MedCrave
Step into the World of Research

MOJ Proteomics & Bioinformatics

The internal oligopeptide sequences missing in crystals are disordererd domains

Abstract

Polypeptide sequences in pdb format are invariably shorter than those in FASTA format. The missing residues are mostly internal oligopeptide strings and few C & N terminal residues. We have compared the panorama of the secondary structure domains generated from both formats by folding *in silico* and find that the missing oligopeptides are mostly from the intrinsically distorted domains.

Keywords: protein crystals, fasta format, pdb fomat, protein secondary structure. disordered domain. α helix, β sheet, internal missing oligopeptides

Research Article

Volume 7 Issue 1 - 2018

Natasha Kelkar,[1] Sohan P Modak[2]
[1]Institute of Bioinformatics and Biotechnology, S. P. Pune University, India
[2]Open Vision, India

Correspondence: Sohan P Modak, Open Vision, 759/75, Deccan Gymkhana, Pune 411004, India, Email spmodak@gmail.com

Received: February 13, 2018 | **Published:** February 23, 2018

Introduction

Prior to their maturation as a biological structure or function, nascent polypeptides fold to form three dimensional structures composed of α helices, β sheets and disordered regions. The amino acid sequence of the processed polypeptide is stored in FASTA format (www.rcsb.org) and it is almost always longer than that in the crystal structure, retrievable in PyMol stored in pdb format, wherein the absence of residues has been noted at the C-terminal, N-terminal and at intra-polypeptide locations of crystals. Indeed, a large number of protein crystals in the data base exhibit internal missing string.[1] Crystallographers generally consider that the missing residues are due to low electron density undetectable in low resolution crystallography. Since some of the gaps at the N and C termini can be attributed to post-translational processing, the presence of missing internal oligopeptides may lead to misinterpretation of the secondary structure domains in the immediate vicinity of the gaps as well as in the flanking segments. While studying the phylogeny of proteins[2] we considered the possibility that the extent of evolutionary conservation of residues defining individual secondary structure domains may be one of the determinants. As we came across the cases of internal missing intra-molecular residues here we analyze their structure and significance.

Materials and methods

Amino acid sequences of 9 proteins were downloaded from RCSB pdb in FASTA and crystal format.[3] (www.rcsb.org) These are, (1) SAICAR synthase from *Saccharomyces cerevisiae*, strain ATCC 204508/S288c (PDB Id : 1A48),[4] (2) SAICAR synthase complexed with ADP, AICAR, and succinate from the same strain as above (www.rcsb.org), (3) Lipoate-protein ligase A from *Streptococcus agalactiae* (PDB Id: 2P0L) (www.rcsb.org), (4) P450 pyrhydroxylase from *Sphingopyxis macrogoltabida* (PDB Id : 3RWL),[5] (5) Hydroxymethylbilane synthase from *Escherichia coli* (K12) (PDB Id : 2YPN),[6] (6) UDP-n-acetylmuramoyl-L-alanine:Dglutamate ligase from *Escherichia coli* (K12) (PDB Id: 1UAG),[7] (7) Glycinamide ribonucleotide synthetase from *Escherichia coli* (K12) (PDB Id: 1GSO)[8] (8) Folypolyglutamate synthetase from *Lactobacillus casei* (PDB Id : 1FGS)[10] and (9) mitochondrial helicase suv3 from *Homo sapiens* (PDB Id : 3RC3).[11]

The amino acid sequences in two formats were aligned and residues missing at the N-terminal, C- terminal and internal regions were detected. Sequences of 9 proteins were folded with JPred 4 (http://www.compbio.dundee.ac.uk/jpred4) and PSSPred.[12-13] From the output we designated residues forming secondary structure domains in different shades, namely light gray (α helix), dark gray (β sheet/loop) and medium gray (disordered domain). The sequences derived from the crystals (pdb format) were similarly shaded.

Results

The sequences in both formats of nine proteins are shown in Figure 1. We noticed that, in contrast to the sequence derived from FASTA file, some amino acids were missing at the termini as well as at internal locations of the polypeptide in the crystal-derived sequences. Upon folding these *in silico* with Jpred4, we find (Figure 1) that each polypeptide gave rise to lawns exhibiting α-helix, β sheet, and disordered domains/random coil (methods). Since the folding pattern with respect to the number and positions of different structural domains was nearly similar with PPSPred, we have restricted this presentation to JPred4 for proteins no 1-9.

Table 1 shows the number of amino acid residues missing in crystal-derived sequences. 2P0L, 3RWL and 3RC3 also exhibit long missing oligopeptides at the termini. Indeed, all crystal-derived sequences contain one or more 3-33 long internal missing oligo-peptides. Table 2 describes the distribution of missing residues in crystals based on their physicochemical properties and number. These were highlighted in sequences from mature protein (FASTA file) in Figure 1. We find that, in 10 cases, more hydrophilic residues are missing in the internal oligopeptide. In the rest 6, the ratio of hydrophobic residues to total number of missing residues is more than 0.5.

(A) 1A48 (Jpred4)

Mature protein

XSITKTELDGILPLVARGKVRDIYEVDAGTLLFVATDRISAYDVIMENSIPEKGILLTKLSEFWFKFLSND
VRNHLVDIAPGKTIFDYLPAKLSEPKYKTQLEDRSLLVHKHKLIPLEVIVRGYITGSAWKEYVKTGTVH
GLKQPQGLKESQEFPEPIFTPSTKAEQGEHDENISPAQAAELVGEDLSRRVAELAVKLYSKCKDYAKEK
GIIIADTKFEFGIDEKTNEIILVDEVLTPDSSRFWNGASYKVGESQDSYDKQFLRDWLTANKLNGVNGV
KMPQDIVDRTRAKYIEAYETLTGSKWSH

