimidazole-4-carboxamide ribonucleotide). Interestingly, IGPS functions at the junction of histidine biosynthesis and *de novo* purine biosynthesis, since AICAR is the entry point to the latter (Figure 1). Thus, IGPS is a key metabolic enzyme, which links amino acid and nucleotide biosynthesis pathways. IGPS has different structural architectures in SC and TT. In TT, IGPS forms a hetero-dimer interface with glutaminase (HisH) and cyclase (HisF) subunits. **[1]** In SC, the two subunits are fused into a single polypeptide an N terminal HisH domain and a C terminal HisF domain forming an interface between HisH-HisF domains. **[11]** The conserved glutamine binding site in IGPS is at the interface of HisH and HisF in both TT and SC. **[12, 13]** Thus, the stability of the interface plays an important role in glutaminase catalysis. The subunit interaction in TT and domain interaction in SC mediate the catalytic activity of glutamine hydrolysis. **[9]** Thus, the fused protein retains the glutaminase active site and a small linker connects HisF and HisH in SC. However, the structure, dynamics and kinetics advantages of this arrangement in fused proteins are not known. Therefore, it is our interest to probe into the structure and dynamics properties of the fused (SC - IGPS) and un-fused (TT - IGPS) structures using molecular dynamics simulation.

Methodology:

Initial IGPS structures for simulation:

We used the IGPS structures for SC (PDB code: 1OX6 – resolution 2.4 Å) **[14]** and TT (PDB code: 1KA9 – resolution 2.3 Å) **[12]** downloaded from PDB. Hydrogen atoms were added to these structures using SYBYL 6.8 (Tripos Associates Inc.).

Molecular dynamics simulation:

All molecular mechanics calculations were carried out using the TRIPOS force field **[15]** in SYBYL (Molecular Modeling Software Package, Version 6.8, Tripos Associates Inc.) running on a Silicon Graphics Workstation. The energy function used in the force field was defined as the sum of six contributions (bond stretching, angle bending, torsion, van der Waals, electrostatic and planarity (for aromatic conjugated systems). Minimizations of the potential energy of the system were carried out using the Simplex algorithm and the Powell torsional gradient algorithm as implemented in SYBYL, terminating when a 0.5 Kcal/molÅ energy gradient shift was obtained. A distance dependent dielectric constant of 1.0 was used to compute electrostatic effects. The non-bonded cutoff distance used was 8 Å and the net atomic charges in the residues were calculated by the Gasteiger-Hucker method. **[16,17]** The *in vacuo* system was simulated at constant temperature, constant volume (NVT) ensemble which is referred to as the canonical ensemble. The system was run at a temperature of 300 K using a coupling constant of 100 femtosecond. The initial atom velocities were employed from a Maxwell-Boltzmann distribution with scaling velocities. The non-bonded pair list was updated every 25 femtosecond and an 8 Å cut-off was applied. During the simulation, the integration step was set up as 1 femtosecond and molecular snapshots were saved for every 1000 steps (1 pico-second). A total of 5000 structures were generated and the simulation properties were derived from the analyses of these snapshots.

Analysis:

We performed a comprehensive analysis of structures in each trajectory to detect structural differences between the two simulated systems. The flexibilities of the different structures were assessed by computing gap volume, gap index, interface area and radius of gyration.

Result:

Figure 2 illustrates the fused and un-fused IGPS structures in SC and TT, respectively. A small linker connects HisH (glutaminase) and HisF (cyclase) in SC and thus IGPS is fused in SC. However, this linker

is absent in TT and HisH – HisF are un-fused in TT. The HisH domain in SC has 47% similarity to the HisH subunit in TT. Similarly, the HisF domain in SC has 48% similarity to the HisF subunit in TT. The HisH and HisF units are homologous and have similar structures in SC and TT.

