Identification of an Ideal-like Fingerprint for a Protein Fold using Overlapped Conserved Residues based Approach

Amit Goyal, Sriram Sokalingam, Kyu-Suk Hwang*, Sun-Gu Lee*

School of Chemical and Biomolecular Engineering, Pusan National University, Busan,
Korea 609-735

Method:

Extraction of OCR-based pattern for Green Fluorescent Proteins

Overlapped conserved residues based on the three aspects of residue conservation, i.e. sequence, structure, and intramolecular interaction were depicted using the general multiple sequence alignment (MSA), structure based alignment (SBA) and supersecondary structure (SSS) -based alignment methods were employed, respectively. The three approaches were implemented independently to the target protein folds and the overlapped residues, among the conserved residues achieved by each method, were chosen. For this, 10 representative Green Fluorescent protein structures, as listed in Supplementary Table S6, were selected. Now, three independent alignment MSA, SBA and SSS were obtained for the representative sequence/structure. Supplementary Table S7, S8 and S9 comprise the alignment for MAS, SBA and SSS, respectively. Alignment using the MSA, SBA and SSS retrieved 55, 44 and 46 conserved positions respectively which are listed below.

MSA: 55 conserved positions

13P, 18L, 22V, 23N, 24G, 28S, 29V, 33G, 35G, 40G, 41K, 43T, 45K, 46F, 50T, 51G, 53L, 55V, 60L, 67G, 71F, 73R, 77H, 79K, 82D, 90E, 91G, 92Y, 101K, 102D, 103D, 106Y, 109R, 111E, 113K, 114F, 117D, 119L, 122R, 124E, 125L, 129D, 131K, 134G, 136I, 138G, 140K, 149N, 153M, 158K, 161I, 170N, 173D, 174G, 180D, 185N, 200Y

MSA-based Sequence pattern

[PGKTRE]-x(4)-[LCMI]-x(3)-[YVL]-[NTG]-[GD]-x(3)-[SQEKT]-[VCLMI]-x(3)-G-x-[GPS]-x(2,4)-G-[KRTDHE]-x-[TKEN]-x-[KREH]-[AFILM]-x(3)-[TSDKE]-[GSK]-x(0,2)-[LV]-x-[VFM]-x(4,10)-[LVI]-x(2,6)-[GS]-x(3)-[FIY]-x-[RTSKG]-x(3)-[HTGED]-x-[KGPSE]-x(0,2)-[DSNH]-x(7)-[EGN]-G-x(0,1)-[YIFL]-x(8)-[KDEQ]-x(0,8)-[DTN]-[GD]-x(2)-[YFLMACV]-

x(2)-[RQNHTSK]-x-[ERDTS]-x-[KTSER]-[FYLI]-x(2)-[DHNTG]-x(1,3)-[LIVF]-x(2)-[RKDNH]-x-[EQKTD]-[LFCIV]-x(3)-[GDN]-x-[HKPQR]-x(2,10)-[GS]-x-[IV]-x-[GQKRT]-x-[KPTQD]-x(6,8)-[NPTGS]-x(3)-[MLV]-x(4)-[KSGPTD]-x(1,2)-[ILV]-x(8)-[NTQKS]-x(2)-[DEGT]-[GP]-x(5,8)-[DHNTS]-x(4)-[NQK]-x(9,14)-[YAF]

SBA: 44 conserved positions

10G, 13P, 18L, 19S, 20G, 22V, 24G, 27F, 28S, 33G, 40G, 41K, 46F, 60L, 71F, 77H, 82D, 89P, 91G, 92Y, 95E, 100F, 102D, 103D, 106Y, 109R, 111E, 113K, 117D, 119L, 122R, 124E, 125L, 139H, 144N, 152I, 159N, 161I, 170N, 173D, 174G, 175S, 183Q, 218M

SBA-based Sequence Pattern

x(0,8)-[GSTEP]-x(2)-[PRKTE]-x(4)-[LVIM]-[DKHETS]-G-x-[VIL]-x-[GPD]-x(2)-F-[SEKT]-x(4)-[GS]-x(6)-G-[KRTDHE]-x(4)-[FIMAL]-x(11,13)-[LIV]-x(3,7)-[FY]-x(5)-[HGED]-x(1,4)-[DKNH]-x(6)-[PGTN]-x(0,1)-[GE]-[YFL]-x(2)-[EQTD]-x(4)-[FY]-x-[DGN]-[DGH]-x(2)-[YLMAVC]-x(2)-[RKNHTS]-x-[EQRDTS]-x-[KSTER]-x(3)-[DGNT]-x-[LIVF]-x(2)-[RSDKNH]-x-[EKTD]-[LCVIF]-x(6,13)-[HPDNKQ]-x(2,4)-[NHKRDG]-x(3)-[IYMVL]-x(4,6)-[NSGT]-x-[IVL]-x(8)-[NTQKS]-x(2)-[DPGT]-[GS]-[SHRGK]-x(7)-[QNTS]-x(24,34)-[MLVY]-x(11)

SSS: 46 SSS-determining residues

13P, 18L, 19D, 20G, 22V, 23N, 24G, 28S, 33G, 40G, 41K, 46F, 50T, 55V, 64F, 66Y, 67G, 71F, 77H, 79K, 90E, 91G, 92Y, 95E, 100F, 102D, 103D, 104G, 106Y, 109R, 111E, 113K, 122R, 124E, 125L, 137L, 138G, 140K, 144N, 152I, 157Q, 161I, 170N, 180D, 191G, 196P

SSS-based Sequence Pattern

[PRKTE]-x(4)-[LVIM]-[DKHETS]-G-x-[LVI]-x(0,2)-[NG]-[GSD]-x(3,8)-[SNEKT]-x(4)-[GS]-x(4,6)-G-[KRTDH]-x(4)-[FIMAL]-x(1,4)-[TDG]-x(2,4)-[VFML]-x(7,10)-[FAMCI]-x-[YI]-G-x(3,6)-[FY]-x(5)-[HGED]-x-[KPS]-x(5,10)-[ETN]-G-x(0,2)-[YFL]-x(2)-[EQTD]-x(4)-[FY]-x(0,2)-[DHENQ]-[DPG]-G-x(1,2)-[YMLAVC]-x(2)-[RKNHTS]-x[EQRTDS]-x-[KSTER]-x(8,13)-[RSDNHK]-x-[EKTD]-[LCIVF]-x(6,11)-[LIFM]-[GPKQRT]-x-[KGTQD]-x(2,3)-[NHKRDGS]-x(7)-[IYMVL]-x(3,4)-[QSGT]-x(1,3)-[IVL]-x(8)-[NTQKS]-x(7,17)-[DQEHSR]-x(8,15)-[GRNKH]-x(4,8)-[PS]

Now, the amino acid positions found to be commonly conserved among the three different alignments were identified. The residues were called "Overlapped Conserved Residues" (OCR) and used to create the OCR fingerprint for the fold detection process. Sequence pattern can be generated using the overlapped positions from any of the sequence alignment.

OCR: 24 overlapped conserved residues

13P, 18L, 22V, 24G, 28S, 33G, 40G, 41K, 46F, 71F, 77H, 91G, 92Y, 102D, 103D, 106Y, 109R, 111E, 113K, 122R, 124E, 125L, 161I, 170N

OCR fingerprnt

[PRKTE]-x(4)-[LVIM]-x(3)-[LVI]-x(1,3)-[GSD]-x(3,8)-[SNEKT]-x(4)-[GS]-x(4,6)-G-[KRTDH]-x(4)-[FIMAL]-x(24,25)-[FY]-x(5)-[HGED]-x(8,13)-G-x(0,2)-[YFL]-x(8,10)-[DHENQ]-[DPG]-x(2,3)-[YMLAVC]-x(2)-[RKNHTS]-x-[EQRTDS]-x-[KSTER]-x(8,13)-[RSDNHK]-x-[EKTD]-[LCIVF]-x(28,35)-[IVL]-x(8)-[NTQKS]

Further, the OCR^S embedded in the beta strand region were identified and were used to generate the OCR^S fingerprint.

OCR^S: 17 overlapped conserved residues

13P, 18L, 22V,28S, 33G, 41K, 46F, 92Y, 106Y, 109R, 111E, 113K, 122R, 124E, 125L, 161I, 170N

OCR^S fingerprint

[PRKTE]-x(4)-[LVIM]-x(3)-[LVI]-x(5,12)-[SNEKT]-x(4)-[GS]-x(5,7)-[KRTDH]-x(4)-[FIMAL]-x(40,45)-[YFL]-x(13,14)-[YMLAVC]-x(2)-[RKNHTS]-x-[EQRTDS]-x-[KSTER]-x(8,13)-[RSDNHK]-x-[EKTD]-[LCIVF]-x(28,35)-[IVL]-x(8)-[NTQKS]

In addition, minimization of OCR^{S} fingerprint, by eliminating the conserved positions individually, produced the OCR^{MIN} fingerprint. The OCR-based fingerprints, such as OCR, OCR^{S}, and OCR^{MIN}, were used to detect the homologous proteins of a target fold, and their fold detection efficiencies were compared with those of the fingerprints obtained by MSA, SBA and SSS.

Figure S1: Conserved sequence residues obtained by MSA, SBA and SSS methods. Protein sequence patterns, as shown in figure from **A**-**L** were obtained using MSA, SBA, SSS and OCR-based approach for 12 protein folds in target dataset. The figure shows the fingerprints for following protein folds: **A** GFP-like protein, **B** Cupredoxin-like proteins, **C** Acid Proteases, **D** Ribosomal Protein L14, **E** β-Grasp (Ubiquitin-like) protein, **F** Nucleoside Triphosphate Hydrolase, **G** RNAase A-like proteins, **H** Cysteine Proteinases, **I** 50S Ribosomal Protein L25, **J** 50S Ribosomal Protein L6P, **K** Difydrofolate reductase-like proteins, **L** Pyruvate kinase N-terminal domain. Here, Distribution of conserved positions in secondary structure elements (SSEs) is shown for each alignment method. Sequence pattern is PROSITE-like pattern similar to Figure 2.

	Strand1	Loop	Strand2	Loop	Strand3
MSA	x(1,2)-[PRKTH]-x(4)-[LVIM]-[DKHETS]-[G]-x-[LVI]-x(0,2)	[NG]-[GSDN]-x(0,4)	x(3,4)-[SNEKT]-x(4)-[GS]-x(2,3)	x(1,4)-G	[KRTDHE]-x(4)-[FIMAL]-x(1,4)
SBA	x-[PGKTRE]-x(4)-[LCMI]-x(3)-[YVL]	[NTG]-[GD]	x(3)-[SQEKT]-[VCLMI]-x(3)-G-x-[GPS]-x(0,1)	x(2,3)-G	[KRTDHE]-x-[TKEN]-x-[KREH]-[AFILM]-x(2)
SSS	x-[PRKTE]-x(4)-[LVIM]-[DKHETS]-G-x-[VIL]	x-[GPD]	x(2)-F-[SEKT]-x(4)-[GS]-x(3)	x(1,3)-G	[KRTDHE]-x(4)-[FIMAL]-x(2)
OCR	x(0,2)-[PRKTE]-x(4)-[LVIM]-x(3)-[LVI]-x(0,2)	x(2,6)	x(3,4)-[SNEKT]-x(2,3)-[GS]-x(2,3)	x(2,5)	[KRTDH]-x(4)-[FIMAL]-x(1,4)

	Loop	Strand4
MSA	x(0,1)-[TDG]-x(2,4)-[VFML]-x(5,10)-[FAMCI]-x-[YI]-G-x(3,6)-[FY]-x(5)-[HGED]-x-[KPS]-x(5,10)-[ETN]-G-x(0,1)	x(0,1)-[YFL]-x(2)-[EQTD]-x(4)-[FY]
SBA	x-[TSDKE]-[GSK]-x(0,2)-[LV]-x-[VFM]-x(4,10)-[LVI]-x(2,6)-[GS]-x(3)-[FIY]-x-[RTSKG]-x(3)-[HTGED]-x-[KGPSE]-x(0,2)-[DSNH]-x(7)-[EGN]-G-x(0,1)	[YIFL]-x(8)
SSS	x(9,11)-[LIV]-x(10)-[LIV]-x(3,7)-[FY]-x(5)-[HGED]-x(1,4)-[DKNH]-x(6)-[PGTN]-x(0,1)-[GE]	[YFL]-x(2)-[EQTD]-x(4)-[FY]
OCR	x(38,43)	x(0,1)-[YFL]-x(8)