Mature protein folded with JPred4

Crystal derived sequence

01 X

SITKTELDGIL PLVARG KVRDIYEVD AG TLLFVAT DR SA YD VIM ENSIPEKGILLTKLSEFWFKFLSNDV RN HLV DI 80

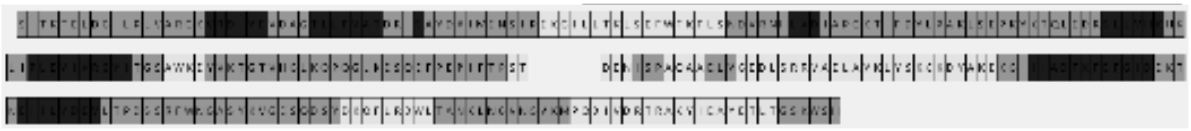
APGKTIFDYL PAKLSEPKYKTQLED RSLLVHK H KLIPLEVIVRGY TGSAWKEYVKT GTV HG LKQ PQGLKE SQEFPEP 157

163 KAEQGEH 170

FTPST D EN SPAQAAELVGEDLSRRVAELAVKLYSKCKDYAKEK GIIIADTKFEFGID EKT NEIILV DEVLTPDSS RF 243

VNG ASYKVGESQDSY DKQFLRDWLTANK LNGVNGVKM PQDIVDRTRAKYIEAYETLT GSKWSH 306

Crystal derived sequence folded with JPred4

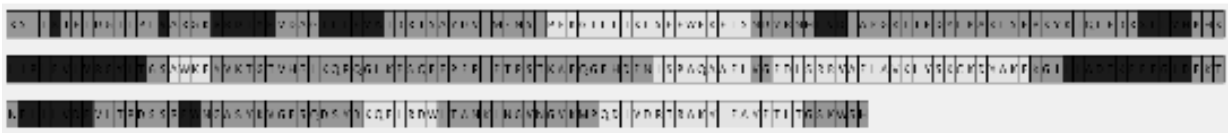


(B) 2CNQ (Suicar synthase complexed with ADP, AICAR, succinate)

Mature protein

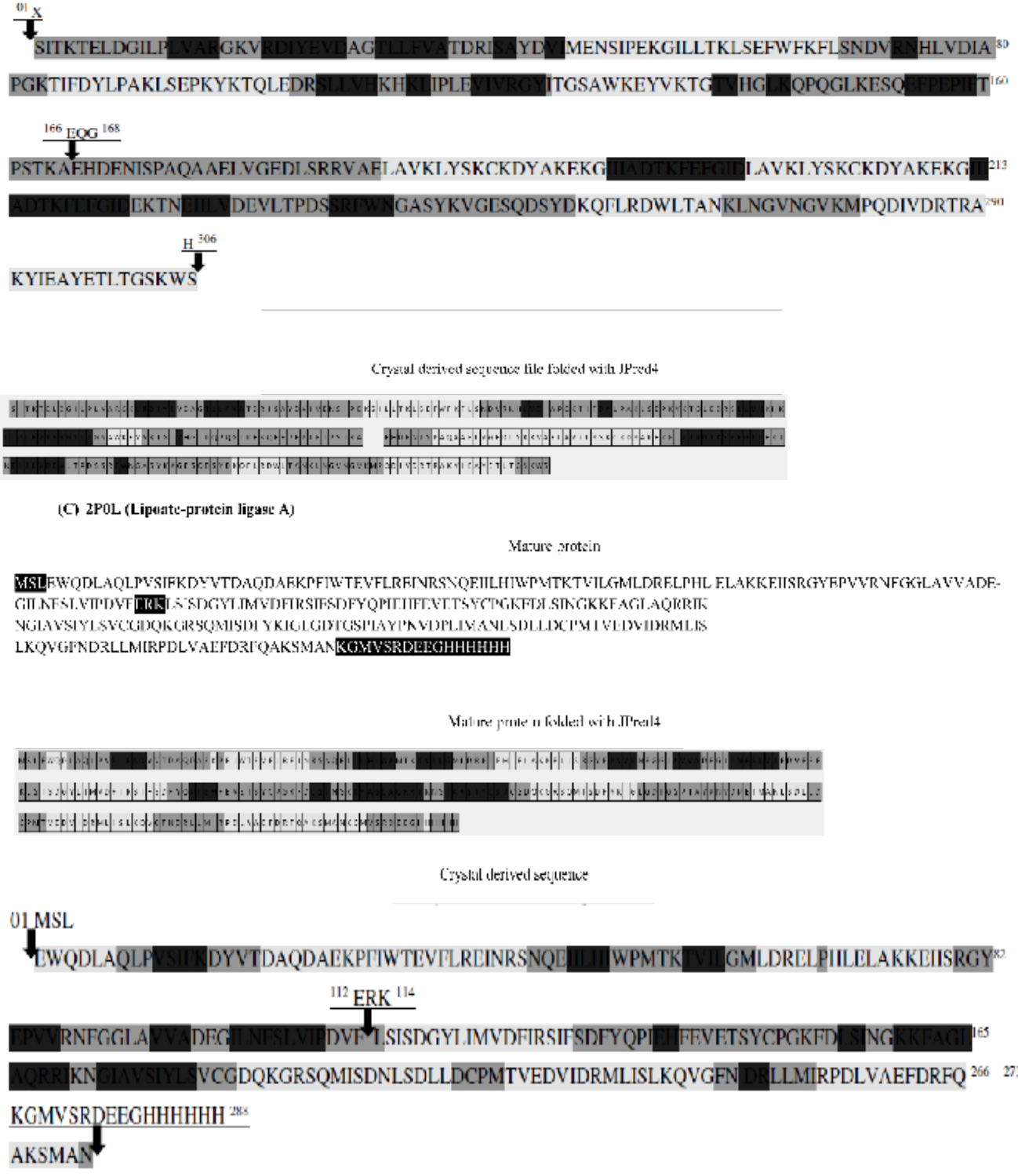
XSITKTELDGILPLVARGKVRDIYEVDAGTLLFVATDRISAYDVIMENSIPEKGILLTKLSEFWFKFLSND
VRNHLVDIAPGKTIFDYLPAKLSEPKYKTQLEDRSLLVHKHKLIPLEVIVRGYITGSAWKEYVKTGTVH
GLKQPQGLKESQEFPEPIFTPSTKAEQGEHDENISPAQAAELVGEDLSRRVAELAVKLYSKCKDYAKEK
KMPQDIVDRTRAKYIEAYETLTGSKWSH