Figure 3 shows the structural snapshots of TT IGPS and SC IGPS at 0 and 5 ns simulation. The HisH and HisF interface in TT and SC is also visualized in Figure 3. The linker connecting HisH and HisF in SC is labeled and this linker is absent in TT. Thus, the interface is formed by HisH and HisF domains in SC and HisH and HisF subunits in TT. This demonstrates an evolutionary transition from a subunit-subunit interface in TT to a domain-domain interface in SC.

Figure 4 shows the interface area (change in solvent accessible surface area upon interface formation between HisH and HisF calculated using NACCESS implemented using Lee and Richard algorithm **[18]**) in TT IGPS and SC IGPS for structures generated over a 5 ns simulation. The interface area between HisH and HisF is significantly larger (> 1000 Å^2) in fused SC IGPS compared to the un-fused TT IGPS throughout the simulation period.

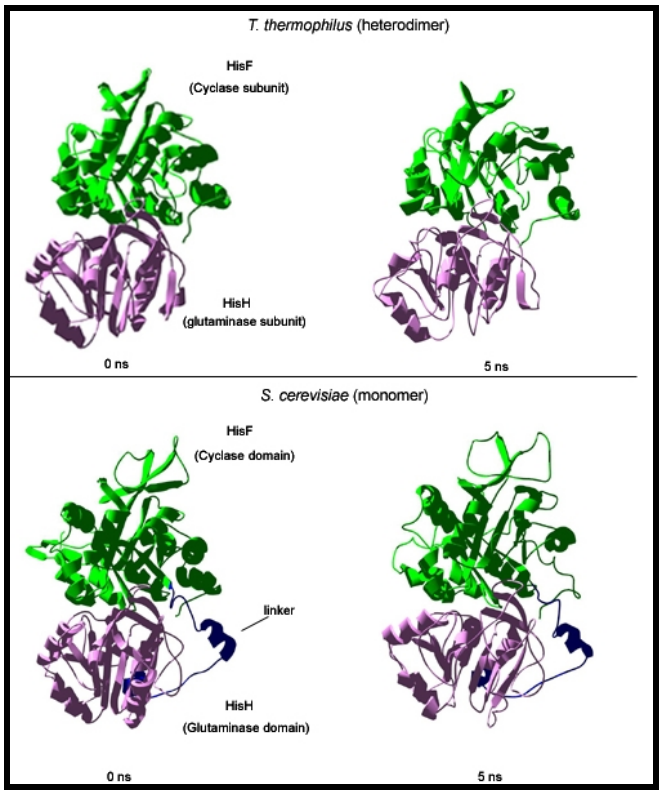

Figure 3: Snapshots of IGPS at 0 ns and 5 ns for SC and TT are shown. The molecules are rendered as a ribbon diagram with contrasting colors for the glutaminase (bottom) and cyclase (top) domains. The figure shows the bacterial IGPS is a heterodimer and the yeast IGPS is a monomer. The C-terminal cyclase domain of yeast IGPS has a longer loop at the top of the barrel than that of the bacteria.

Table 1: Residue conservation at the interface of IGPS in TT and SC

HisF (TT | SC)	Total	Interior	Interface	Surface
(a) No. of conserved residue	113	14	20	61
(b) No. of residues	317	34 | 43	42 | 43	241 | 231
(a)/(b)	35%	41% | 33%	47%	26%
HisH (TT | SC)	**Total**	**Interior**	**Interface**	**Surface**
(c) No. of conserved residue	69	9	17	23
(d) No. of residues	205	36 | 41	36 | 53	133 | 111
(c)/(d)	34%	25% | 22%	47% | 32%	17% | 21%

Data shows that interface residues are more conserved than surface residues for HisF and HisH between TT and SC. The number of conserved residues for HisF is 113 (> 95 == (14+20+61)) and the remaining 18 conserved residues are located at different regions (interior/interface/surface) in the two structures from TT and SC. This explanation holds true for the HisH structures in TT and SC.