	Loop	Strand5	Loop	Strand6
MSA	x(0,2)-[DHENQ]-[DPG]-G-x(0,1)	x(0,1)-[YMLAVC]-x(2)-[RKNHTS]-x-[EQRTDS]-x-[KSTER]-x(0,2)	x(2,10)	x(2,4)-[RSDKNH]-x-[EKTD]-[LCIVF]-x(3,4)
SBA	[KDEQ]-x(0,8)-[DTN]-[GD]-x	x-[YFLMACV]-x(2)-[RQNHTSK]-x-[ERDTS]-x-[KTSER]-[FYLI]-x	x-[DHNTG]	x(1,3)-[LIVF]-x(2)-[RKDNH]-x-[EQKTD]-[LFCIV]-x(3)
SSS	x-[DGN]-[DG]-x	x-[YLMAVC]-x(2)-[RKNHTS]-x-[EQRDTS]-x-[KSTER]-x(2)	x-[DGNT]	x-[LIVF]-x(2)-[RSDKNH]-x-[EKTD]-[LCVIF]-x(3)
OCR	x(4,6)	x(0,1)-[YMLAVC]-x(2)-[RKNHTS]-x-[EQRTDS]-x-[KSTER]-x(0,2)	x(2,10)	x(2,4)-[RSDNHK]-x-[EKTD]-[LCIVF]-x(3,4)

	Loop	Strand7	Loop	Strand8	Loop
MSA	x(2,8)-[LIFM]-[GPKQRT]-x-[KGTQD]-x(2,3)-[NHKRDGS]-x(3)	x(4)-[IYMVL]-x(2,3)	x(0,1)-[QSGT]-x(0,2)	x-[ILV]-x(8)-[NTQKS]-x(0,2)	x(3,10)
SBA	[GDN]-x-[HKPQR]-x(2,10)-[GS]-x-[IV]-x-[GQKRT]-x-[KPTQD]-x(5,7)	x-[NPTGS]-x(3)-[MLV]-x(2)	x(2)-[KSGPTD]x(0,1)	x-[ILV]-x(8)-[NTQKS]	x(2)-[DEGT]-[GP]-x(1,4)
SSS	x(3,10)-[HPDNKQ]-x(2,4)-[NHKRDG]-x(3)	x(4)-[IYMVL]-x(3)	x(1,3)-[NGST]	x-[IVL]-x(8)-[NTQKS]	x(2)-[DPGT]-[GS]-[SHRGK]
OCR	x(12,18)	x(8)	x(2,4)	x-[IVL]-x(8)-[NTQKS]-x(0,2)	x(2,10)

	Strand9	Loop	Strand10	Loop	Strand11
MSA	x(4,5)-[DQEHSR]-x(7,9)	x(1,9)-[GRNKH]-x(3,8)-[PS]-x(0,2)	x(10,14)	x(3,8)	x(9,11)
SBA	x(4)-[DHNTS]-x(4)-[NQK]-x(2)	x(6,11)	x-[YAF]-x(8)	x(8)	x(11)
SSS	x(7)-[QNTS]-x(4)	x(7,11)	x(10)	x(2,8)	x-[MLVY]-x(9)
OCR	x(12,14)	x(10,15)	x(10,14)	x(3,8)	x(9,11)

A

	Strand1	Loop1	Strand2	Loop2	Strand3
MSA	x-[VAY]-x(4)	x-[MILF]-[KSTN]-[YVIAL]-x(2,3)	x(0,5)	x(0,1)-[GSPR]-x	x-[VILF]-x(3)-[NPK]
SBA	x(2)-V-x-[LIMA]-x	x(5)-[TPQ]-x	x-[LIF]-x-[VIAL]-x	x(3)	x-[VIL]-x-[WFLV]-x-[NP]
SSS	V-x-[ILMV]-x(1,4)	x(0,3)-[DE]-x(1,4)-[YF]-x-[TPQ]-x(0,1)	x-[LFI]-x-[VIAL]-x(0,1)	x(1,2)-[GN]-[DET]-x(0,1)	x(0,2)-[VI]-x-[WFLV]-x(2,4)
OCR	x(4,10)	x(2,8)-[YF]-x(3,0)	x(4,5)	x(2,4)	x(1,2)-[VI]-x(4,6)

	Loop3	Strand4	Loop4	Strand5
MSA	x(3,4)-[PEGKQ]-x(7,11)-[LIV]-x(0,3)-[AVF]-x(2,11)-[MLYVF]-x(4,19)	[QEK]-x(5)-[FLV]	[TSKR]-x(2,4)	G-x-[YF]-[YFVLI]-x
SBA	x(3,4)-[PEGSQ]-[HE]-[NTG]-x(3,5)-[GDKN]-x(0,1)-[LIV]-x(2,8)-[MLYVI]-x(1,4)	[QEK]-x(5)-[FLVA]	[TSKDR]-[ENTPQK]-x	G-x-[YF]-x[YFVLI]-x-C
SSS	x(4,8)-[NES]-x(5,7)-[GDK]-x-[IVLF]-x(10,25)	x(4,8)	x(1,2)[ESPT]-x(1,3)	G-x-[YF]-x-[YFVLI]-x(0,1)
OCR	x(13,18)-[IVLF]-x(14,20)	x(4,8)	x(3,5)	G-x-[YF]-x-[YFVLI]-x(0,1)

	Loop5	Strand6
MSA	C-x(2,4)-[HP]-x(3,4)-[ML]	x(3)-[VIL]-x-V-x
SBA	C-x(2)-H-x(2)-[ML]	x(3)-[VIL]-x-V-x
SSS	x(0,1)-C-x(2,4)-H-x(3,7)	x(2,3)[VIL]-x-V-x(0,1)
OCR	x(0,1)-C-x(2,4)-H-x(3,7)	x(2,3)[VIL]-x-V-x(0,1)

B

	Loop1	Strand1	Loop2	Strand2
MSA	x(1,6)-[QT]-x-[TSN]-L-x-[KQR]-R-P	x(3)-[IVA]-x-[IV]	[GEN]-[GD]	x(3)-[EKN]-[AVF]-L-L
SBA	x-[QT]-x-[TSN]-L-x-[KQR]-R-P	x(3)-[IVA]-x-[IV]	[GEN]-[GD]	x(3)-[EKN]-[AVF]-L-L
SSS	x(1,6)-[QT]-x-[TSN]-L-x-[KQR]-R-P	x(3)-[IVA]-x-[IV]	[GEN]-[GD]	x(3)-[EKN]-[AVF]-L-L
OCR	x(9,14)	x(3)-[IVA]-x-[IV]	x(2)	x(3)-[EKN]-[AVF]-L-L

	Loop3	Strand3	Loop4
MSA	[DN]-T-G-A-D-x-[TS]-[VI]-[ILFV]-x-[EGRT]-x-[SNEDH]-[LFY]-[PGQN]-x(0,6)-[GNPR]-[RNHGK]-x	[KTSQ]-x(1,3)-[IVM]-x-G	[IV]-G
SBA	[DN]-T-G-A-D-x-[TS]-[VI]-[ILFV]-x-[EGRT]-[MVIFY]-[SNEQ]-x(3)-[RNHE]-x	x-[PRQ]-[KQG]-x-[IVM]-x-G	[IV]-G
SSS	[DN]-T-G-A-D-x-[TS]-[VI]-[ILFV]-x-[EGRT]-x(1,6)-[SNEDK]-[LFY]-[PGQR]-x(0,1)-[GNPK]-x(2)	x(0,3)-[KTSG]-[PRT]-[KQG]-x(2,4)	x(2,6)
OCR	x(18,23)	x(5,8)	x(2,6)

	Strand4	Loop5	Strand5
MSA	G-x(3,4)-[RKT]-x(3)-[QNP]-[IV]-x-[IL]-E-x	x-[GND]	[HKER]-[KRNH]-[AVIY]-x(2)-[TQR]-[VICM]-[LMI]-x(2)
SBA	G-x(2)-[KNRE]-x(2)-[QENS]-x(2,3)-[IV]-x-[IL]-[EK]-x	x-[GNT]	[HKQR]-x-[AVI]-x(2)-[TNR]-[VIM]-[LMC]-x(2)
SSS	x(3,51)-[RKT]-x(3)-[QNP]-[IV]-x-[IL]-[EK]-x	x(1,7)-[GHNQT]	x(1,2)-[AVI]-x(2)-[TSNR]-[VIM]-[LMC]-x(1,2)
OCR	x(8,10)-[IV]-x-[IL]-[EK]-x	x(2,8)	x(1,2)-[AVI]-x(2)-[TSNR]-[VIM]-[LMC]-x(1,2)

	Loop6
MSA	[PDN]-x(4,12)-[IVL]-[IFL]-G-R-[ND]-x-[LM]-x(2)-[ILF]-[GN]-[ACMI]-[TSRK]-L-x-[FLM]
SBA	[PDS]-x(3)-[NPT]-[IVL]-[IFL]-G-R-[ND]-x-[LM]-x(2)-[ILF]-[GN]-[ACMI]-[TSRK]-L-x-[FLM]
SSS	x(1,3)-[PDS]-x(3)-[NPT]-[IVL]-[IFL]-G-R-[ND]-x-[LM]-x(2)-[ILF]-[GN]-[ACMI]-[TSRK]-L-x-[FLM]
OCR	x(21,24)

C

	Loop1	Strand1	Loop2	Strand2
MSA	[IL]-[QPE]-x-[EQG]-x	x-[LIMV]-[KENT]-[VC]	A-D-N-[ST]-G	A-R-[EKRN]-[VIL]-x-[VCI]- I-x-V
SBA	[QPE]-x-[EQG]-x	x-[LIMV]-[KENT]-[VC]	A-D-N-[ST]-G	A-[RK]- [EKRN]-[VIL]-x-[VCI]-I-x-V
SSS	[EQG]-x	x (1,3)-[LIMV]-[KENT]-x(0,1)	x(0,1)-A-D-N-x(0,1)-[GT]-x(0,3)	x(0,2)-[EKRN]-[VIL]-x(1,7)
OCR	x(4,24)	x (1,3)-[LIMV]-[KENT]-x(0,1)	x(5,8)	x(0,2)-[EKRN]-[VIL]-x(1,7)

	Loop3	Strand3	Loop4	Strand4
MSA	x-[GK]-x-[SKGH]-[GNHS]-x(3,7)-A-x[IVL]-G-D	x-[VI]-x-[AIV]-[TS]-[VI]-[KT]-[DEK]-x	x-P-[GRE]-x(1,4)-[KER]-[GKR]-[QDK]	V-x-[KPEN]-A-[VI]-[VI]-[VI]-R
SBA	x-[GK]-x-[SKGRH]-x(1,2)-[RKP]-x-A-x-[IVL]- G-D	x-[VI]-x-[AIV]-[TS]-[VI]-[KT]-[DEK]-x	x(5,8)- [GKR]-[QDK]	V-x-[KPEN]-A-[VI]-[VI]-[VI]-R
SSS	x(1,5)-G-x(3,6)-[RK]-x(1,3)-A-x-[IVL]-G-x(1,2)	x(0,1)-[VI]-x-[AIV]-[TS]-x(0,4)	x(1,4)-P-[GRE]-x(2,5)-[GKR]-x	x(1,2)-[KPEN]-A-x(0,4)
OCR	x(13,18)	x(0,1)-[VI]-x-[AIV]-[TS]-x0,4)	x(7,13)	x(1,2)-[KPEN]-A-x(0,4)

	Loop5	Strand5
MSA	[TQ]-x-[RK]-[GESP]-[VIYW]-[RK]-R-[PKR]-D-G-x(2)-[IVL]-x-F-[DE]-[EDG]-N	A-x-V-[IL]-[ILVA]
SBA	[TQ]-x-[RK]-[GESP]-[VIYW]-[RK]- R-[PKR]-D-G-x(2)-[IVL]-x-F-[DE]-[EDG]-N	A-x-V-[IL]-[ILVA]
SSS	x(0,4)-[TQ]-x-[RK]-[GESP]- [VIYW]-[RK]- R-[PKR]-D-G-x(2)-[IVL]-x-F-[DE]-[EDG]-x(1,4)	x(0,3)-[IL]-x(0,1)
OCR	x(18,25)	x(0,3)-[IL]-x(0,1)

	Loop6
MSA	[RND]-[DNEQ]-x(0,1)-[DQSKN]-x(2)-G-[TS]-x-[IVL]-x-G-P-[VI]-x-[RK]-E-[LCVA]-x-[DESQ]-x(1,2)-[FW]-x-[KRS]-[IV]-x-[ST]-x(2)-[PGT]-x-[VI]
SBA	[RND]-[DNEQ]-[DQSKN]-x(2)-G-[TS]-x-[IVL]-x-G-P-[VI]- x-[RK]-E-[LCVA]-x-[DESQ]-x(1,2)-[FW]-x-[KRS]-[IV]-x-[ST]- x(2)-[PGT]-x-[VI]
SSS	x-[DNE]-x(2,3)-[SEDQ]-x(2)- G-[TS]-x-[IVL]-x-G-P-[VI]- x-[RK]-E-[LCVA]-x(1,2)-[KEDQ]-x-[FW]-x-[KRS]-[IV]-x-[ST]- x(2)-[PGT]-x(1,2)-[ILAV]
OCR	x(33,35)