Mature protein folded with JPred4



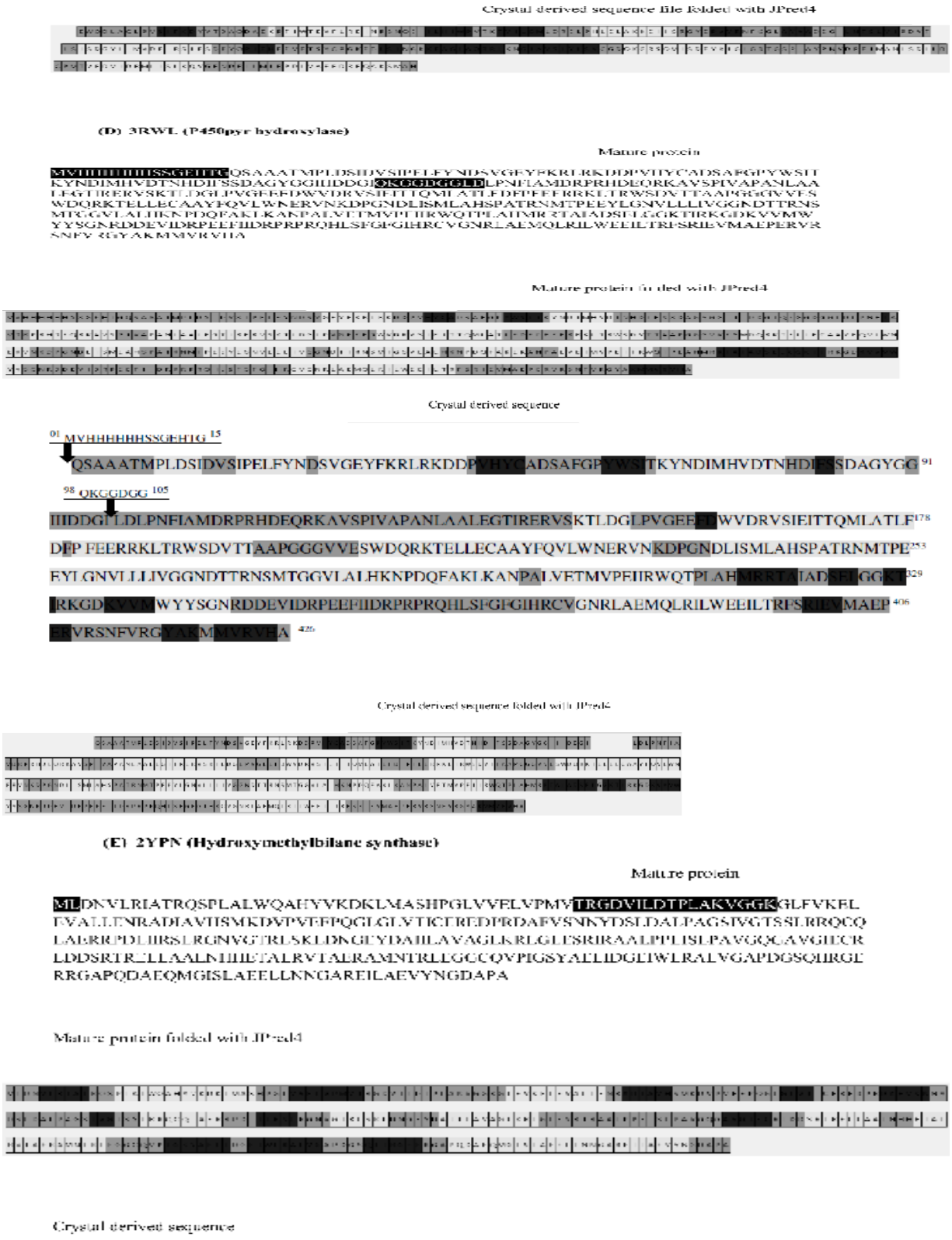
Crystal derived sequence

(Figure 1 continues)

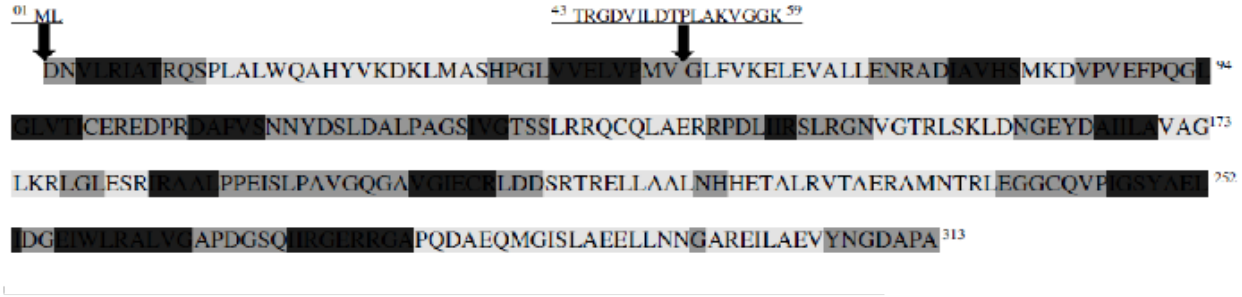


Citation: Kelkar N, Modak SP, The internal oligopeptide sequences missing in crystals are disordererd domains. *MOJ Proteomics Bioinform.* 2018;7(1): 00215. DOI: 10.15406/mojpb.2018.07.00215

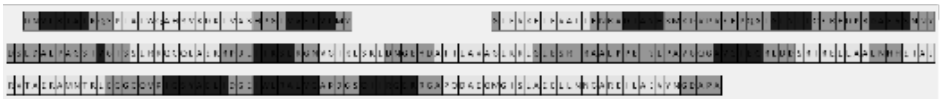
(Figure 1 continues)



(Figure 1 continues)



