

PROFILES OF MOLECULAR FUNCTION - GENOMIC ENZYMOLOGY

JOHN L. ANDREASSI AND THOMAS S. LEYH*

Department of Biochemistry, The Albert Einstein College of Medicine, 1300 Morris Park Ave.,
Bronx, New York 10461-1926, U.S.A.

E-Mail: *leyh@acom.yu.edu

Received: 16th March 2004 / Published: 1st October 2004

ABSTRACT

A worldwide initiative, the goal of which is to place all of Nature's globular protein domains within modelling distance of a known three-dimensional structure, is underway. The tens of thousands of structures slated to be delivered to the scientific community by the Initiative over the ensuing decade will create an acute need for a complementary program to characterize the functions of these proteins. It is timely to consider the design of such a program.

INTRODUCTION

The rate at which protein structures are archived is remarkable (~ 5000/year in 2003) and continuing to increase. A primary aim of the global protein structural initiative has been to place all protein folds within modelling distance of a representative structure [1]. This endeavour is driven by a desire to understand the molecular functions of these structures. The coverage-of-protein-fold-space problem appears to be approaching a watershed from which the objectives of the Initiative are turning toward articulating the differences within a single domain type. With function in-hand, the domain architecture of a protein becomes a hierarchy of functional entities carefully positioned to accomplish a specific molecular task. As our understanding of the differences among domain-structures deepen, so does our need to understand how these differences produce changes in molecular function.

When a molecular biologist becomes interested in a specific protein, among the first pieces of information (s)he seeks are its structure, its conserved residues, and an understanding of the functions of those residues.

This information is fundamental in the sense that it is the information that the community naturally reaches for when it begins to pursue molecular explanations for biology - it is the information with which one begins to consider and to control (through mutagenesis) the molecular architecture that underlies biological function.

Often, though not exclusively, it is the conserved residues that lie at or near the surface of the protein that are of the greatest general interest. These residues, preserved through evolution, are the residues with which a protein family "senses" and interacts with its environment. Defining how these residues function establishes a signature, or *profile*, that describes the molecular operations of a given family. *This information is not generally available, and should be created for every family for which it is feasible.*

A database of family profiles would prove valuable in many areas of biology. For example, SNP (single nucleotide polymorphism) databases, currently under development for the human, mouse, fly and nematode genomes, are mined routinely to identify which mutations, from the millions that have been catalogued, are linked to disease and other interesting phenotypes. Cross-referencing disease-linked SNPs with a database of family-function profiles will associate the disease-causing mutations with specific alterations in protein function. In a second broad application, understanding how particular mutations effect function enables the geneticist to select the functional properties of the protein that will be bred into the genetic background of an organism. A database of profiles would provide an off-the-shelf resource for designing such transgenic experiments. As a third and final example, bioinformatics projects often aim at identifying the residues on which the features of a protein family pivot. Sophisticated models are being developed to scrutinize primary sequences and identify putative critical residues. A profile database will not only confirm the predictions of the models, but, importantly, will also provide data sets on which the models can be trained, tested and developed further.

Creating a database of profiles will require that the conserved residues positioned at the surface of a protein be identified (this can now be done computationally), that their functions be determined (through mutagenesis and functional characterization), and that the data be incorporated into the database. A database format that seems particularly intuitive for the user is one in which the conserved residues, presented in CPK on a rotatable structure, can be "clicked" to present a table describing the results of the functional studies on that residue (cf., Table 1, below). It would be valuable to have the table include links to the results of findings associated with a particular residue in other members of the same family.