D

	Strand1	Loop1	Strand2	Loop2	Helix1
MSA	[IL]-x-[VI]-x(1,2)	x(3,4)- [KRDSQ]	x-[IWVL]-x-[LIVF]-[EDNKS]	[VIL]-x(2)-[STEHD]-[DST]-[TKHP]	[IVL]-x(2)-[VIFL]-x(3)-[IVYF]-x-[DEQK]-x(2)
SBA	[IL]-x-[VI]-x-[TSGD]	x(2,3)- [KRDSQ]	x-[IWVL]-x-[LIVF]-[EDNKS]	[VIL]-x(2)-[STEHD]-[DST]-[TKHP]	[IVL]-x(2)-[VIFL]-x(3)-[IVYF]-x-[DEQK]-x(2)
SSS	[IL]-x-[VI]-x(1,2)	x(3,5)	x(3,4)-[LIVF]-[EDNKS]-x(0,2)	x(0,2)-[PTRK]-[STEHD]-x(0,3)	x(2,3)-[VIFL]-x(3)-[IVFY]-x-[DEKQ]-x(0,2)
OCR	[IL]-x-[VI]-x(1,2)	x(3,5)	x(3,4)-[LIVF]-[EDNKS]-x(0,2)	x(4,6)	x(9,12)

	Loop3	Strand3	Loop4
MSA	G-x-[PSE]-x(2)-[QNSK]	x-[RS]-[LF]-x-[FY]	x-G-x(2)-[LMI]-x(2)-[GEDNTK]-[RKQHDE]-x(3)-[DESQ]-[YL]-[NKHGED]-[IML]-x-[KNGDES]-[EHGN]-[SHD]
SBA	G-x-[PSE]-x(1,2)-[QNSK]	x-[RS]-[LF]-x-[FY]	x-G-x(5)- [GEDNTK]-[RKQHDE]-x(3)-[DESQ]-[YL]-[NKHGED]-[IML]-x-[KNGDES]-[SHD]
SSS	x(0,2)-G-x-[PSE]-x(2)-[QNSK]-x(0,1)	x(0,1)-[RS]-[LF]-x-[FY]	x(1,6)-G-x(2)-[LMI]-x(2)-[GEDNT]-x-[TG]-x(2)-[DSEQ]-x(0,5)-[EHGN]-x(0,1)
OCR	x(6,8)	x(0,1)-[RS]-[LF]-x-[FY]	x(19,20)

	Strand4
MSA	x-[LVI]-x-[LVA]
SBA	x-[LVI]-x-[LVA]
SSS	x(0,1)-[LVI]-x-[LVA]
OCR	x(0,1)-[LVI]-x-[LVA]

E

	Strand1	Loop1	Helix1	Loop2	Strand2	Loop3
MSA	[KR]-LVI-x-[IVLM]-[VLIF]-G	x(4)-G	K-[TS]-x-[LI]-x(6)	[DNKGES]-[QHKTDE]-x(6,7)-[TE]-x	x(1,2)-[NSDRT]-[YFIV]-x(2,6)-[VICL]	[DNET]-x
SBA	[KR]-LVI-x-[IVLM]-[VLIF]-G	x(4)-G	K-[TS]-x-[LI]-x(5,6)	[DNKG]-x(4,7)	x-[EDGN]-x(7,8)	x(0,2)
SSS	[KR]-LVI-x-[IVLM]-[VLIF]-x(0,1)	x(5,6)	K-[TS]-x-[LI]-x(3,6)	x(0,2)-[DNKG]-x(7,8)-[TE]-x(1,3)	x(0,1)-[EDGN]-x(3,7)	x(2,6)
OCR	[KR]-LVI-x-[IVLM]-[VLIF]- x(0,1)	x(6)	x(7,10)	x(11,14)	x(5,10)	x(2,6)

	Strand3	Loop4	Helix2	Loop5
MSA	[KERQS]-x-[VCYFLI]-x-[LMF]-x-D-x	x-G-x	[EPHKTD]-[DERQSK]-[YFLI]-[DSQRP]-x(1,6)	[YMAIF]-x(3)-[DEMHQ]
SBA	(1,2)-[VCYFLI]-x-[LMF]-x-[LIV]-x-D-x	x-G-x(0,1)	x(4,10)	[YMAIF]-x(4)
SSS	x(1,4)-[LMF]-x-[LIV]-x-D-x(0,1)	x(1,2)-G-x(1,8)	x(6,11)	x(2,4)
OCR	x(1,4)-[LMF]-x-[LIV]-x-D-x(0,1)	x(3,11)	x(6,11)	x(2,4)

	Strand4	Loop6	Helix3	Loop7
MSA	x-[IFCVL]-[LVI]-[MCLFWY]-[CV]-[FYV]-x	x(3)	[PTEKR]-x-[STR]-[LFMIV]-[EKTQGN]-x(3)-[EQSNPRE]-x(9,13)	x(2)-[NDEGK]-[VFAI]-x
SBA	x-[IFCVL]-[LVI]-[MCLFWY]-[CV]-[FYV]-x	x(3)	[PTEKR]-x-[STR]-[LFMIV]-[EKTQGN]-x(3,4)-[KQSNPE]-x(3,4)-[VIFLM]-x(3)	x(2)-[NDPEGK]-[VFAI]-x
SSS	x-[IFCVL]-[LVI]-[MCLFWY]-[CV]-[FYV]-x	x(3,6)	x(0,6)-[ILVM]-x(8,16)	x(3,9)
OCR	x-[IFCVL]-[LVI]-[MCLFWY]-[CV]-[FYV]-x	x(3,6)	x(15,18)	x(3,9)

	Strand5	Loop8	Helix4	Loop9	Strand6	Loop10
MSA	[IMFLV]-[IVLM]-[LVI]-[VLF]-x-[NT]	[KQ]-x(15,18)-[MLWV]-x(0,6)	x(1,9)-[RSKDQ]-x	x(2)-[FICW]	x-[YFI]-x(2)	x(2)-A-x(4)-[GND]
SBA	[IMFLV]-[IVLM]-[LVI]-[VLF]-x-[NT]	[KQ]-x(14,18)-[VILM]-x	x(10,11)	x(2,3)	x-[YFI]-x(2)	x(2)-A-x(4)-[GND]
SSS	x(0,1)-[IMFLV]-[IVLM]-[LVI]-[VLF]-x-[NT]	x(1,2)-D-x(6,20)	x(4,8)-[NRQKED]-x(1,5)	x(2,5)	x(0,2)-[YFI]-x(0,3)	x(0,4)-A-x(4)-[GND]
OCR	x(0,1)-[IMFLV]-[IVLM]-[LVI]-[VLF]-x-[NT]	x(1,23)	x(10,12)	x(2,5)	x(0,2)-[YFI]-x(0,3)	x(7,10)

	Helix5
MSA	[VL]-x-[EDQRP]-x-[FIMAL]-x(2)-[ALIV]-x-[RKQTDEN]
SBA	[VL]-x-[EDQRP]-x-[FIML]-x(2)-[ALI]-x-[RKQTDEN]
SSS	x(0,3)-[VAIL]-x(3)-[ALIVM]
OCR	x(11,14)

F

	Helix1	Loop1	Helix2	Loop2	Strand1	Loop3
MSA	x(0,9)	x(0,11)-[NRPQ]	x-C-[NETD]-x-[MAIV]-M-x(2)	x-[NKGR]-x(4,9)-C-K-[PDNE]	x-N-T-F-[VIL]	[HRQ]-x-[STN]
SBA	x(0,9)	x(0,11)	x(1,2)-C-[NETD]-x-[MAIV]-M-x(1,2)	x(3,4)-[KSQNRG]-x(1,2)-C-K-[PDNE]	x-N-T-F-[VIL]	[HRQ]-x-[STN]
SSS	x(0,11)	x(0,12)	x(0,3)-C-[NETD]-x-[MAIV]-M-x(2,4)	x(2,8)-[KTQRG]-x(2)-C-K-[PDNE]	x-N-T-F-[VIL]	[HRQ]-x-[STN]
OCR	x(0,11)	x(0,12)	x(7,11)	x(8,14)	x-N-T-F-[VIL]	x(3)

	Helix3	Loop4	Strand2	Loop5	Strand3	Loop6
MSA	x(2)-[DNSHR]-[VI]-x(2)-[VI]-C	x(2,3)	[SGET]-[QKNGD]	x(2,7)-[TRNE]-[ND]	[CL]-x(2)	S-x(2)
SBA	x(2)-[DNSHR]-[VI]-x(2)-[VI]-C	[SGET]-[QGN]	[KSQNT]-[NPSR]-x	x(1,7)-[ND]	[CL]-x(2)	S-x(2)-[THRSQP]
SSS	x(2)-[DNSHR]-[VI]-x(2)-[VI]-x(0,2)	x(1,4)	x(0,4)	x(2,10)	x(2,4)	x(0,2)-[SQP]-x(0,3)
OCR	x(7,9)	x(1,4)	x(0,4)	x(2,10)	x(2,4)	x(3,4)

	Strand4	Loop7	Strand5
MSA	[THRSQP]-[MFV]-[SQRPT]-[IVL]-x-[DTH]-C-[RTDK]-x	x(0,2)-[TKPHR]-[GS]-x-[SQEN]-[KPNR]-x(2)-[NP]-C-x	Y-x(2)-[TSRK]-x(3)-[KR]
SBA	[MFV]-[SQRPT]-[IVL]-x-[DTH]-C-[RTDK]-x	x(1,3)-[SG]-x(2)-[PSR]-[NP]-C-x	Y-x(2)-[TSRK]-x(3)-[KR]
SSS	[MFV]-[SQRPT]-[IVL]-x(2)-C-[RTDK]-x(1,3)	x(1,5)-[SQEN]-[KPNR]-x-P-[NP]-C-x(0,1)	x(0,1)-Y-x(2)-[TSRK]-x(3)-[KR]-x(0,1)
OCR	[MFV]-[SQRPT]-[IVL]-x(2)-C-[RTDK]-x(1,3)	x(7,12)	x(0,1)-Y-x(2)-[TSRK]-x(3)-[KR]-x(0,1)

	Loop8	Strand6	Loop9	Strand7
MSA	x	[IYV]-[IV]-[VIL]-x-C-[EGD]	[GNK]-[NGKRE]-x(0,10)	[VWL]-P-[VA]-H-[FLY]-[DE]-x-[STG]-[VFI]
SBA	x	[IYV]-[IV]-[VIL]-x-C-[EGD]	[GNK]-x(1,3)	[VWL]-P-[VA]-H-[FLY]-[DE]-[STG]
SSS	x(0,2)	x(0,2)-[IV]-[VIL]-x-C-[EGD]-x(0,1)	x(0,4)-[GD]-x(0,5)	x(0,1)-[VWL]-P-[VA]-H-[FLY]
OCR	x(0,2)	x(0,2)-[IV]-[VIL]-x-C-[EGD]-x(0,1)	x(2,10)	x(0,1)-[VWL]-P-[VA]-H-[FLY]

G

	Strand1	Loop5	Helix4	Loop6	Strand2
MSA	[TGDNS]-x(1,2)-[NRQDES]	[VIL]-x(3,4)	x(1,3)-[LMI]-x(6,7)	[QGNH]-x(1,2)	[VI]-x(3)-[LVIF]
SBA	[TGDNS]-x(2)-[NRQES]	[VIL]-x(2,4)	x(2.3)-[LMI]-X(3)-[VIL]-x(2)	[QGNS]-x	[VI]-x(3)-[LVIF]
SSS	x(2,5)[NRQDES]-x(0,1)	x(1,4)-[NQD]-x(0,1)	x(0,3)-[LMI]-x(3)-[VIL]-x(2,3)	x(0,1)-[QGNH]-x(1,2)	[VI]-x(3)-[LVIF]
OCR	x(2,5)[NRQDES]-x(0,1)	x(3,5)	x(7,11)	x(2,3)	[VI]-x(3)-[LVIF]

	Loop7	Strand3	Loop8
MSA	x(4,7)-[FW]-x(2)-Y-x-[SGEK]-[GR]-x-[FYMI]-x-[GESQH]-x(7,11)	H-x-[IV]-x-[IAVL]-[VL]-G-[YW]-[GNS]-x(0,3)	x(0,11)
SBA	x(4,6)-[FW]-x(2)-Y-x-[SGEKR]-[GT]-[IV]-[FYMI]-x-[GESRH]-x(3,7)	H-x-[IV]-x-[IAL]-[VL]-G-[YFW]-x(0,3)	x(0,2)
SSS	x(2,6)-[FW]-x(2)-Y-x-[SGEK]-[GR]-x-[FYMI]-x-[GESQH]-x(3)-[TNSHDE]-x(0,3)	H-x-[IV]-x-[IAVL]-[VL]-G-x(1,4)	x(0,1)-[GDP]
OCR	x(17,27)	H-x-[IV]-x-[IAVL]-[VL]-G-x(1,4)	x(1,2)