Crystal derived sequence folded in JPred4

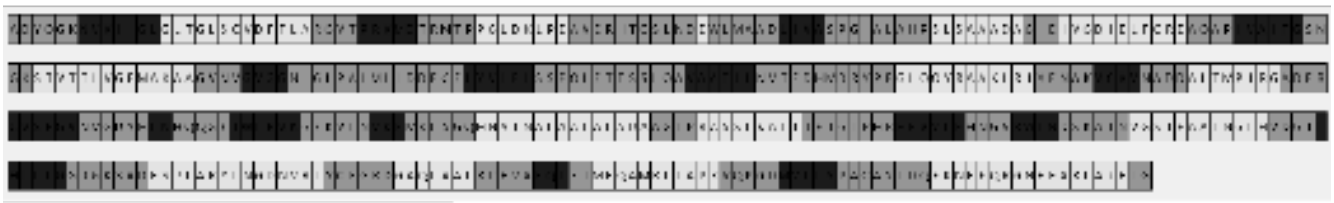


(F) 1UAG (UDP-N-acetylmuramoyl-L-alanine:D-glutmate ligase)

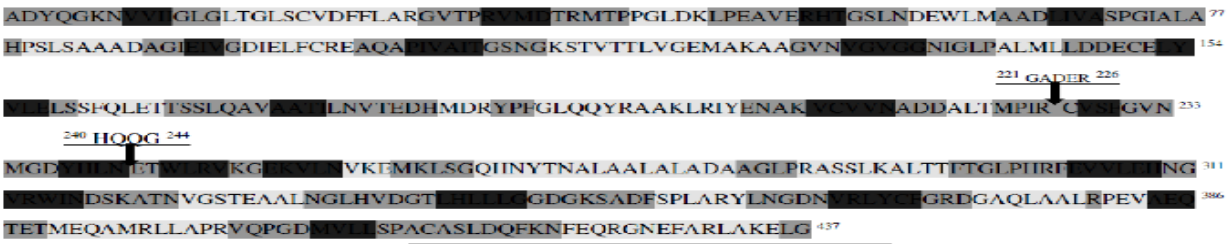
Mature protein

ADYQGKNVVTIGLGLTGLSCVDFFLARGVTPRVMDTRMTPPGLDKLPEAVERHTGSLNDEWLMAADL
IVASPGIALAHPSLSAAADAGIEIVGDIELFCREAQAPIVAITGSNGKSTVTTLVGEMAKAAGVNVGVGG
NIGLPALMLLDDECELYVLELSSFQLETTSSLQAVAATILNVTEDHMDRYPFGLQQYRAAKLRIYENAK
VCVVNADDALTMPIRGADER CVSFGVNMGDYHLNHQQG ETWLRVKGEKVLNVKEMKLSGQHNYTN
ALAAIALADAAGLPRASSLKALTTFTGLPHRFEVVLEHNGVRWINDSKATNVGSTEAALNGLHVDGT
LHLLLGGDGKSADFSPLARYLNGDNVRLYCFGRDGAQLAALRPEVALQTETMLQAMRLLAPRVQPG
DMVLLSPACASLDQFKNFEQRGNEFARLAKELG