Profiles of Molecular Function - Genomic Enzymology

Table 1. Effects of mutation on the initial-rate constants of the PMK-catalysed reaction.

| Protein | K_i (P-mev) (μM) | K_m (P-mev) (μM) | K_i (ATP) (mM) | K_m (ATP) (mM) | k_{cat} (sec^{-1}) |
|--------------------|---------------------------------|---------------------------------|------------------|------------------|--|
| Wild Type | ^a 22 (0.38) | 2.4 (0.13) | 2.85 (0.15) | 0.31 (0.007) | 18.3 (0.09) |
| | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| D150E | 39 (0.28) | 2.4 (0.0085) | 4.0 (0.016) | 0.25 (0.0015) | 1.54 (0.001) |
| | 1.8 | 1.0 | 1.4 | 0.8 | 0.084 |
| ¹ S291A | 1650 (172) | 45 (1.0) | 3.0 (0.074) | 0.083 (0.009) | 11.9 (0.083) |
| | 75 | 19 | 1.05 | 0.27 | 0.65 |
| ² A293T | ^b ND | 42,000 (5.8) | ND | 4.2 (0.7) | 0.28 (0.027) |
| | | 17,500 | | 14 | 0.015 |

^a A kinetic constant, and its standard error, are listed above the line; beneath it, the constant is normalized to that of the wild-type enzyme. ^bValue could not be determined.

Correlates

1) Mutation in human galactokinase at the position analogous to S291 in PMK causes type II galactosemia. (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>). [11]

2) Mutation in the human mavalonate kinase at the position analogous to A293 kinase causes mevalonate aciduria. (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>). [12]

RESULTS AND DISCUSSION**A case study-phosphomevalonate kinase (PMK)**

A preliminary exploration of the feasibility and efficacy of a large-scale effort to study protein function was carried out using the enzyme phosphomevalonate kinase, cloned from the R6 strain of *Streptococcus pneumoniae*. PMK was obtained from the Rockefeller University contingent of the New York Structural Genomics Research Consortium [2]. PMK was selected for structural characterization by the consortium because it is a member of the GHMP kinase family for which, at the time, there was no representative structure. The consortium uses sequence uniqueness as a criterion in selecting proteins for structure determination, which strongly biases the outcome toward novel structural folds, and directs the endeavour toward creating a diverse structural library that is representative of the folds found in Nature. As anticipated, the sequences unique to the GHMP kinase family manifest as novel structures which, in this case, are intimately involved catalysing phosphoryl-transfer [3, 4].

Phosphomevalonate kinase (ATP: 5-phosphomevalonate phosphotransferase, EC 2.7.4.2) catalyses transfer of the γ -phosphoryl group of ATP to (R)-5-phosphomevalonate, forming the pyrophosphoryl-linkage found in diphosphomevalonate (Reaction 1).



Reaction 1 is the second step in the so-called mevalonate pathway, which is comprised of a four-step sequence of reactions the end-product of which, isopentenyl diphosphate (IPP), provides the 5-carbon building blocks used in the biosynthesis of isoprenoids—a complex family of metabolites encompassing more than 23,000 compounds [5] including cholesterol, steroids, vitamin K₁₂, and the prenyl-moiety used to post-translationally modify and target proteins to the membranes [6]. In certain prokaryotes, IPP is linked into the linear undecaprenyl chains (C₅₅) that translocate peptidoglycans across the cell membrane during cell wall biosynthesis [5, 7].

Highly conserved, solvent-accessible GHMP kinase family residues were identified in several ways. The primary sequences of the GHMP kinase family were compared to identify both family-wide and PMK-specific conservation. These residues were mapped onto the PMK structure (PDB 1K47) to assess their surface accessibility. In an alternative strategy, GHMP kinase structures were aligned, using the *Vector Alignment Search Tool* (VAST) [8, 9], to identify surface accessible residues that are well-conserved in three-dimensional space, but not in sequence-space. These analyses identified a set of forty-seven candidate residues, twenty of which appear to be solvent accessible when mapped onto the PMK crystal structure (PDB ID 1k47) see Fig. 1. The structure suggested that nine of the twenty were integral to the structural core, and therefore removed from further consideration. The remaining eleven were characterized, and three of these are discussed here (D150E, S291A and A293T).