	Strand4	Loop9	Strand5	Loop10	Strand6
MSA	x(2)-Y-[WIY]-[IL]-[VI]-[KRG]	N-S-W-[DGTN]-[TQEP]-x-W-G-x-[EQKNGR]-G	[YF]-x(2)-[IMF]-x	(5,9)-[CM]-x-[ILV]-x(1,4)	x(4)
SBA	x(0,2)-Y-[WI]-[IL]-[VI]-[KRG]	N-S-W-[DGTN]-[TQEP]-x-W-G-x-[EQKNGR]-G	[YF]-x(2)-[IMF]-x	x(3,6)-[GNDR]-x-[CM]-x-[ILV]-x(4)	x(4)
SSS	x(0,3)-[WIY]-[IL]-[VI]-[KRG]	N-S-W-[DGTN]-[TQEP]-x-W-G-x-[EQKNGR]-G	[YF]-x(2)-[IMF]-x(0,1)	x(3,8)-[GNRD]-x-[CM]-x-[ILV]-x(3,4)	x(3,4)
OCR	x(0,3)-[WIY]-[IL]-[VI]-[KRG]	x(11)	[YF]-x(2)-[IMF]-x(0,1)	x(11,15)	x(3,4)

H

	Loop1	Strand1	Loop2	Helix1	Loop3	Strand2
MSA	[FYM]	[TRE]-[IL]-[NKT]-A-x(2)	R-x(0,4)	x(0,1)-[GKT]-x-[SK]—x(2,3)-L-[RD]	x(2,3)	x-[FLI]-x(2)-[IV]-[IMA]-Y-[GN]
SBA	x(0,2)	x(0,1)-[ILM]-[NKER]-[AL]-x(0,1)	x(3)-[QETR]-x(0,1)	x(0,3)-[SP]-[RSK]-x(3)	x(2)-[NGS]	x-[FLI]-x(2)-[IV]-[IMA]-Y-[GN]
SSS	x(2,3)	x(2,8)	x(0,4)-[ER]-[QET]-[GP]-x(0,4)	x(2,6)-[RQ]-x(0,4)	x(1,2)-[NGE]-x(0,3)	x(2,5)-[IV]-[IMA]-Y-x(0,1)
OCR	x(2,3)	x(2,8)	x(4,11)	x(4,9)	x(1,5)	x(4,7)-[IV]-[IMA]-Y-x(0,1)

	Loop4	Strand3	Helix2	Loop5	Strand4
MSA	x(0,1)-[KR]	[EH]-x-[PN]-x(2)-[IVF]-x-[LV]-D	x(3)-[VF]-x-[NKR]-[MVA]-x	x-[KQ]-x-[ES]-x(2)-[SHG]-x-[VIF]-x	x(2)-[VL]-x
SBA	[GRK]-x(0,1)	x(1,2)-[PN]-x(2)-[IVF]-x-[LV]-D	x(3)-[VF]-x-[NKR]-[MVA]-x	x-[KQ]-x(3,4)-[SHT]-x-[VLI]-[LIFV]	x(4)
SSS	x(0,1)-[KGR]-x(0,4)	x(1,6)-[IVF]-x(1,3)	x(1,6)D-[KR]-[VA]-x(0,4)	x(0,7)-[ESG]-[FIL]-x(1,2)-[EHD]-x(0,10)	x(2,5)
OCR	x(2,6)	x(1,6)-[IVF]-x(1,3)	x(3,8)-[VA]-x(0,4)	x(4,20)	x(2,5)

	Loop6	Strand5	Loop7	Strand6	Loop8
MSA	[DG]-G-[KQE]-[EST]-[ILF]-[KP]	x(2)-[AV]-[QRK]-x-V-[QN]-x	[HD]-[PK]-x-[KR]-[PR]-x	x(2)-H-[IV]-D-F-[VFY]-x	[ALV]
SBA	[DPE]-[GD]-[KQE]-[EST]-[ILF]-[KP]	x(2)-[AV]-[QRK]-x-V-[QN]-x	[HD]-[PK]-x-[KR]-[PR]-x	x(1,2)-H-[IV]-D-F-[VFY]-x	[ALV]
SSS	x(0,4)-[DK]-[GA]-x(1,6)	x(2,12)	x(0,8)-[PK]-x-[KR]-[PR]-x(0,6)	x(0,6)-F-[VFY]-x(0,1)	x(0,1)-[ALV]
OCR	x(0,4)-[DK]-[GA]-x(1,6)	x(2,12)	x(0,8)-[PK]-x-[KR]-[PR]-x(0,6)	x(0,6)-F-[VFY]-x(0,1)	x(0,1)-[ALV]

I

	Strand1	Loop1	Strand2	Loop2	Strand3	Loop3
MSA	[RSEK]-x-[ETNQP]-[LMVI]-x	[IV]-P-x-[DGK]-V	[DTS]-[AVI]-[EQTSK]-x(2)	x(2)	x-[IV]-[TEK]-[VI]-x	G-[DPK]-x
SBA	x(0,2)-[ETNQP]-[LMVI]-x	[IV]-P-x-[DGK]-V	[DTS]-[AVI]-[EQTSK]-x(2)	x(2)	x-[IV]-[TEK]-[VI]-x-G	[DPK]-x
SSS	x(0,7)-[LMVI]-x	x(0,2)-[IV]-P-x-[DGK]-V-x(0,2)	x(0,2)-[EQTSK]-x(1,2)	x(2,3)	x-[IV]-x-[TEK]-x(0,3)	x(0,3)-[DPK]-x(1,4)
OCR	x(0,7)-[LMVI]-x	x(5,7)	x(0,2)-[EQTSK]-x(1,2)	x(2,3)	x-[IV]-x-[TEK]-x(0,3)	x(2,9)

	Strand4	Loop4	Strand5	Loop5	Strand6	Loop6
MSA	G-[SKTE]-[VLI]-[TKE]-x(2)	[LFV]-x(2,3)-[ILVM]	[DTRE]-[VFI]-x(0,2)	x(3,5)-[DRQNG]	x-[VLI]-x-[IVLF]-x	x(5)
SBA	G-[SKTE]-[VLI]-[TKE]-x(2)	[LFV]-x(2,3)-[ILVM]	[DTRE]-[VFI]-x(2)	x(2)-[DRNG]	x-[VLI]-x-[IVLF]-x	x(5)
SSS	x(0,3)-[TKE]-x(2)	x(5,6)	x(2,6)	x(1,5)-[GDN]-x(0,2)	x-[VIL]-x-[IVF]-x(1,5)	x(1,5)
OCR	x(0,3)-[TKE]-x(2)	x(5,6)	x(2,6)	x(2,8)	x-[VIL]-x-[IVF]-x(1,5)	x(1,5)

	Helix1	Loop7	Strand7	Loop8
MSA	x-[KRHD]-[TEHG]-x(4)-[GQR]-x(5)-[IVL]-x-[NS]-[MLA]-[FIV]-x-G	V-[TS]-[EK]-G	[WYF]-[EKRTS]-x(2)-[ML]-x-[VLFYI]-x	x(2)-[HG]-[FY]-[PR]-[MIA]
SBA	x(0,1)-[KRHD]-[TEHG]-x(4)-[GQR]-x(5)-[IVL]-x-[NS]-[MLA]-[FIV]-x-G	V-[TS]-[EK]-G	[WYF]-[EKRTS]-x(2)-[ML]-x-[VLFYI]-x	x(2,4)-[PG]-[MIY]
SSS	x-[KRHD]-x-[TEHG]-x(4)-[GRQ]-x(5)-[IVL]-x-[NS]-[MLA]-[FIV]-x-G	V-[TS]-[EK]-G-x(0,1)	x(0,1)-[ERKTS]-x(2)-[ML]-x-[VFLYI]-x	x(2)-[HG]-[FY]-x(0,2)
OCR	x(20)	x(4,5)	x(0,1)-[ERKTS]-x(2)-[ML]-x-[VFLYI]-x	x(4,6)

	Strand8	Loop9	Strand9	Loop10	Strand10	Loop11
MSA	x(3,5)	[GKNE]-[DSKNR]-x(0,4)	x(1,3)-[IVL]-x	x(2)-[LV]-G-x-[KS]-x	x(4)	[TDQPE]-x(3)-[DG]-x
SBA	[QNR]-x(4)	[GKNE]-[DSKNR]	x-[VIL]-x-[IVL]-[LV]-[ERKST]	x(6,8)	x(3)-[TDQPE]	x(3)-[DG]-x
SSS	x(3,5)	x(0,3)-[GN]-x(1,2)	x(0,1)-[VIL]-x-[IVL]-x(0,1)	x(1,2)-[LV]-G-x-[KS]-x(1,2)	x(2,5)	x(2,4)-G-x(0,2)
OCR	x(3,5)	x(2,6)	x(0,1)-[IVL]-x(0,1)	x(6,8)	x(2,5)	x(4,5)

	Strand11	Loop12	Strand12	Loop13	Helix2
MSA	[DTKE]-[VIAF]-x(3,6)	GKQP]-[EDNT]	[ETKR]-[LIV]-x-[VL]-[SDQK]-G	x-[DS]	x-[ENQ]-x-[VI]-[GS]-[QRE]-x-[AC]-A-x-[IL]-[ENHQR]-x
SBA	[DTKE]-[VIAF]-x(2)-[DSEP]	[GKQST]-[EDNT]	[ETKR]-[LIV]-x-[VL]-[SDQK]-G	x-[DS]	x-[ENQ]-x-[VI]-[GS]-[QRE]-x-[AC]-A-x-[IL]-[ENHQR]-x
SSS	x(0,1)-[DTKE]-[VIAF]-x(1,4)	x(0,4)-[EKQP]-x(0,4)	[ETKR]-[LIV]-x-[VL]-[SDQK]-G	x-[DS]	x-[ENQ]-x-[VI]-[GS]-[QRE]-x-[AC]-A-x-[IL]-[ENHQR]-x
OCR	x(0,1)-[DTKE]-[VIAF]-x(1,4)	x(2,7)	[ETKR]-[LIV]-x-[VL]-[SDQK]-G	x(2)	x(13)

	Loop14	Strand13	Loop15
MSA	x(4)-[NKRES]-x(4,8)-[DK]	G-[VI]-x-[IVY]-x-[RDHG]	[KE]-x(4)-[DENK]
SBA	x(9,13)-[DK]	G-[VI]-x-[IVY]-x-[RDHG]	x(6,10)
SSS	x(4)-[NKRES]-x(2)-[DKH]-x-[RK]-x(0,4)	G-[VI]-x-[IVY]-x(0,5)	x(1,7)-[NTK]-[REP]-x(0,7)
OCR	x(11,15)	G-[VI]-x-[IVY]-x(0,5)	x(5,16)

J

	Strand1	Loop1	Strand2	Loop2	Helix1	Loop3
MSA	x(4)-[AWVF]-x(2)	x(4)	x-[IL]-G	x-[EDNKQ]-[NG]-x-[ML]—P-W-[NHSPKR]-x(0,2)	x(2)-[DE]-x(3)-[FY]-[KRSQ]-x(3)	x(2,47)-[KND]-x
SBA	x(3)-[ILV]-[AWVC]-A-x	x(4)	x-[IL]-G	x(7,9)	x(2)-[DE]-x(3)-[FY]-[KRSQ]-x(3)	x-[DGKQ]-[KND]-x
SSS	x(2,4)-[ILV]-[AWVC]-A-x(1,4)	x(3,18)	x(1,3)-[IL]-G-x(0,1)	x(0,1)-[EDNKQ]-x-[ML]-P-W-[NHSKPR]-x(1,4)	x(1,2)-[DEK]-x(2)-[WLFY]-x(2)-[RKSQ]-x(0,4)	x(4,48)
OCR	x(3,5)-[AWVC]-x(2,5)	x(3,18)	x(1,3)-[IL]-G-x(0,1)	x(9,11)	x(7,12)	x(4,48)

	Strand3	Helix2	Loop4	Strand4
MSA	[VIML]-[IV]-[MVL]-G	[RK]-[HNRKST]-[TN]-[WFY]-x-S-[IFM]	x(0,1)-[GKREQ]-x(0,2)-[RKD]-[PD]-L-[PK]-[GENK]-[RS]-x	x-[IVL]-[IV]-[LVM]
SBA	[VIML]-[IV]-[MVL]-G	[RK]-[HNRKST]-[TN]-[WFY]-x-S-[IFM]	[GKREQ]-[RKD]-x(2)-[PKG]-x(0,1)-[RS]-x	x-[IVL]-[IV]-[LVM]
SSS	x(0,3)-[VIML]-[IV]-[MVL]-G	[RK]-[HNRKST]-[TN]-[WFY]-x(1,3)	x(1,5)-[RKD]-[PD]L-[PK]-[GENK]-x(0,2)	x(0,2)-[IVL]-[IV]-[LVM]
OCR	x(0,3)-[VIML]-[IV]-[MVL]-G	x(5,7)	x(6,9)	x(0,2)-[IVL]-[IV]-[LVM]