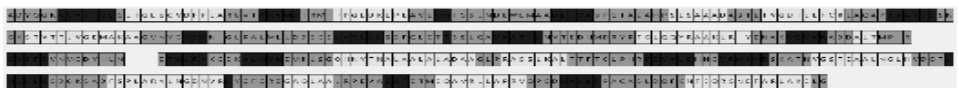
Mature protein folded with JPred4



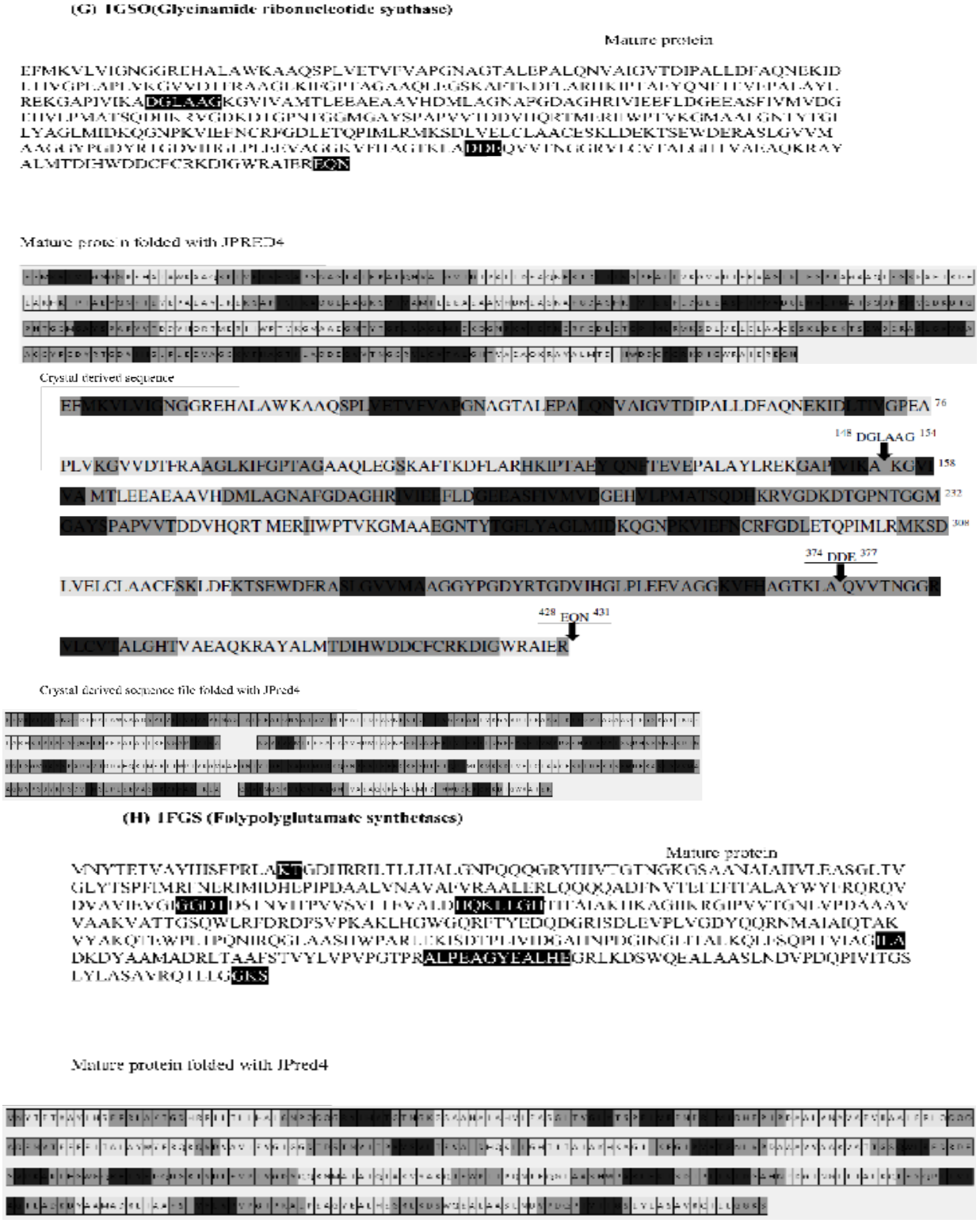
Crystal derived sequence



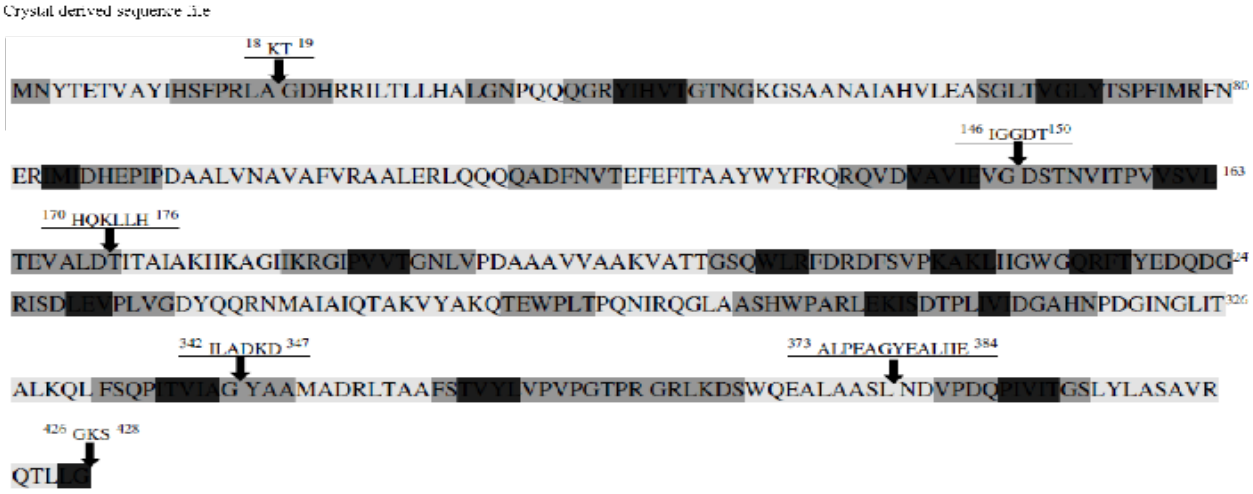
Crystal derived sequence folded with JPred4



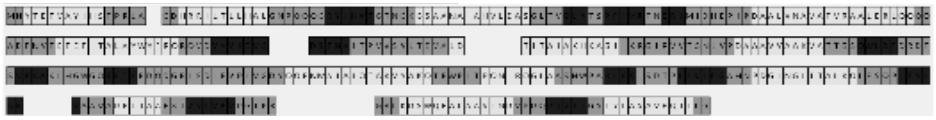
(Figure 1 continues)



(Figure 1 continues)



Crystal derived sequence 3.e folded with JPred4

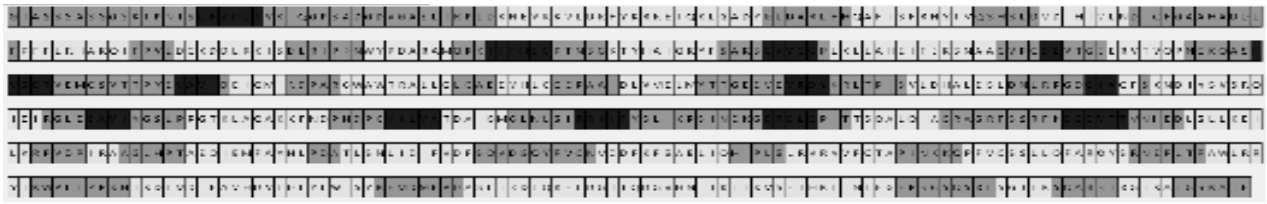


(I) 3RC3 (Human Mitochondrial Helicase Suv)

Mature protein

GTASSSASGGSKIPNTSLFVPLTVKPQGPSADKDVGIAILTRPLDKNFVKKVLDKFYKRKFIQKLGADY
GLDARLFHQAFISFRNYIMQSHSLDVDIHIVLNDICFGAAHADDLFPFFLRHAKQIFPVLDCKDDLRKIS
DLRIIPNWYYPDARAMQRKIIFISGPTNSGKTYHAIQKYFSAKSGVYCGPLKLLAHEIFKSNAAGVPCD
LVTGEERVTVQPNGKQASHVSCTVEMCSVTTPYEVAVIDEIQMIRDPARGWAWTRALLGLCAEEVHL
CGEPAAIDLVMELMYTTGETVEVRDYKRLTPISVLDHALESLDNLRPGDCIVCFSKNDIYSVSRQIEIRG
LESAVIYGSLPPGTKLAQAKKFNDPNDPCKILVATDAIGMGLNLSIRRIIFYSLIKPSINEKGERELLPITTS
QAIQIAGRAGRFSSRFKFGFVTTMNHEDLSLLKELLKRPVDIPIRAAGLHPTLAEQIEMFAYHLPDATLSN
LIDIFVDFSQVDGQYFVCNMDDFKFSAELIQHIPLSLRVRYVFCTAPINKKQPFVCSSLLQFARQYSRNF
PLTFAWLRRYIKWPLLPPKNIKDLMDLEAVHDVLDLYLWLSYREMDVFPDASLIRDLQKFLDGHIQDG
VHNITKLIKMSETHKLLNLEGTPSGSQSRLSGTLKSQARRTRGTKALGSKATL