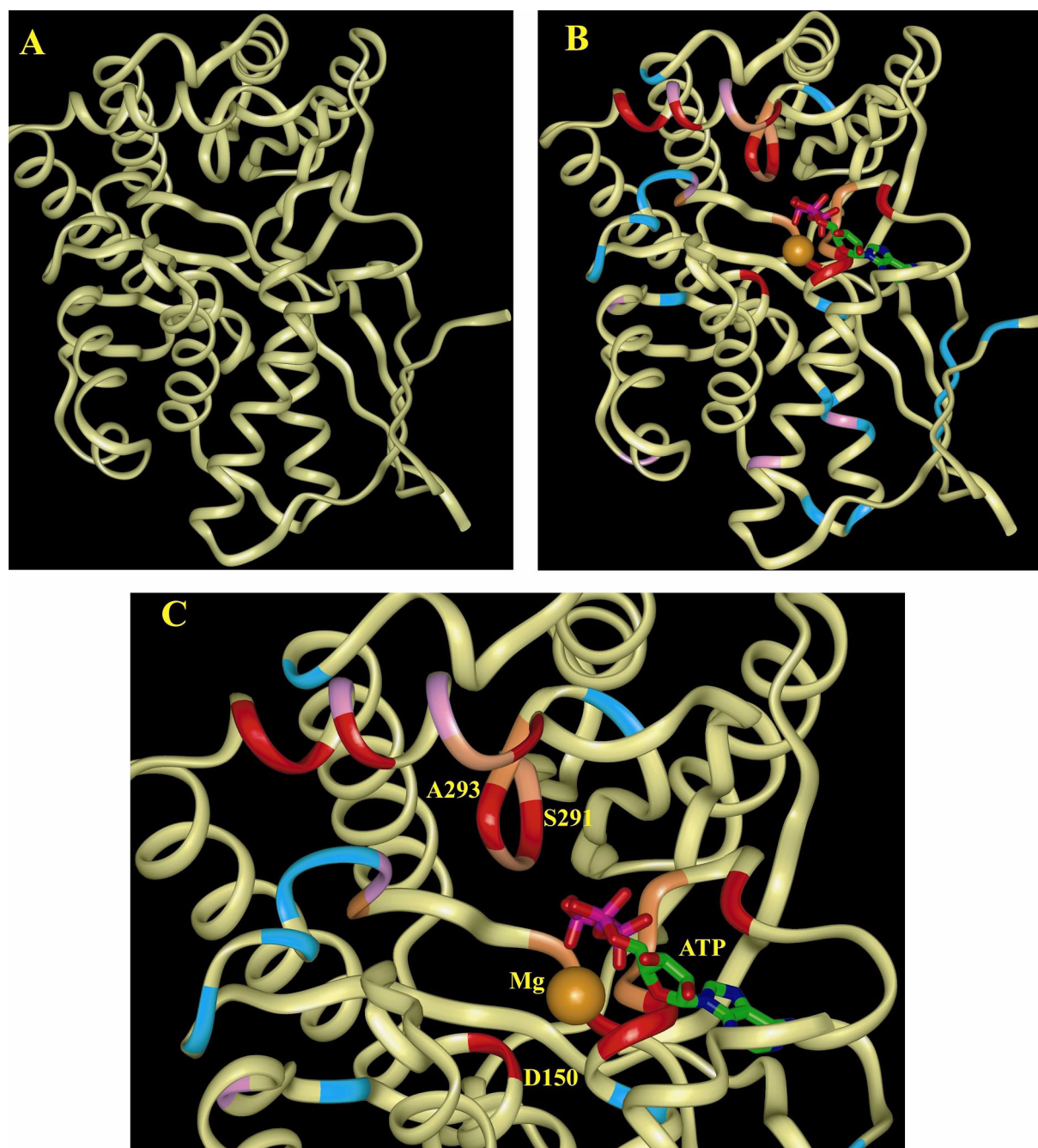


Figure 1. Conserved amino acids in PMK. (A.) A backbone representation of the structure of the wild-type, apo-PMK (PDB ID 1k47). (B.) The forty-seven, highly-conserved GHMP kinase residues are highlighted on the apo-PMK structure. Twenty of the forty-seven conserved residues are part of the hydrophobic core of the structure (Cyan), and 7 are involved in internal salt-bridge interactions. Of the remaining twenty residues, all of which were solvent accessible, nine appeared to be involved primarily in maintaining secondary structure (Orange), the remaining 11 residues (Red) were characterized. (C.) A close-up of the PMK active-site. The residues discussed in the text are labelled according to their sequence position. Mg²⁺·ATP was modelled into the PMK structure using the structure of the mevalonate kinase·Mg²⁺·ATP complex (PDB ID 1kvk) as a template. The figures were rendered using Insight II.