	Loop5	Strand5	Loop6	Helix3	Loop7	Strand6
MSA	[ST]-[SRHK]-[QSET]-x-[GDPKTR]-x(5,10)	x(1,3)	x(0,2)	x(2,4)-[IVFL]-x	x(3,10)-[DPSK]-x(3)	[ILVCA]-[MVFWY]-[VI]
SBA	[ST]-[SRHK]-[QSET]-x(1,2)-[TQDPS]-x(3)-[VAI]	x(3)	x-[STDN]	[VLI]-x(3)-[IVFL]-x	x(7)	[ILVCA]-[MVFWY]-[VI]
SSS	x(0,5)-[SQE]-x-[PDKTS]-x(4,10)	x(3,4)	x(0,1)-[STDN]-x(0,1)	[VLI]-x(3,9)	x(0,3)-[DNEKS]-x(1,2)-[ENK]-x(0,3)	x(0,3)-[ILVC]-[MVFWY]-[VI]-x(0,2)
OCR	x(10,12)	x(3,4)	x(2)	x(4,10)	x(3,9)	x(0,3)-[ILVC]-[MVFWY]-[VI]-x(0,2)

	Loop8	Helix4	Loop9	Strand7	Loop10	Strand8
MSA	[IACV]-G-G	x(2)-[VI]-YF]-x(3)	x(5,7)	[KRT]-[LIWM]-[YLFV]-LVFI]-TS]-[HRK]-[ILV]	x(0,8)	x(0,2)
SBA	[IACV]-G-G	x(2)-[VI]-YF]-x(3)	x(4,5)	[KRTQH]-[LIWM]-[YLFV]-LVFI]-TS]-[HRK]-[ILV]	x(4)-[EQP]-x-D	x(2)
SSS	x(0,1)-G-x(0,1)	x(2)-[VI]-YF]-x(3,7)	x(1,4)-[QDK]	x(0,2)-[KRTQH]-[LIWM]-[YLFV]-LVFI]-TS]-[HRK]-x(0,5)	x(2,3)-[ETSD]-x(0,6)	x(2,3)
OCR	x(1,3)	x(6,10)	x(2,5)	x(0,2)-[KRTQH]-[LIWM]-[YLFV]-LVFI]-TS]-[HRK]-x(0,5)	x(3,8)	x(2,3)

	Loop11	Strand9	Loop12	Strand10
MSA	x(1,2)-[DPE]-[YLIFW]-[ENPD]-x(3)	x(4,9)	x(1,52)-[SNTKQG]	x(2)-[FYL]-x(2)-[LWFY]-x-[RK]
SBA	[FMLY]-x-[DPEK]-x-[ENPD]-x(2)-[DNTKER]	[WFY]-x(8)	x(1,8)-[SNTKQG]	x(2)-[FYL]-x(2)-[LWFY]-x-[RK]
SSS	x(0,21)-P-x(0,3)-[DNEQ]-[YWILF]-[EPD]-x(1,10)	x(5,16)	x(2,9)	x(1,4)-[FYL]-x(2)-[LWFY]-x(0,3)
OCR	x(8,38)	x(5,16)	x(2,9)	x(1,4)-[FYL]-x(2)-[LWFY]-x(0,3)

K

	Loop1	Strand1	Loop2	Strand2	Loop3	Strand3	Loop4	Helix1
MSA	P-E	I-R	T-[GMH]-x-[ILFM]-x-[GD]-[SGH]-x(1,4)	[VAI]-x-[LM]	x(2)-G-x	[TQE]-[LVCF]-x-[IVF]-x	x-[DN]	x(0,3)
SBA	P-E	I-R	T-x(2)-[ILFM]-x(4,7)	[VAI]-x-[LM]	x(2)-G-x	[TQE]-[LVCF]-x-[IVF]-x	x-[DN]	x(0,3)
SSS	P-E	I-R	T-x(2)-[ILFM]-x-[GD]-[SGH]-x(0,4)	x(0,1)-[VAI]-x(1,2)	x(2,3)-G-x	[TQE]-[LVCF]-x-[IVF]-x	x-[DN]-x(0,1)	[NPSE]-[ALV]-x
OCR	x(2)	I-R	x(7,11)	x(0,1)-[VAI]-x(1,2)	x(4,5)	[TQE]-[LVCF]-x-[IVF]-x	x(2,3)	x(3)

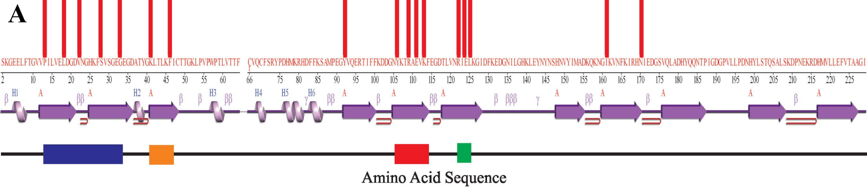
	Loop5	Helix2	Loop6	Strand4	Loop7
MSA	x(4)-[DNT]-x-[NDET]-x-[LVFI]-x-[LVIC]-[DTS]-Y-[KPQEG]-x	[ILF]-x-[KRTQN]	x-[VL]-[DPRSKEQ]-x-G-[SGN]	x-[VI]-[YL]-[VIL]	x-D-G-x
SBA	x(2,4)-[DNT]-x-[NDET]-x-[LVFI]-x-[LVIC]-[DTS]-Y-[KPQEG]-x	[ILF]-x-[KRTQN]	x-[VL]-[DPRSKEQ]-x-G-[SGN]	x-[VI]-[YL]-[VIL]	x-D-G-x
SSS	x(0,3)-[DNT]-x-[NDET]-x-[LVFI]-x-[IVC]-[DTS]-Y-x(0,2)	x(0,2)-[ILF]-x(1,3)	x(0,2)-[VL]-[DPRSKEQ]-x-G-x(0,1)	x(1,2)-[VI]-[YL]-[VIL]	x-D-G-x
OCR	x(12,15)	x(3,4)	x(5,7)	x(1,2)-[VI]-[YL]-[VIL]	x(4)

	Strand5	Loop8	Strand6	Loop9	Strand7	Loop10
MSA	[IL]-x-[LMV]-x-V-x(2)	x-[GED]-x(2,4)	x(0,1)-[LV]-x(2)-[EQTKED]-[VAI]-x-N	[GSNT]	x(2)-[LI]	[GSK]-[SDEN]-[KRN]-[KR]
SBA	[IL]-x-[LMV]-x-V-x(2)	x-[GDE]-x-[DEQNK]	x-[LV]-x(2)-[EQTKD]-[VAI]-x-N	[GSNT]	x(2)-[LI]	[GSK]-[SDEN]-[KRN]-[KR]
SSS	[IL]-x-[LMV]-x-V-x(2,5)	x(2,5)	x(0,1)-[LV]-x(2)-[EQTKD]-[VAI]-x(0,2)	x(0,2)-[GN]-x(0,1)	x(2,4)	x(0,1)-[SDEN]-[KRN]-[KR]-x(0,1)
OCR	[IL]-x-[LMV]-x-V-x(2,5)	x(2,5)	x(0,1)-[LV]-x(2)-[EQTKD]-[VAI]-x(0,2)	x(1,4)	x(2,4)	x(4,5)

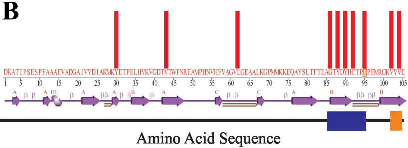
	Strand8	Loop11
MSA	[GN]-[VM]-N	L-P-[GN]-[AC]-x-[VI]-x-L
SBA	[GN]-[VM]-N	L-P-[GN]-[AC]-x-[VI]-x-L
SSS	x(0,1)-[VM]-N	L-P-[GN]-[ACV]-x-[VI]-x(2)
OCR	x(0,1)-[VM]-N	x(8)

L

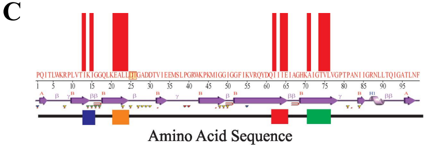
Figure S2: Distribution of OCR^S across the protein sequence. Conserved positions in OCR^S fingerprints of target dataset, **A-L,** are plotted across the entire protein sequence length for easy visualization. The figure shows conserved positions for following protein folds: **A** GFP-like protein, **B** Cupredoxin-like proteins, **C** Acid Proteases, **D** Ribosomal Protein L14, **E** β-Grasp (Ubiquitin-like) protein, **F** Nucleoside Triphosphate Hydrolase, **G** RNAase A-like proteins, **H** Cysteine Proteinases, **I** 50S Ribosomal Protein L25, **J** 50S Ribosomal Protein L6P, **K** Difydrofolate reductase-like proteins, **L** Pyruvate kinase N-terminal domain. Conserved residues are not distributed equally across the sequence length but forms multiple conserved blocks.



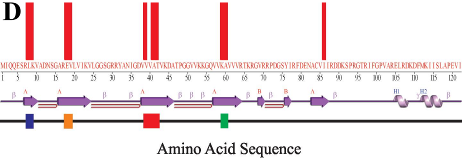
[PRKTE]-x(4)-[LVIM]-x(3)-[LVI]-x(5,12)-[SNEKT]-x(4)-[GS]-x(5,7)-[KRTDH]-x(4)-[FIMAL]-x(40,45)-[YFL]-x(13,14)-[YMLAVC]-x(2)-[RKNHTS]-x-[EQRTDS]-x-[KSTER]-x(8,13)-[RSDNHK]-x-[EKTD]-[LCIVF]-x(28,35)-[IVL]-x(8)-[NTQKS]



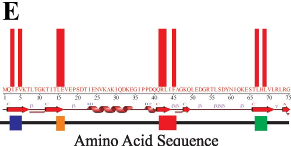
[YF]-x(3,7)-[VI]-x(16,22)-[IVLF]-x(22,38)-G-x-[YF]-x-[YFVLI]-x-C-x(2,4)-H-x(6,9)-[VIL]-[VICA]



`[IVA]-x-[IV]-x(5)-[EKN]-[AVF]-L-L-x(37,42)-[IV]-x-[IL]-[EK]-x(5,10)-[AVI]-x(2)-[TSNR]-[VIM]-[LMC]`

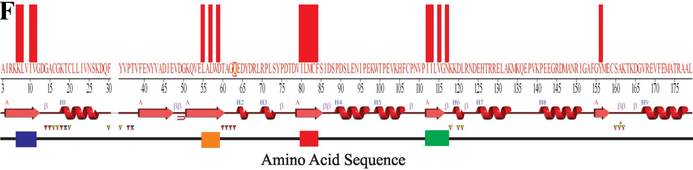


[LIMV]-[KENT]-x(8)-[EKRN]-[VIL]-x(19,23)-[VI]-x-[AIV]-[TS]-x(13,16)[KPEN]-A-x(25)-[IL]

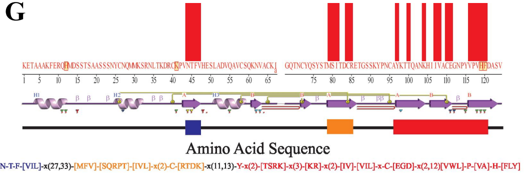


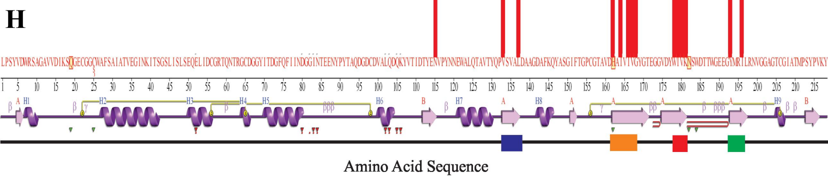
[IL]-x-[VI]-x(8,10)-[LIVF]-[EDNKS]-x(25)[RS]-[LF]-x-[FY]-x(21)-[LVI]-x-[LVA]

F



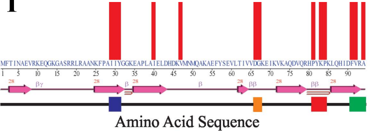
[KR]-LVI-x-[IVLM]-[VLIF]-x(39,43)-[LMF]-x-[LIV]-x-D-x(15,20)-[IFCVL]-[LVI]-[MCLFWY]-[CV]-[FYV]-x(26,31)-[IMFLV]-[IVLM]-[LVI]-[VLF]-x-[NT]-x(34,42)-[YFI]