Mature protein folded with JPred4



Crystal derived sequence

(Figure 1 continues)

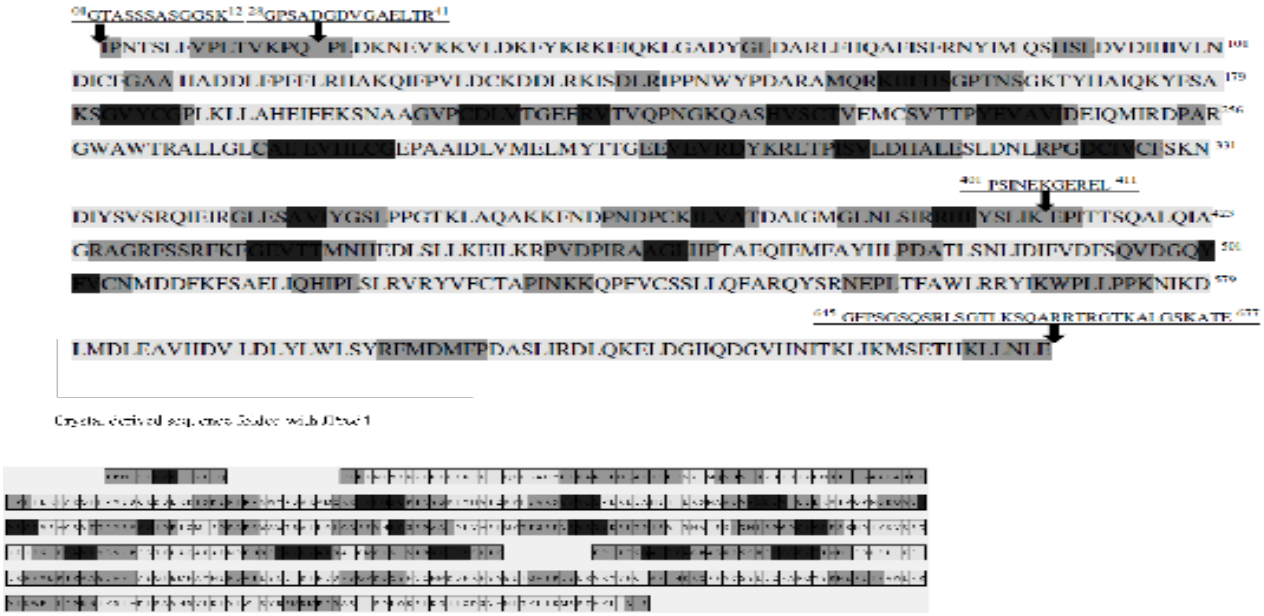


Figure 1 Panorama of secondary structures from mature protein sequence (FASTA) and crystal structure sequence folded with JPred4: The different secondary domains are coloured according to gray scale. The crystal derived sequence is also colour coded according to secondary structures in grey scale and the arrows indicate the missing region. The missing oligopeptide region in crystal derived sequence in highlighted in the FASTA sequence.

Table 1 Missing oligopeptide in the crystal structure derived sequence

Protein	PDB Id	Missing residues in crystal derived sequence		
		N terminal	C terminal	internal strings
Saicar synthase	1A48	1	0	7
Saicar synthase	2P0L	3	16	3
Lipoate-protein ligase A	3RWL	15	0	7
P450 pyr hydroxylase	2YPN	2	0	17
Hydroxymethylbilane synthase	2CNQ	1	0	3
UDP-n-acetylmuramoyl- Lalanine: D-glutamate ligase	1UAG	0	0	5, 4
Glycinamide ribonucleotide synthetase	1GSO	0	0	6, 3
Folypolyglutamate synthetase	1FGS	0	0	32, 5, 7, 6, 12
Mitochondrial helicase suv3	3RC3	12	0	14, 11, 33

The secondary structures predicted for internal missing oligopeptide and their flanking tripeptides from both FASTA and PyMol (crystal) formats are shown in Table 3. Surprisingly, 10 out of 16 internal missing oligopeptides form the disordered domains (DD). Among the remaining 6, two disordered domains adjoin terminal residue from β sheet, one adjoins α helix and 3 are from putative helix. In the tripeptides flanking the internal missing stings, we find that, at N-terminal, 10 out of 16 forms IDD, 3 form beta sheets 2 are from α helix and 1 form a junction between beta sheet and random coil. In the C terminal tripeptide, 7 are from disordered domain, 3 form a junction between disorder domain and α helix, 2 β sheet- DD junctions and 2 each from β sheet and α- helix. Thus, clearly, all missing strings are part of original disordered domains

Table 2 Distribution of missing residues in crystal structure

Protein	PDB Id	Missing oligopeptide	Proline residues	glycine residues	charged residue	Polar uncharged	hydro-phobic	Total amino acid
Saicar synthase	1A48	KAEQGEH	0	1	4	1	2	7
Saicar synthase	2CNQ	EQG	0	1	1	1	1	3
Lipoate-protein ligase A	2P0L	ERK	0	0	3	0	0	3
P450 pyr hydroxylase	3RWL	QKGGDGG	0	4	2	1	4	7
Hydroxymethylbilane synthase	2YPN	TRGDVILDTPLAKVGGK	1	3	5	2	10	17
UDP-n-acetylmuramoyl-L-alanine-D-glutamate ligase	1UAG	GADER	0	1	3	0	2	5
"	"	HQQG	0	1	1	2	1	4
Glycinamide ribonucleotide synthetase	1GSO	DGLAAG	0	2	1	0	5	6
"	"	DDE	0	0	3	0	0	3
Folypolyglutamate synthetase	1FGS	KT	0	0	1	1	0	2
"	"	IGGDT	0	2	1	1	3	5
"	"	HQKLLGH	0	1	3	1	3	7
"	"	ILADKD	0	0	3	0	3	7
"	"	ALPEAGYEALHE	1	1	4	0	7	12
Mitochondrial Helicase suv 3	3RC3	GPSADGDVGAELTR	0	3	4	2	8	14
"	"	PSINEKGEREL	1	1	5	2	4	11

Note: only one aromatic residue (tyrosine) was seen in the internal missing oligopeptide string in IFGS.