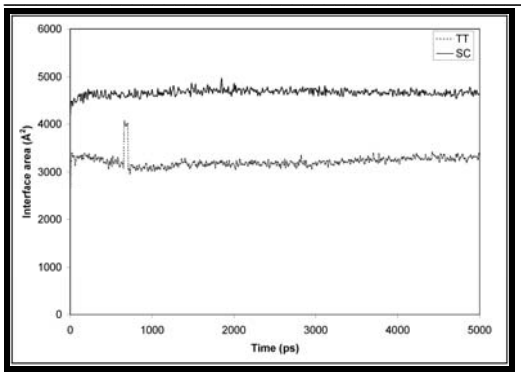

Figure 4: Interface area between HisH and HisF is given for IGPS from SC and TT over a 5 ns molecular dynamics simulation. The domain-domain interface area in SC is larger than TT throughout the simulation period.

Figure 5 shows the gap volume (calculated using SURFNET **[19]**) between HisH and HisF in SC IGPS and TT IGPS for structures generated over a 5 ns simulation. Similar to interface area, the gap volume is consistently larger in SC IGPS compared to TT IGPS throughout the simulation period.

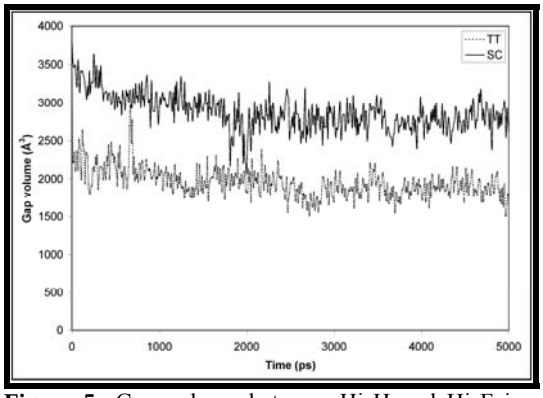

Figure 5: Gap volume between HisH and HisF is given for IGPS from SC and TT over a 5 ns molecular dynamics simulation. The domain-domain gap volume in SC is larger than TT throughout the simulation period.

Figure 6 shows the gap index (ratio of gap volume to interface area) between HisH and HisF in SC IGPS and TT IGPS for structures generated over a 5 ns simulation. Unlike interface area and gap volume, gap index is steadily similar throughout the simulation period.

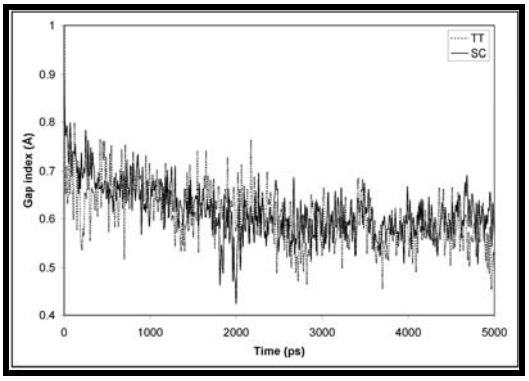

Figure 6: Gap index (ratio of volume to interface area) between HisH and HisF is given for IGPS from SC and TT over a 5 ns molecular dynamics simulation. The gap index is similar for the interface between HisH and HisF from SC and TT.

Figure 7 shows the radius of gyration for SC IGPS and TT IGPS for structures generated over a 5 ns simulation. Similar to interface area and gap volume, the radius of gyration for SC IGPS is considerably larger compared to TT IGPS throughout the simulation period.