G
KETAAAKFERQHMDSSTSAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADVQAVCSQKNVACKJ
a
GQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV
1 5 10 15 20 25 30 35 40 45 50 55 60 65 75 80 85 90 95 100 105 110 115 120
H1
H2
H3
β
A
B
Amino Acid Sequence
N-T-F-[VIL]-x(27,33)-[MFV]-[SQRPT]-[IVL]-x(2)-C-[RTDK]-x(11,13)-Y-x(2)-[TSRK]-x(3)-[KR]-x(2)-[IV]-[VIL]-x-C-[EGD]-x(2,12)[VWL]-P-[VA]-H-[FLY]

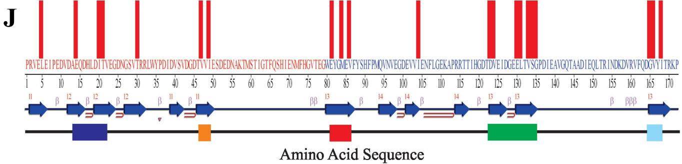


[NRQDES]-x(15,19)-[VI]-x(3)-[LVIF]-x(17,24)-H-x-[IV]-x-[IAVL]-[VL]-G-x(2,9)-[WIY]-[IL]-[VI]-[KRG]-x(11)-[YF]-x(2)-[IMF]

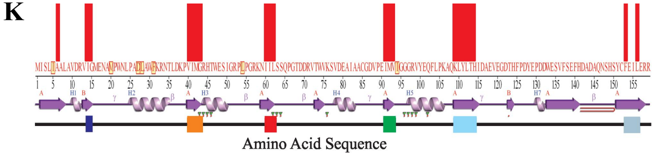
I



[IV]-[IMA]-Y-x(7,8)-[IVF]-x(6,10)-[VA]-x(12,24)-[DK]-[GA]-x(4,13)-[PK]-x-[KR]-[PR]-x(6)-F-[VFY]-x-[ALV]

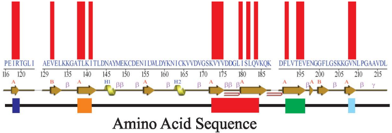


[LMVI]-x(8,10)-[EQTSK]-x(5)-[IV]-x-[TEK]-x(8)-[TKE]-x(15,18)-[VIL]-x-[IVF]-x(29,31)-[ERKTS]-x(2)-[ML]-x-[VFLYI]-x(11,17)-[IVL]-x(16,18)-[DTKE]-[VIAF]-x(5,8)-[ETKR]-[LIV]-x-[VL]-[SDQK]-G-x(25,29)-G-



[AWVC]-x(7,24)-[IL]-G-x(25,69)-[VIML]-[IV]-[MVL]-G-x(16,21)-[IVL]-[IV]-[LIVM]-x(28,40)-[ILVC]-[MVFWY]-[VI]-x(15,17)-[KRTQH]-[LIWM]-[YLFV]-[LVFI]-[TS]-[HRK]-x(40,56)-[FYL]-x(2)-[LWFY]

L



I-R-x(8,11)-[VAI]-x(6)-[TQE]-[LVCF]-x-[IVF]-x(28,31)-[VI]-[YL]-[VIL]-x(4)-[IL]-x-[LMV]-x-V-x(7,9)-[LV]-x(2)-[EQTKD]-[VAI]-x(11)-[VM]-N

Table S1: Selection of representative Immunoglobulin V-set domain for Alignment

Sr. No.	PDB ID	Name	Sequence Length	Sequence identity	rmsd	#Strands (n)
1	1g6vK	Immunoglobulin V-Set Domain	126	100	0.0	13
2	2x1pD	Gelsolin Nanobody	118	68	1.3	12
3	2cjuH	Nq16-113.8 Anti-Phox Antibody	121	58	1.2	11
4	1a7oH	Igg1-Kappa D1.3 Fv (Light Chain)	116	49	1.3	11
5	3dv4B	Antibody Fv Fragment Sag506-01	119	58	1.4	12
6	4kdtA	Nanobody24	129	74	2.0	11
7	3p0gB	Beta-2 Adrenergic Receptor, Lysozyme	121	76	1.1	12
8	2ibzX	Ubiquinol-Cytochrome-C Reductase Complex Core	127	46	1.4	11
9	1g9eA	Antigen-Free Heavy Chain Variable Domain (VHH)	117	75	1.9	13
10	3lrgA	Anti-Huntingtin Vl Domain	114	26	1.8	11

Here, table lists 10 representative Immunoglobulin V-set domain protein structures used for the various alignment approaches. For each candidate, table represents the PDB ID, Sequence length and number of β-strand in structure. Also, sequence and structure deviation w.r.t. first structure, i.e. sequence identity and rmsd are listed for the remaining structural candidates.

Table S2: Minimization of the OCR^S fingerprint size for IgV-set domain

Sequence Pattern	#Hits	TP	FP	EFF
OCR^S(-V^{812})	529	486	43	81
OCR^S(-L^{820})	570	486	84	77
OCR^S(-S^{821})	666	498	168	68
OCR^S(-C^{822})	653	486	167	67
OCR^S(-W^{836})	599	502	97	77
OCR^S(-F^{837})	554	542	12	95
OCR^S(-R^{838})	677	492	185	66
OCR^S(-E^{846})	696	531	185	71
OCR^S(-V^{848})	485	388	97	59
OCR^S(-S^{869})	520	485	35	82
OCR^S(-V^{877})	547	537	35	91
OCR^S(-L^{879})	537	492	45	82
OCR^S(-A^{890})	564	486	78	77
OCR^S(-Y^{892})	564	486	78	77
OCR^S(-C^{894})	564	486	78	77

Here, table lists the fold detection efficiencies of the fingerprints obtained by eliminating the conserved positions, one at a time, from the OCR^S fingerprints for immunoglobulin V-set domain. Conserved residues at specific positions are listed using the 1g6vK as representative structure. First column list the conserved position eliminated from the OCR^S fingerprint. Here, #Hits, TP, FP and EFF indicates total structural hits, true positive hits, false positive hits and fold detection efficiency, respectively.

Table S3: Database Scan results using various fingerprint for Target Dataset

Fold Class and Title	#Fold	MSA			SBA			SSS			OCR			OCR^S			OCR^{MIN}		
		TP	FP	EFF	TP	FP	EFF	TP	FP	EFF	TP	FP	EFF	TP	FP	EFF	TP	FP	EFF
All beta Proteins																			
GFP-like protein	273	9	0	33	12	0	44	268	0	98	271	0	99	273	0	100	273	0	100
Cupredoxin-like proteins	120	0	0	0	29	0	24	42	0	35	117	0	91	120	580	17	120	1	99
Acid Proteases	578	438	0	76	385	0	67	448	0	78	462	0	80	485	0	84	517	0	89
Ribosomal Protein L14	291	190	8	64	210	10	70	218	8	73	248	8	83	260	8	87	286	26	90
Alpha and beta (α+β) Proteins																			
β-Grasp (Ubiquitin-like) protein	400	235	0	58	249	0	62	284	0	70	302	0	75	363	0	90	363	0	90
Nucleoside Triphosphate Hydrolase	530	180	0	34	248	0	47	304	0	57	482	0	91	492	0	93	517	19	94
RNAase A-like proteins	319	266	0	83	257	0	81	280	0	88	296	0	93	312	0	98	312	0	98
Cysteine Proteinases	198	65	0	33	68	0	34	79	0	40	120	0	61	170	0	86	198	23	90
Alpha and beta (α/β) Proteins																			
50S Ribosomal Protein L25	111	96	0	86	98	0	88	98	0	88	111	4	97	111	208	35	111	4	97
50S Ribosomal Protein L6P	317	181	0	57	197	0	62	235	0	74	256	0	81	282	0	89	312	0	98
Difydrofolate reductase-like proteins	272	152	0	56	157	0	57	182	0	67	182	0	67	248	0	91	261	0	96
Pyruvate kinase N-terminal domain	65	31	0	48	39	0	60	9	0	14	31	0	48	61	0	94	65	0	100

Here, table lists the fold detection efficiencies of the respective fingerprints for the target dataset. Here #Fold, TP, FP and EFF indicates total number of homologous structure for each fold, true positive hits, false positive hits and fold detection efficiency, respectively.

Table S4: Summary of the α-helical rich protein dataset

Fold Class and Title	Sequence Length	β-Strands		α-Helix		Loop #res	Min SEQ ID	#FOLD
		#Strand	#res	#Helix	#res			
α-helix rich Proteins								
Globin-like proteins	131~172	0	0	6~9	91~104	~50	14	997
Nuclear Receptor Ligand Binding Domain	211~249	2	6	11~12	~143	~70	16	548
GST C-terminal domain	143~150	2	4	7~8	~125	~25	25	45
Ribosomal Protein S7	139~155	4	12	6	87	~40	42	220

Here, table lists the 4 α-helical rich structurally similar sequence dissimilar fold used in the study. Fold class and title is listed in first column, second column shows the sequence length of representative structures of each fold. For each fold, secondary structure elements information, i.e. total number of residues involved in strand, helix and loop, are listed. Here, #Fold indicates the total number of structural homologs present in the PDB.

Table S5: Database Scan results using various fingerprint for α-helical rich protein

Fold Class and Title	MSA		SBA		OCR		OCR^H		OCR^{MIN}	
	%res	EFF	%res	EFF	%res	EFF	%res	EFF	%res	EFF
α-helix rich Proteins										
Globin-like proteins	22	52	26	52	21	52	16	75	9	82
Nuclear Receptor Ligand Binding Domain	32	69	30	69	27	76	21	96	7	100
GST C-terminal domain	27	44	25	64	25	64	20	80	6	98
Ribosomal Protein S7	58	96	50	92	47	96	35	96	7	100

Here, table lists the percentage of the sequence residues involved in the generated fingerprints as well as the fold detection efficiencies of the respective fingerprints for the dataset of 4 α-helical rich protein. Here %res indicates percentage of the conserved positions used to generate each fingerprints and EFF indicates the fold detection efficiency.

Table S6: Selection of representative Green Fluorescent Proteins for Alignment

Sr. No.	PDB ID	Name	Sequence Length	Sequence identity	rmsd	#Strands (n)
1	1emaA	GFP from *Aequorea victoria*, Q80R	229	100	0.0	11
2	3u0mA	Fluorescent protein mRuby	232	28	1.6	11
3	2dd7A	GFP from marine copepod *Chiridius poppei*	218	17	1.6	11
4	1xqmA	Kindling Fluorescent Protein	232	22	1.8	11
5	2wisB	Killer-red from *Anthomedusae sp*	259	25	1.3	11
6	2rh7B	GFP from *Renilla reniformis*	241	22	1.5	11
7	1g7kA	Fluorescent Protein FP583 from *Discosoma genus*	236	23	1.5	11
8	2g6xB	green fluorescent protein from marine *copepod Pontellina plumata*	219	18.9	1.6	11
9	2p4mA	GFP-like non-fluorescent chromoprotein from *Montipora efflorescens*	221	23	1.7	11
10	3cfaA	GFP-like Fluorescent Protein from *Anemonia sulcata*	169	18	1.8	8

Here, table lists 10 representative Green Fluorescent Protein structures used for the various alignment approaches. For each candidate, table represents the PDB ID, Sequence length and number of β-strand in structure. Also, sequence and structure deviation w.r.t. first structure, i.e. sequence identity and rmsd are listed for the remaining structural candidates.