Aspartate 150

GHMP kinase structures support that this enzyme family uses a carboxylate-carrying residue (Asp or Glu) to catalytically position the divalent cation (Mg^{2+}) that is coordinated to the tripolyphosphate chain of ATP [10] - aspartic acid 150 in PMK. To explore whether and how this residue might participate in catalysis, a methylene group (-CH₂-) was inserted into the aliphatic chain from which the carboxylate "dangles" by replacing Asp150 with glutamic acid. The effects of lengthening the R-group on the kinetic parameters of the forward reaction are compiled in Table 1. The steady-state affinities of the substrates for various forms of the enzyme are influenced very little by the substitution (0.8 - 1.8-fold effects are observed). k_{cat} , however, decreases 12-fold when the chain is lengthened. Thus, the insertion selectively affects a rate-determining step(s) in the catalytic cycle - which is consistent with a repositioning of Mg^{2+} .

Serine 291

GHMP kinases active sites contain a small, glycine-rich loop that appears well positioned to interact with the substrate that receives the γ -phosphoryl group from ATP. To assess how this loop functions in catalysis, two of the loop's residues (S291 and A293) were altered, via mutagenesis, and the effects of the perturbations were evaluated. Replacing the hydroxyl group of S291 with a proton (S291A) causes $K_{m(P_{mev})}$ to increase 19-fold. Clearly, the S291 hydroxyl is important in recognition of phosphomevalonate during steady-state turnover. Interestingly, the substitution enhances the steady-state affinity of $Mg^{2+} \cdot ATP$ for the enzyme ~ 5 -fold.

The γ -phosphoryl group is the closest ATP moiety to the S291 hydroxyl, and these two groups are separated by ~ 11 Å. Notwithstanding a surprising reorganization of the active-site during catalysis, it appears that the hydroxyl is involved in a communication network between the S241 side chain and the determinants involved in the steady-state recognition of ATP that spans a distance of at least ~ 11 Å. While the S291A substitution affects the steady-state affinities of both substrates, its influence on k_{cat} is small, 1.5-fold.

Position 291 is conserved across mevalonate kinases, homoserine kinases and galactokinases as either Ser or Thr. Interestingly, mutating threonine 334 (the Ser 291 homologue) in human galactokinase results in a galactokinase deficiency that causes cataract formation in children not maintained on a lactose free diet [11]. Our PMK findings suggest that the galactokinase deficiency may be caused by a decreased steady-state affinity of galactokinase for galactose.

Alanine 293

Alanine 293, also in the glycine-rich substrate-recognition loop, was mutagenized to a threonine, and the effect of this substitution on catalysis was studied. The influence on $K_{m(\text{Pmev})}$ was profound (17,000-fold increase) while the effect on $K_{m(\text{ATP})}$ was relatively small (14-fold). Thus, we again see that perturbing the loop causes highly selective effects on the substrate affinities, with the greater effect on the affinity of the non-ATP substrate. Unlike the S291A mutation, the A293T substitution resulted in a large decrease (64-fold decrease) in k_{cat} . While it is not clear whether A293 itself contributes to the energetics of rate-determining steps in the catalytic cycle, it is clear that alterations at this site can influence the relative energetics/structures of the ground- and transition-state(s) of such steps.

In addition to its conservation among PMKs, alanine 293 is also highly conserved in mevalonate kinase across species. Mutating the Ala 334 codon of the human mevalonate kinase (the PMK A293 homologue) causes a mevalonate kinase deficiency that results in elevated levels of mevalonic acid in plasma and can lead to early fatality [12-14]. A human recombinant mevalonate kinase A334T mutant exhibited a k_{cat} decrease comparable to that seen for the PMK A293T mutant [12]. The mutagenesis results support the notion that a second, highly conserved, loop in the active site of GHMP kinases functions to recognize and position the non-nucleotide substrate for catalysis.