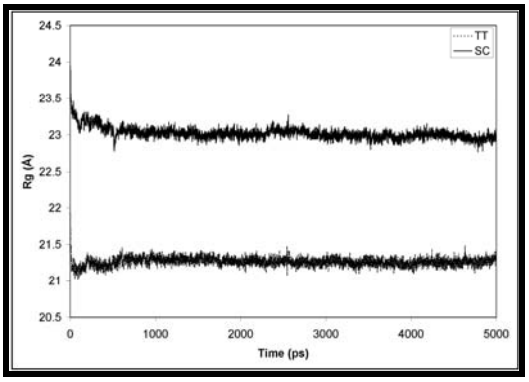

Figure 7: Radius of gyration (measure of unfolding and flexibility) for IGPS from SC and TT is given over a 5 ns molecular dynamics simulation. The radius of gyration is larger for SC IGPS is larger than TT IGPS throughout the simulation.

Table 2: Structural properties of IGPS in TT and SC is given for initial and final structures

Parameters	Initial crystal structure		Final structure after simulation (5 ns)		Difference between initial crystal and final structures	
	TT	SC	TT	SC	TT	SC
Interface area ($Å^2$)	2691.5	3940.3	1652.7291	2617.3474	-1039	-1323
Gap volume ($Å^3$)	3606	3952	3363.6	4627.1	-242	675
Gap index (Å)	0.746	0.997	0.491357207	0.565656	-0.256	-0.432
Radius of gyration (Å)	25.52	21.59	21.25	22.91	-4.27	1.32

Discussion:

Gene fusion is an important evolutionary phenomenon for the formation of proteins with new structural architectures. **[1-8]** Comparative sequence analysis between closely and distantly related species shows evidence for gene fusion/fission. **[1, 2, 3]** Therefore, it is of great significance to document the selection force generating such proteins with fused structural architectures. However, there is no documentation for structural evidence supporting the dynamics of these fused structures in the evolution of orthologous proteins.

The interface residues between HisF and HisH in TT and SC are more conserved than surface residues (Table 1). The interface residues similarities imply catalytic conservation at the interface. The structural properties for IGPS in TT and SC are given for initial and final structures (Table 2). The interface area, gap volume and gap index are greater in SC than TT in both initial and final structures. These values increased relatively due to simulation in both SC and TT. However, the radius of gyration in TT is larger than SC for the initial structure unlike the final structure (Table 2). Interestingly, the radius of gyration increased in SC and decreased in TT due to simulation.

The results given in Figure 3 to Figure 7 demonstrate the structure dynamics of fused IGPS in SC compared to the un-fused IGPS in TT. The IGPS in SC forms a domain-domain interface between HisH and HisF compared to a subunit-subunit interface in TT. The transition from a subunit-subunit interface in TT to a domain-domain interface in SC is interesting. The domain-domain interface area in SC is larger than the subunit interface area in TT over a 5 ns molecular dynamics simulation. The interface area in SC is 1400 $Å^2$ greater than in TT. The larger interface area in SC facilitates better domain-domain interactions compared to subunit interactions in TT (Figure 4). The amount of interface area determines the degree of atomic interaction at the interface. Larger HisH and HisF interface in SC imply better interaction between these two domains. Better interaction between HisF and HisH facilitates greater stability and kinetics in SC. This is assisted largely by the linker segment connecting HisF and HisH domains in SC.

The gap volume between HisF and HisH domains from SC IGPS is larger than that between HisF and HisH subunits in IGPS from TT (Figure 5). The increased gap volume in SC IGPS may aid in substrate flow into the active sites formed by HisH and HisF domains. However, this flow of substrate is relatively restricted in TT IGPS in exchange for interface stability formed by subunit interaction. Larger gap volume in SC IGPS is partly helped by the linker between HisH and HisF which provides enhanced flexibility for these two domains. Interestingly, the increased gap volume in SC IGPS does not affect gap index (ratio of gap volume to interface area) in both SC IGPS and TT IGPS (Figure 6). This suggests that increased gap volume is proportional to the increased interface in SC compared to that in TT.