Citation: Kelkar N, Modak SP, The internal oligopeptide sequences missing in crystals are disordererd domains. *MOJ Proteomics Bioinform.* 2018;7(1): 00215. DOI: 10.15406/mojpb.2018.07.00215

Table 3 Predicted secondary structure of the internal missing oligopeptide and the flanking residues

Protein name	Protein Id	Missing oligopeptide	Secondary structure of residues after folding FASTA sequence in JPred4		
			missing	Tripeptide flanking the missing oligopeptide region	
				N terminal	C terminal
SAICAR synthase	IA48	KAEQGEH	random coil	random coil	random coil
SAICAR synthase	2CNQ	EQG	random coil	random coil	random coil
Lipoate-protein ligase A	2P0L	ERK	random coil	random coil	random coil and α helix
P450 pyr hydroxylase	3RWL	QKGGDGG	random coil	random coil	random coil
Hydrox-ymethylbilane svnthase	2YPN	TRG DVILDTPLAKVGGK	β sheet and random coil	β sheet	random coil and α helix
UDP-n-acetylmuramoyl L-alanine D-glutamate ligase	1UAG	GAD ER	β sheet and random coil	α helix	β sheet
"	"	HQQG	random coil	β sheet	β sheet
Glycinamide ribonucleotide synthetase	1GSO	DGLAAG	random coil	β sheet and random c oil	random coil
"	"	DDE	random coil	random coil	random coil and β sheet
Folypolyglutamate synthetase	IFGS	KT	random coil	random coil	random coil and α helix
"	"	IGGDT	α helix and random coil	α helix	random coil
"	"	HQKLLGH	α helix and random coil	random coil	α helix
"	,,	ILADKD	random coil	β sheet	α helix
"	"	ALPEAGYEALHE	α helix and random coil	random coil	random coil
Mitochondrial Helicase suv 3	3RC3	GPSADGDVGAELTR	random coil	random coil	random coil
"	"	PSINEKGEREL	β sheet and random coil	random coil	β sheet and random coil

Comparing panoramas of secondary structures derived from the crystal structure to those computed by folding sequences from both, mature proteins and crystals with JPred4, we find (Table 4) that for each type of secondary structure crystals give an underestimate of the number of disordered domains as well as the number of residues therein. Indeed, a combined analysis of 9 proteins reveal the ratio (number of secondary structure domains: number of amino acid residues) is comparable for α helices and β sheets, but substantially reduced for disordered domains in crystals than *in silico* folded mature protein. Similar results were obtained by folding with PSSpred (not shown).

We find that only 2 crystals reveal histidine-rich oligopeptides at N or C terminal and none in the internal missing oligopeptide strings (data not shown).

Citation: Kelkar N, Modak SP, The internal oligopeptide sequences missing in crystals are disordererd domains. *MOJ Proteomics Bioinform.* 2018;7(1): 00215. DOI: 10.15406/mojpb.2018.07.00215

Table 4 Secondary structure from the mature protein sequence and crystal structure derived sequence folded using Jpred4 and the original crystal sequence

Protein name	PDB ID	Source	Alpha Helix [no. of motifs (no. of amino acids)]			Beta sheets [no. of motiffs(no. of amino acids)]			Random coils [no. of motiffs(no. of amino acids)]		
			mature protein	crystal structure		mature protein	crystal structure		mature protein	crystal structure	
			JPred4	derived	JPred4	Jpred	derived	JPred4	JPred4	derived	JPred4
Saicar synthase	1A48	*Saccharomyces cerevisiae* ATCC 204508	6 (82)	7 (116)	7(85)	10 (56)	15 (106)	8 (53)	17(168)	20 (84)	16(160)
Saicar synthase	2CNQ	*Saccharomyces cerevisiae* ATCC 204508	6 (82)	6 (127)	5(82)	10 (56)	15 (69)	10 (57)	17(168)	17 (106)	16(163)
Lipoate protein ligase A	2P0L	*Sphingopyxisma-crogoltabida*	10 (105)	9 (118)	8(93)	10 (57)	11 (55)	10 (59)	21(126)	20 (93)	19(114)
P450 pyr Hydroxylase	3RWL	*Sphingopyxis macrogoltabida*	14 (76)	14 (234)	14(177)	6 (36)	12 (40)	6 (34)	19(214)	24 (130)	20 (193)
Hydroxymethylbilane synthase	2YPN	*Escherichia coli* K12	8 (112)	11 (112)	8 (113)	11 (69)	13 (76)	11(62)	20(132)	21 (106)	20(119)
UDP-n-acetylmuramoyl L-alanine D-glutamate ligase	1UAG	*Escherichia coli* K12	15 (161)	20 (161)	20(152)	17 (83)	20 (89)	20(88)	32(193)	38 (178)	33(188)
Glycinamide ribonucleotide synthetase	1GSO	*Escherichia coli* K12	11 (127)	16 (128)	12(130)	20 (97)	16 (99)	12(99)	32(207)	33 (192)	32(190)
Folypolyglutamate synthetase	1FGS	*Lactobacillus casei*	16 (192)	15 (172)	13(166)	14 (67)	16 (62)	13(70)	31(169)	29 (159)	27(157)
Mitochondrial helicase suv3	3RC3	*Homo sapiens*	31 (441)	26 (444)	26(366)	13 (60)	16 (69)	14(65)	43(176)	42 (164)	37(246)

Table 2 lists the relative distribution of Proline, Glycine, charged and hydrophobic residues in the internal missing strings. Thus, there is a high concentration of flexible (Glycine, 20/107) and charged residues (43/107) in these strings, while rigid Proline is of rare occurrence (3/107). Similarly, there is only 1 aromatic residue in the missing strings (not shown).