CONCLUSION

The findings described above illustrate well the value of family profiling. The behaviour of two of the mutants, S291A and A293T, in conjunction with the PMK structure, suggests that the active sites of GHMP kinase family members exhibit a conserved, substrate-recognition loop, and that the loop participates in an allosteric network that establishes an energetic poise between the substrate binding pockets. Furthermore, these findings provide precise descriptions of the metabolic lesions that are caused by these disease-linked mutations. The remaining mutant, D150A, is remarkable in that its effects are nearly exclusively focused at the rate-determining step(s) in the reaction, and its position in the PMK structure suggests that the effects are likely to be mediated through coordination of the ATP-bound Mg^{2+} ion. D150A seems an excellent candidate for reverse genetic exploration of the effects of down-regulating flux through the mevalonate pathway in model organisms.

It is impossible to predict the myriad of ways that the information available in a *Database of Family Profiles* will be applied by molecular biologists to their individual situations. What is clear, however, is that the profiling method has worked, and that the impact of the information that it yields is considerable. There are thousand of proteins for which this work has yet to be done, and the literature suggests that it will not likely be accomplished using *status-quo* mechanisms. The creation of such a database will require a multidisciplinary program.

ACKNOWLEDGMENTS

Supported by the National Institutes of Health Grants GM54469 and the Albert Einstein College of Medicine.

REFERENCES

- [1] Burley, S.K., Bonanno, J.B. (2002) Structuring the universe of proteins. *Annu. Rev. Genomics Hum. Genet.* **3**: 243-262.
- [2] Romanowski, M.J., Bonanno, J.B., Burley, S.K. (2002) Crystal structure of the *Streptococcus pneumoniae* phosphomevalonate kinase, a member of the GHMP kinase superfamily. *Proteins* **47**(4): 568-571.
- [3] Burley, S.K., Bonanno, J.B. (2002) Structural genomics of proteins from conserved biochemical pathways and processes. *Curr. Opin. Struct. Biol.* **12**(3): 383-391.
- [4] Bork, P., Sander, C., Valencia, A. (1993) Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein Sci.* **2**(1): 31-40.
- [5] Kharel, Y., Koyama, T. (2003) Molecular analysis of cis-prenyl chain elongating enzymes. *Nat. Prod. Rep.* **20**(1): 111-118.
- [6] Wang, K.C., Ohnuma, S. (2000) Isoprenyl diphosphate synthases. *Biochim. Biophys. Acta* **1529**(1-3): 33-48.
- [7] Fujihashi, M. et al. (2001) Crystal structure of cis-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase. *Proc. Natl. Acad. Sci. USA* **98**(8): 4337-4342.
- [8] Gibrat, J.-F., Midej, T., Bryant, S.H. (1996) Surprising similarities in protein structure. *Curr. Opin. Struct. Biol.* **6**: 377-385.
- [9] Madej, T., Gibrat, J.-F., Bryant, S.H. (1995) Threading a database of protein cores. *Protein Struct. Funct. Genet.* **23**: 356-369.
- [10] Fu, Z. et al. (2002) The structure of a binary complex between a mammalian mevalonate kinase and ATP: insights into the reaction mechanism and human inherited disease. *J. Biol. Chem.* **277**(20): 18134-18142.
- [11] Asada, M. et al. (1999) Molecular characterization of galactokinase deficiency in Japanese patients. *J. Hum. Genet.* **44**(6): 377-382.

- [12] Hinson, D.D. et al. (1997) Identification of an active site alanine in mevalonate kinase through characterization of a novel mutation in mevalonate kinase deficiency. *J. Biol. Chem.* **272**(42): 26756-26760.
 - [13] Hoffmann, G. et al.(1986) Mevalonic aciduria--an inborn error of cholesterol and nonsterol isoprene biosynthesis. *New Engl. J. Med.* **314**(25): 1610-1614.
 - [14] Hoffmann, G.F et al. (1993) Clinical and biochemical phenotype in 11 patients with mevalonic aciduria. *Pediatrics* **91**(5): 915-921.
-