Radius of gyration in proteins is a measure of their size and implies their compactness. The radius of gyration for IGPS from SC and TT given in Figure 7 describes the unfolding of the structure during simulation. The flexibility rendered by the linker between HisF and HisH in the case of SC IGPS is shown by the increased radius of gyration compared to that in TT throughout the simulation period over 5 ns. The difference in the average radius of gyration between SC and TT IGPS is about 1.76 Å. This provides the explanation for the increased stability leading to greater kinetics of IGPS caused by the linker in the fused structure of SC IGPS.

The raise and fall in interface area, gap volume and gap index in TT during simulation is unusual. This may be due to the high interface movement between the weakly associated subunits. The proposed hypothesis driving the formation of fused proteins by gene fusion is the structural determinant providing increased stability, dynamics and kinetics facilitated during evolutionary selection. This is evident by the structure and dynamics of IGPS as described using interface area, gap volume and radius of gyration in SC and TT.

Conclusion:

A number of fusion proteins have been identified by comparative genome analysis using sequence comparison. This suggests that gene fusion is common in evolutionary phylogeny. However, the selection force driving gene fusion in organism

evolution is not fully evident due to the lack of structure, dynamics and kinetics data supporting this phenomenon. Despite the growth in structures at PDB, the number of structural pairs illustrating fusion/fission in distant phylogeny is limited. Here, we show the importance of fused protein by probing the fused IGPS structure in SC as against the un-fused structure in TT using molecular dynamics simulation. The simulation shows larger interface area and radius of gyration in SC IGPS compared to TT IGPS. Thus, fused IGPS in SC have better structural features than un-fused IPGS in TT. This finding provides meaningful insight for gene fusion in establishing optimal dynamics and kinetics. This is an extremely interesting one and is likely to become more important as the international structural genomics efforts increase significantly their production of structures

References:

[1] Y. Yiting *et al., Front. Biosci.* 9:2964 (2004) [PMID: 15353329]

[2] M. K. Sakharkar *et al., Front. Biosci.* 10:1070 (2005) [PMID: 15769606]

[3] K. Truong & M. Ikura, *BMC Bioinformatics* 4:16 (2003) [PMID: 12734020]

[4] I. Yanai *et al., Proc. Natl. Acad. Sci. U.S.A.* 98: 7940 (2001) [PMID: 11438739]

[5] E. Marcotte *et al., Science* 285:751 (1999) [PMID: 10427000]

[6] Long M. *Genome Res.* 10:1655 (2000) [PMID: 11076848]

[7] F. Katzen *et al., EMBO J.* 21:3960 (2002) [PMID: 12145197]

[8] E. Berthonneau & M. Mirande, *FEBS Lett.* 470:300 (2000) [PMID: 10745085]

[9] T. J. Klem & V. J. Davisson, *Biochemistry* 32**:**5177 (1993) [PMID: 8494895]

[10] T. J. Klem *et al., Bacteriol.* 183 :989 (2001) [PMID: 11208798]

[11] S. V. Chittur *et al., Protein Expr. Purif.* 18 **:**366 (2000) [PMID: 10733892]

[12] R. Omi *et al., J. Biochem.* 132 :759 (2002) [PMID: 12417026]

[13] B. N. Chaudhuri *et al., Structure* 9 :987 (2001) [PMID: 11591353]

[14] B. N. Chaudhuri *et al., Biochemistry* 42 :7003 (2003) [PMID: 12795595]

[15] J. G. Vinter *et al., J. Comput. Aided. Mol. Des.* 1 :31 (1987) [PMID: 3505586]

[16] M. Marsili & J. Gasteiger, *Croat. Chem. Acta.* 53 :601 (1980)

[17] J. Gasteiger & M. Marsili, *Tetrahedron* 36 :3219 (1980)

[18] B. Lee & F. M. Richards, *J. Mol. Biol.* 55:379 (1971) [PMID: 5551392]

[19] R. A. Laskowski, *J. Mol. Graph.* 13:323 (1995) [PMID: 8603061]

Edited by K. Gunasekaran

Citation: Yiting *et al.,* Bioinformation 1(3): 99-104 (2006)