Table S7: Multiple sequence alignment of GFP-like proteins using ClustalW

PDB-ID		Strand												Loop			Strand2													Loop			
1EMA		V	P	I	L	V	E	L	D	G	D	V		N	G		H	K	F	S	V	S	G	E	G	E	G	D		A	T	Y	G
3U0M		M	R	M	K	V	V	L	V	A	S	V		N	G		H	Q	R	K	C	T	G	E	G	E	N	G		P	Y	M	G
2DD7		F	K	I	E	S	R	I	H	G	N	L		N	G		E	K	F	E	L	V	G	G	G	V	G	-		-	E	E	G
1XQM		M	P	F	K	T	T	I	E	G	T	V		N	G		H	C	F	K	C	I	G	K	G	E	G	N		P	F	E	G
2WIS		M	T	F	K	I	F	I	D	G	E	V		N	G		Q	K	F	T	I	V	A	D	G	S	S	K		F	P	H	G
2RH7		M	P	T	K	I	N	L	E	G	L	V		G	D		H	A	F	S	M	E	G	V	G	E	G	N		I	L	E	G
1G7K		M	R	F	K	V	R	M	E	G	T	V		N	G		H	E	F	E	I	E	G	E	G	E	G	R		P	Y	E	G
2G6X		M	E	I	E	C	R	I	T	G	T	L		N	G		V	E	F	E	L	V	G	G	G	E	G	T		P	E	Q	G
2P4M		-	-	-	-	-	-	-	-	-	-	-		-	-		-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-
3CFA		M	T	Y	K	V	Y	M	S	G	T	V		N	G		H	Y	F	E	V	E	G	D	G	K	G	K		P	Y	E	G

PDB-ID		Strand3									Loop																						
1EMA		K	L	T	L	K	F	I	C		T	T	G	-	K	L	P	V	P	W	P	T	L	V	T	T	F	G	Y	G	V	Q	C
3U0M		T	Q	T	M	R	I	K	V		G	S	S	-	Q	V	P	V	V	F	…	Y	V	V	M	N	H	G	R	S	Y	T	A
2DD7		R	L	E	I	E	M	K	T		K	D	K	-	P	L	A	F	S	P	F	L	L	S	H	C	M	G	Y	G	F	Y	H
1XQM		T	Q	E	M	K	I	E	V		I	E	G	G	P	L	P	F	A	F	H	I	L	S	T	S	C	M	Y	G	S	K	T
2WIS		D	F	N	V	H	A	V	C		E	T	G	-	K	L	P	M	S	W	K	P	I	C	H	L	I	Q	Y	G	E	P	F
2RH7		T	Q	E	V	K	I	S	V		T	K	G	A	P	L	P	F	A	F	D	I	V	S	V	A	F	G	Y	G	N	R	A
1G7K		H	N	T	V	K	L	K	V		T	K	G	G	P	L	P	F	A	W	D	I	L	S	P	Q	F	Q	Y	G	S	K	V
2G6X		R	M	T	N	K	M	K	S		T	K	G	-	A	L	T	F	S	P	Y	L	L	S	H	V	M	G	Y	G	F	Y	H
2P4M		-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	Y	G	S	K	T
3CFA		E	Q	T	V	K	L	T	V		T	K	G	G	P	L	P	F	A	W	D	I	L	S	P	Q	C	Q	Y	G	S	I	P

PDB-ID		Loop																							Strand4								
1EMA		F	S	R	Y	P	D	H	M	K	R	H	D	F	F	K	S	A	M	P	E	G	-		Y	V	Q	E	R	T	I	F	F
3U0M		I	S	T	I	P	E	T	V	G	-	Y	S	L	L	P	L	A	P	I	G	G	I		F	T	W	M	F	V	T	R	Q
2DD7		F	A	S	F	P	K	G	T	K	-	-	N	I	Y	L	H	A	A	T	N	G	G		Y	T	N	T	R	K	E	I	Y
1XQM		F	I	K	Y	V	S	G	I	P	-	-	D	Y	F	K	Q	S	F	P	E	G	-		F	T	W	E	R	T	T	T	Y
2WIS		F	A	R	Y	P	D	G	I	S	-	-	H	F	A	Q	E	C	F	P	E	G	-		L	S	I	D	R	T	V	R	F
2RH7		Y	T	G	Y	P	E	E	I	S	-	-	D	Y	F	L	Q	S	F	P	E	G	-		F	T	Y	E	R	N	I	R	Y
1G7K		Y	V	K	H	P	A	D	I	P	-	-	D	Y	K	K	L	S	F	P	E	G	-		F	K	W	E	R	V	M	N	F
2G6X		F	G	T	Y	P	S	G	Y	E	-	-	N	P	F	L	H	A	I	N	N	G	G		Y	T	N	T	R	I	E	K	Y
2P4M		F	I	K	Y	V	S	G	I	P	-	-	D	Y	F	K	Q	S	F	P	E	G	-		F	T	W	E	R	T	T	T	Y
3CFA		F	T	K	Y	P	E	D	I	P	-	-	D	Y	V	K	Q	S	F	P	E	G	-		F	T	W	E	R	I	M	N	F

PDB-ID		Loop						Strand5												Loop			Strand6										
1EMA		K	-	D	D	G		N	Y	K	T	R	A	E	V	K	F	E		G	D		T	L	V	N	R	I	E	L	K	G	I
3U0M		E	…	D	G	G		V	I	T	V	Q	Q	D	T	S	L	L		D	G		C	L	V	Y	H	Q	Q	V	R	G	V
2DD7		E	-	D	G	G		I	L	E	V	N	F	R	Y	T	Y	E		F	N		K	I	I	G	D	V	E	C	I	G	H
1XQM		E	-	D	G	G		F	L	T	A	H	Q	D	T	S	L	D		G	D		C	L	V	Y	K	V	K	I	L	G	N
2WIS		E	-	N	D	G		T	M	T	S	H	H	T	Y	E	L	D		D	T		C	V	V	S	R	I	T	V	N	C	D
2RH7		Q	-	D	G	G		T	A	I	V	K	S	D	I	S	L	E		D	G		K	F	I	V	N	V	D	F	K	A	K
1G7K		E	-	D	G	G		V	V	T	V	T	Q	D	S	S	L	Q		D	G		C	F	I	Y	K	V	K	F	I	G	V
2G6X		E	-	D	G	G		V	L	H	V	S	F	S	Y	R	Y	E		A	G		R	V	I	G	D	F	K	V	M	G	T
2P4M		E	-	D	G	G		F	L	T	A	H	Q	D	T	S	L	D		G	D		C	L	V	Y	K	V	K	I	L	G	N
3CFA		E	-	D	G	A		V	C	T	V	S	N	D	S	S	I	Q		G	N		C	F	T	Y	H	V	K	F	S	G	L

PDB-ID		Loop																		Strand7									Loop		
1EMA		D	F	K	E	D	G	N	I	L	G	H	K	L	E	Y	…	S		H	N	V	Y	I	M	A	D		K	Q	K
3U0M		P	S	N	…	E	G	A	V	M	Q	I	K	T	K	G	…	N		R	G	Y	T	H	M	A	L		D	G	G
2DD7		G	F	P	S	Q	S	P	I	F	K	D	T	I	V	K	S	C		P	T	V	D	L	M	L	P		M	S	G
1XQM		N	F	P	A	D	G	P	V	M	Q	N	K	V	G	R	W	E		P	G	T	E	I	V	Y	E		V	D	G
2WIS		G	F	Q	P	D	G	P	I	M	R	D	Q	L	V	D	I	L		P	N	E	T	H	M	F	P		H	G	P
2RH7		D	L	R	R	M	G	P	V	M	Q	Q	D	I	V	G	M	Q		P	S	Y	E	S	M	Y	T		N	V	T
1G7K		N	F	P	S	D	G	P	V	M	Q	K	K	T	M	G	W	E		A	S	T	E	R	L	Y	P		R	D	G
2G6X		G	F	P	E	D	S	V	I	F	T	D	K	I	I	R	S	N		A	T	V	E	H	L	H	P		M	G	D
2P4M		N	F	P	A	D	G	P	V	M	Q	N	K	A	G	G	W	E		P	S	C	E	I	L	Y	E		V	D	G
3CFA		N	F	P	P	N	G	P	V	M	Q	K	K	T	Q	G	W	E		P	S	S	E	R	L	F	A		R	G	G

PDB-ID		Strand8												Loop						Strand9											
1EMA		G	I	K	V	N	F	K	I	R	H	N		I	E	D	G	S		V	Q	L	A	D	H	Y	Q	Q	N	T	P
3U0M		G	L	G	R	Y	F	V	T	T	Y	R		K	M	P	G	…		G	H	P	S	H	T	H	I	Y	Q	W	R
2DD7		I	I	A	S	S	Y	A	R	A	F	Q		L	K	D	G	…		A	E	V	K	N	N	I	D	F	K	N	P
1XQM		V	L	R	G	Q	S	L	M	A	L	K		C	P	G	G	…		C	H	L	H	T	T	Y	R	S	K	K	P
2WIS		A	V	R	Q	L	A	F	I	G	F	T		T	A	D	G	…		G	H	F	D	S	K	M	T	F	N	G	S
2RH7		S	V	I	G	E	C	I	I	A	F	K		L	Q	T	G	…		Y	H	M	R	T	V	Y	K	S	K	K	P
1G7K		V	L	K	G	E	I	H	K	A	L	K		L	K	D	G	…		V	E	F	K	S	I	Y	M	A	K	K	P
2G6X		D	L	D	G	S	F	T	R	T	F	S		L	R	D	G	…		S	V	V	D	S	H	M	H	F	K	S	A
2P4M		V	L	C	G	Q	S	L	M	A	L	K		C	P	G	G	…		C	R	L	H	T	T	Y	R	S	K	K	P
3CFA		M	L	I	G	N	N	F	M	A	L	K		L	E	G	G	…		C	E	F	K	T	T	Y	K	A	K	K	P

PDB-ID		Loop											Strand10											Loop							
1EMA		I	…	G	P	V	L	L	P	D	N		H	Y	L	S	T	Q	S	A	L	S		K	D	P	N	E	K	R	D
3U0M		K	…	T	V	G	N	I	A	H	D		H	A	V	S	H	R	L	E	R	L		-	-	S	V	N	E	M	-
2DD7		I	-	S	F	S	K	S	G	-	P		M	F	T	H	R	R	V	E	E	T		H	T	K	E	N	L	A	M
1XQM		A	-	S	A	L	K	M	P	G	F		H	F	E	D	H	R	I	E	I	M		E	E	V	E	K	G	K	C
2WIS		R	-	-	A	I	E	I	P	G	P		H	F	V	T	I	I	T	K	Q	M		R	D	T	S	D	K	R	D
2RH7		-	-	-	V	E	T	M	P	L	Y		H	F	I	Q	H	R	L	V	K	T		N	V	D	T	A	S	G	Y
1G7K		-	-	-	-	V	Q	L	P	G	Y		Y	Y	V	D	S	K	L	D	I	T		S	H	N	E	D	Y	T	I
2G6X		I	…	S	I	L	Q	N	G	G	P		M	F	A	F	R	R	V	E	E	D		H	S	N	T	E	L	G	I
2P4M		A	-	S	A	L	K	M	P	E	F		H	F	E	D	H	R	I	E	V	K		-	E	V	Q	K	G	K	H
3CFA		-	-	-	-	V	K	M	P	G	Y		H	Y	V	D	R	K	L	D	V	T		N	H	N	K	D	Y	T	S

PDB-ID		Strand11										
1EMA		H	M	V	L	L	E	F	V	T	A	A
3U0M		F	V	V	Q	R	E	H	A	V	A	K
2DD7		V	E	Y	Q	Q	V	F	N	S	A	P
1XQM		Y	K	Q	Y	E	A	A	V	G	R	Y
2WIS		H	V	C	Q	R	E	V	A	Y	A	H
2RH7		V	V	Q	H	E	T	A	I	A	A	H
1G7K		V	E	Q	Y	E	R	T	E	G	R	H
2G6X		V	E	Y	Q	H	A	F	K	T	P	D
2P4M		Y	E	Q	Y	E	A	A	V	A	R	Y
3CFA		V	E	Q	C	E	I	S	I	A	R	K

Here, PDB code of proteins is provided in the first column, chain name – in the 2nd column (‘chain’), and the beginning and end position of each sequence in a domain – in the 3rd and 4th (‘start’ and ‘end’) columns, respectively. The sequences of the longer loops are not shown in their entirety, and the ‘skipped’ segments of loops are indicated by ellipses (“…”). The spaces within the sequences separate between secondary-structure units (strands and loops). The conserved hydrophobic and hydrophilic positions are shown in blue and green respectively.