Discussion

According to Djinovic-Carrugo & Carrugo,[1] most crystallographic data reveal incidence of internal missing strings of oligopeptides. Here we describe in detail 9 such strings and analyses of their position in the overall panorama of secondary structure domains of a polypeptide sequence. To study this aspect we have adopted the strategy of folding *in silico* sequences for the same protein representing the post-translationally processed polypeptide and that derived from the crystal. The issue here is that when the crystal structure is obtained at low resolution, a number of residues fail to be detected due to low electron density. Therefore, by comparing two amino acid sequences of the same protein, we find the missing residues missing in crystal-derived sequence.

Comparison of the panorama of secondary structure domains revealed that after folding the sequences *in silico*, allowed us to detect secondary structure domains to which the missing residue belong and we conclude that most are disordered domains. This is further supported by the fact that these are rich in flexible amino acid Glycine and poor in rigid Proline. We find that out of the 16 cases, the proportion of hydrophobic residue is less than 0.5 in 10 cases and in remaining, it is less than 0.6. These disordered oligopeptides contain high concentration of charged residues and nearly 20% glycines. We conclude that the apparent loss or delectability in crystals of large internal oligopeptide strings involve highly disordered domains which probably accounts for the difficulty crystallographers face in designating a signature domain to the missing internal string. In fact, since these strings are not actually absent in polypeptides, the inability to detect leads to an incomplete crystal structure. Clearly, in most cases the problem can be solved by comparing the *in silico* folded amino acid sequences of mature proteins to those derived from crystals.

Finally, one must consider the structural and functional relevance of the apparently missing segments. To that effect we are now assessing the propensity of various Triads involved in defining

functionally important sites for enzyme-substrate interactions as well as other protein: protein binding. Another possible approach is to examine the missing oligopeptides alone and with flanking regions in Ramachandran plots. The question, therefore, remains as to how one should solve the crystal structure beyond the offerings of crystallography. In any case, it is unlikely that proteins exist in crystalline form in vivo and probably exhibit a metastable state with variable mobility of flexible regions depending on the intracellular environment.

Indeed, polymeric structures, namely, micelles, membranes and globular proteins exhibit a hydrophobic core and hydrophilic exterior that are differentially sensitive to perturbations by osmotic pressure, ionic strength and temperature and exhibit differential movements such that the kinetic energy between the two domains is conserved.[9]

Acknowledgments

We are grateful to Prof. C. Ramakrishnan, Prof. T. Ramasarma, Prof, N,V, Joshi and Prof. V. Sitaramam for critical comments and suggestions. We also thank Dr. Vaijayanti Tamhane and Prof. Ameeta Ravi Kumar, Head, IBB, for encouragement and support.

Conflict of interest

The author declares that there is no conflict of interest regarding the publication of this article.

References

1. Djinovic–Carugo K, Carugo O. Missing strings of residues in protein crystal structures. *Intrinsically Disord Proteins*. 2015;3(1):e1095697.
2. Modak SP, Milner Kumar M, Bargaje. *Molecular Phylogenetic Trees: Topology of multiparametric poly-genic/phenic Tree exhibit Higher Taxonomic Fidelity than Uniparametric Trees for Mono–Genic/Phenic traits in Evolutionary Biology: Mechanisms and trends*. Springer verlag; 2012. p. 79–102.
3. www.rcsb.org
4. Levdikov VM, Barynin VV, Grebenko AI, et al. The structure of SAICAR synthase: an enzyme in the de novo pathway of purine nucleotide biosynthesis. *Structure.* 1983;6(3):363–376.
5. Pham SQ, Pompidor G, Liu J, et al. Evolving P450pyr hydroxylase for highly enantioselective hydroxylation at non-activated carbon atom. *Chemical Communications*. 2012;48(38):4618–4620.
6. Nieh YP, Raftery J Weisgerber S, Habash J, et al. Accurate and highly complete synchrotron protein crystal Laue diffraction data using the ESRF CCD and the Daresbury Laue software. *Journal of Synchrotron Radiation*. 1999;6:995–1006.
7. Bertrand JA, Auger G, Fanchon E, et al. Crystal structure of UDP-N-acetylmuramoyl-L-alanine: D-glutamate ligase from Escherichia coli. *EMBO J*. 1997;16(12):3416–3425.
8. Wang Weiru T, JosephKappock, JoAnne Stubbe, et al. "X-ray crystal structure of glycinamideribonucleotide synthetase from Escherichia coli." Biochemistry.1998;37(45):15647-15662.
9. Madhavarao CN, Sauna ZE, Srivastava A, et al. Osmotic perturbations induce differential movements in the core and periphery of proteins, membranes and micelles. *Biophys Chem* 2011;90(3):233–248.
10. Sun X, Bognar AL, Baker EN, et al. Structural homologies with ATP-and folatebinding enzymes in the crystal structure of folylpolyglutamatesynthetase. *Proc Natl Acad Sci U S A*. 1998;95(12): 6647–6652.
11. Jedrzejczak R, Wang J, Dauter M, et al. Human Suv3 protein reveals unique features among SF2 helicases. *Acta Crystallographica Section D: Biological Crystallography*. 2011;67(11):988–996.
12. http://zhanglab.ccmb.med.umich.edu/PSSpred/
13. Drozdetskiy A, Cole C, Procter J, et al. JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 2015;43(W1):W389–94.
14. Dunker AK, Babu MM, Barbar E, et al. What's in a name? Why these proteins are intrinsically disordered: Why these proteins are intrinsically disordered. Intrinsically Disordered Proteins. 2013;(1):e24157.
15. http://www.compbio.dundee.ac.uk/jpred/