Table S8: Structure based sequence alignment of GFP-like proteins using Dali server

PDB-ID	Chain	Start	End		Strand1												Loop			Strand2											
1EMA	A	12	227		V	P	I	L	V	E	L	D	G	D	V		N	G		H	K	F	S	V	S	G	E	G	E	G	D
3U0M	A	4	222		M	R	M	K	V	V	L	E	G	S	V		F	G		H	Q	F	K	C	T	G	E	G	E	V	N
2DD7	A	4	212		F	K	I	E	S	R	I	H	G	N	L		N	G		E	K	F	E	L	V	G	G	G	V	G	E
1XQM	A	9	220		M	P	F	K	T	T	I	E	G	T	V		N	G		H	C	F	K	C	I	G	K	G	E	G	N
2WIS	B	12	223		M	T	F	K	I	F	I	D	G	E	V		N	G		Q	K	F	T	I	V	A	D	G	S	S	K
2RH7	B	11	221		M	P	T	K	I	N	L	E	G	L	V		G	D		H	A	F	S	M	E	G	V	G	E	G	N
1G7K	A	12	220		M	R	F	K	V	R	M	E	G	T	V		N	G		H	E	F	E	I	E	G	E	G	E	G	R
2G6X	B	4	215		M	E	I	E	C	R	I	T	G	T	L		N	G		V	E	F	E	L	V	G	G	G	E	G	T
2P4M	A	12	220		-	-	-	-	-	-	-	-	-	-	-		-	-		-	-	-	-	-	-	-	-	-	-	-	-
3CFA	A	88	219		M	T	Y	K	V	Y	M	S	G	T	V		N	G		H	Y	F	E	V	E	G	D	G	K	G	K

PDB-ID		Loop					Strand3									Loop																	
1EMA		A	T	Y	G		K	L	T	L	K	F	I	C		…	L	P	V	P	W	P	T	L	V	T	T	F	V	Q	C	F	S
3U0M		P	Y	M	G		T	Q	Y	M	A	I	K	V		…	P	L	-	P	F	A	F	A	V	D	I	L	A	T	S	F	M
2DD7		E	-	-	G		R	L	E	I	E	M	K	T		…	L	A	F	S	P	F	L	L	S	H	C	M	F	Y	H	F	A
1XQM		P	F	E	G		T	Q	E	M	K	I	E	V		…	L	P	F	A	F	H	I	L	S	T	S	C	S	K	T	F	I
2WIS		F	P	H	G		D	F	N	V	H	A	V	C		…	L	P	M	S	W	K	P	I	C	H	L	I	E	P	F	F	A
2RH7		I	L	E	G		T	Q	E	V	K	I	S	V		…	L	P	F	A	F	D	I	V	S	V	A	F	N	R	A	Y	T
1G7K		P	Y	E	G		H	N	T	V	K	L	K	V		…	L	P	F	A	W	D	I	L	S	P	Q	F	S	K	V	Y	V
2G6X		P	E	Q	G		R	M	T	N	K	M	K	S		…	L	T	F	S	P	Y	L	L	S	H	V	M	F	Y	H	F	G
2P4M		-	-	-	-		-	-	-	-	-	-	-	-		…	-	-	-	-	-	-	-	-	-	-	-	-	S	K	T	F	I
3CFA		P	Y	E	G		E	Q	T	V	K	L	T	V		…	L	P	F	A	W	D	I	L	S	P	Q	C	S	I	P	F	T

PDB-ID																				Strand4										Loop			
1EMA		R	Y	P	D	H	M	…	D	F	F	K	S	A	M	P	E	G		Y	V	Q	E	R	T	I	F	F		K	D	D	G
3U0M		V	R	T	F	K	Y	-	K	G	G	P	D	F	G	G	-	E		F	T	W	E	A	V	T	R	Y		L	D	G	G
2DD7		S	F	P	K	G	T	…	N	I	Y	L	H	A	A	T	G	G		Y	T	N	T	R	K	E	I	Y		E	D	G	G
1XQM		K	Y	V	S	G	I	…	D	Y	F	K	Q	S	F	P	E	G		F	T	W	E	R	T	T	T	Y		E	D	G	G
2WIS		R	Y	P	D	G	I	…	H	F	A	Q	E	C	F	P	E	G		L	S	I	D	R	T	V	R	F		E	N	D	G
2RH7		G	Y	P	E	E	I	…	D	Y	F	L	Q	S	F	P	E	G		F	T	Y	E	R	N	I	R	Y		Q	D	G	G
1G7K		K	H	P	A	D	I	…	D	Y	K	K	L	S	F	P	E	G		F	K	W	E	R	V	M	N	F		E	D	G	G
2G6X		T	Y	P	S	G	Y	…	N	P	F	L	H	A	I	N	G	G		Y	T	N	T	R	I	E	K	Y		E	D	G	G
2P4M		K	Y	V	S	G	I	…	D	Y	F	K	Q	S	F	P	E	G		F	T	W	E	R	T	T	T	Y		E	D	G	G
3CFA		K	Y	P	E	D	I	…	D	Y	V	K	Q	S	F	P	E	G		F	T	W	E	R	I	M	N	F		E	D	G	A

PDB-ID		Strand5												Loop			Strand6												Loop				
1EMA		N	Y	K	T	R	A	E	V	K	F	E		G	D		T	L	V	N	R	I	E	L	K	G	I		…	L	G	H	K
3U0M		V	I	T	V	K	Q	D	T	S	G	I		D	G		C	L	V	Y	H	T	Q	V	R	G	V		…	N	F	P	-
2DD7		I	L	E	V	N	F	R	Y	T	Y	E		F	N		K	I	I	G	D	V	E	C	I	G	H		…	F	K	D	T
1XQM		F	L	T	A	H	Q	D	T	S	L	D		G	D		C	L	V	Y	K	V	K	I	L	G	N		…	M	Q	N	K
2WIS		T	M	T	S	H	H	T	Y	E	L	D		D	T		C	V	V	S	R	I	T	V	N	C	D		…	M	R	D	Q
2RH7		T	A	I	V	K	S	D	I	S	L	E		D	G		K	F	I	V	N	V	D	F	K	A	K		…	M	Q	Q	D
1G7K		V	V	T	V	T	Q	D	S	S	L	Q		D	G		C	F	I	Y	K	V	K	F	I	G	V		…	M	Q	K	K
2G6X		V	L	H	V	S	F	S	Y	R	Y	E		A	G		R	V	I	G	D	F	K	V	M	G	T		…	F	T	D	K
2P4M		F	L	T	A	H	Q	D	T	S	L	D		G	D		M	L	V	Y	K	V	K	I	L	G	N		…	M	Q	N	K
3CFA		V	C	T	V	S	N	D	S	S	I	Q		G	N		C	F	T	N	R	I	E	L	K	G	I		…	L	G	H	K

PDB-ID		Loop								Strand7									Loop				Strand8										
1EMA		L	E	Y	N	Y	N	S		H	N	V	Y	I	M	A	D		K	…	N		G	I	K	V	N	F	K	I	R	H	N
3U0M		-	S	N	G	A	V	M		Y	T	E	M	M	H	P	A		D	-	G		G	L	R	G	Y	T	H	M	A	L	K
2DD7		I	V	-	K	S	C	P		T	V	D	L	M	L	P	M		S	…	N		I	I	A	S	S	Y	A	R	A	F	Q
1XQM		V	G	-	R	W	E	P		G	T	E	I	V	Y	E	V		D	-	G		V	L	R	G	Q	S	L	M	A	L	K
2WIS		L	V	-	D	I	L	P		N	E	T	H	M	F	P	H		G	…	N		A	V	R	Q	L	A	F	I	G	F	T
2RH7		I	V	-	G	M	Q	P		S	Y	E	S	M	Y	T	N		V	-	T		S	V	I	G	E	C	I	I	A	F	K
1G7K		T	M	-	G	W	E	A		S	T	E	R	L	Y	P	R		D	-	G		V	L	K	G	E	I	H	K	A	L	K
2G6X		I	I	-	R	S	N	A		T	V	E	H	L	H	P	M		G	…	N		D	L	D	G	S	F	T	R	T	F	S
2P4M		A	G	-	G	W	E	P		S	C	E	I	L	Y	E	V		D	-	G		V	L	C	G	Q	S	L	M	A	L	K
3CFA		L	E	Y	N	Y	N	S		H	N	V	R	L	F	A	R		G	-	G		M	L	I	G	N	N	F	M	A	L	K

PDB-ID		Loop						Strand9													Loop										
1EMA		I	E	D	G	S		V	Q	L	A	D	H	Y	Q	Q	N	T	P		I	G	D	G	P	V	L	L	P	D	N
3U0M		V	D	G	G	G		G	H	L	S	C	S	F	Q	T	T	Y	R		-	K	T	V	G	-	-	N	I	K	M
2DD7		L	K	D	G	S		F	Y	T	A	E	V	K	N	N	I	D	F		K	-	-	N	P	I	H	S	G	-	P
1XQM		C	P	G	G	R		H	L	T	C	H	L	H	T	T	Y	R	S		K	K	P	S	A	L	K	M	P	G	F
2WIS		T	A	D	G	G		L	M	M	G	H	F	D	S	K	M	T	F		N	G	S	R	A	I	E	I	P	G	P
2RH7		L	Q	T	G	K		H	F	T	Y	H	M	R	T	V	Y	K	S		K	K	-	P	V	E	T	M	P	L	Y
1G7K		L	K	D	G	G		H	Y	L	V	E	F	K	S	I	Y	M	A		K	K	-	-	P	V	Q	L	P	G	Y
2G6X		L	R	D	G	G		Y	Y	S	S	V	V	D	S	H	M	H	F		K	S	-	-	A	I	H	N	G	G	P
2P4M		C	P	G	G	R		H	L	N	C	R	L	H	T	T	Y	R	S		K	K	P	S	A	L	K	M	P	G	Y
3CFA		L	E	G	G	G		H	Y	L	C	E	F	K	T	T	Y	K	A		K	K	-	-	P	V	K	M	P	E	F

PDB-ID		Strand10											Loop					Strand11										
1EMA		H	Y	L	S	T	Q	S	A	L	S		K	D	…	D		H	M	V	L	L	E	F	V	T	A	A
3U0M		H	Q	V	S	H	R	L	E	R	L		Y	D	…	N		F	V	Q	Q	R	L	H	A	V	A	K
2DD7		M	F	T	H	R	R	V	E	E	T		H	-	-	E		N	L	A	M	V	E	Y	Q	Q	V	F
1XQM		H	F	E	D	H	R	I	E	I	M		E	E	…	K		C	Y	K	Q	Y	E	A	A	V	G	R
2WIS		H	F	V	T	I	I	T	K	Q	M		R	D	…	D		H	V	C	Q	R	E	V	A	Y	A	H
2RH7		H	F	I	Q	H	R	L	V	K	T		N	V	-	-		Y	V	V	Q	H	E	T	A	I	A	A
1G7K		Y	Y	V	D	S	K	L	D	I	T		S	H	…	T		I	V	E	Q	Y	E	R	T	E	G	R
2G6X		M	F	A	F	R	R	V	E	E	D		H	-	-	T		E	L	G	I	V	E	Y	Q	H	A	F
2P4M		H	F	E	D	H	R	I	E	V	K		E	V	…	K		H	Y	E	Q	Y	E	A	A	V	A	R
3CFA		H	Y	V	D	R	K	L	D	V	T		N	H	…	Y		S	V	E	Q	C	E	I	S	I	A	R

Here, PDB code of proteins is provided in the first column, chain name – in the 2nd column (‘chain’), and the beginning and end position of each sequence in a domain – in the 3rd and 4th (‘start’ and ‘end’) columns, respectively. The sequences of the longer loops are not shown in their entirety, and the ‘skipped’ segments of loops are indicated by ellipses (“…”). The spaces within the sequences separate between secondary-structure units (strands and loops). The conserved hydrophobic and hydrophilic positions are shown in blue and green respectively.

Table S9: SSS-based sequence alignment of GFP like proteins

PDB-ID	Chain	Start	End		Strand1														Loop				Strand2									
1EMA	A	12	227		-	V	P	I	L	V	E	L	D	G	D	V	-		N	G	-		…	F	S	V	S	G	E	G	E	G
3U0M	A	4	222		N	M	R	M	K	V	V	L	E	G	S	V	…		N	G	…		…	H	N	F	K	C	T	S	E	V
2DD7	A	4	212		-	F	K	I	E	S	R	I	H	G	N	L	-		N	G	-		…	F	E	L	V	G	G	G	V	G
1XQM	A	9	220		-	M	P	F	K	T	T	I	E	G	T	V	-		N	G	-		…	F	K	C	I	G	K	G	E	G
2WIS	B	12	223		-	M	T	F	K	I	F	I	D	G	E	V	-		N	G	-		…	F	T	I	V	A	D	G	S	S
2RH7	B	11	221		V	M	P	T	K	I	N	L	E	G	L	V	-		G	D	-		…	F	S	M	E	G	V	G	E	G
1G7K	A	12	220		-	M	R	F	K	V	R	M	E	G	T	V	-		N	G	-		…	F	E	I	E	G	E	G	E	G
2G6X	B	4	215		-	M	E	I	E	C	R	I	T	G	T	L	-		N	G			…	F	E	L	V	G	G	G	E	G
2P4M	A	12	220		-	M	T	Y	K	V	Y	M	S	G	T	V	-		N	G	-		…	F	E	V	E	G	D	G	K	G
3CFA	A	88	219		-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-		…	-	-	-	-	-	-	-	-	-

PDB-ID		Loop			Strand3							Loop																						
1EMA		Y	G		K	L	T	L	K	F		T	T	G	K	L	P	V	P	…	F	G	Y	G	…	F	S	R	Y	P	D	H	M	K	R
3U0M		M	G		T	Q	Y	M	A	I		-	G	P	L	P	F	A	F	…	Y	P	I	G	…	F	K	W	S	F	P	E	G	K	W
2DD7		E	G		R	L	E	I	E	M		K	D	K	P	L	A	F	S	…	M	G	Y	G	…	F	A	S	F	P	K	G	T	K	N
1XQM		E	G		T	Q	E	M	K	I		-	G	G	P	L	P	F	A	…	C	M	Y	G	…	F	I	K	Y	V	S	G	I	P	D
2WIS		H	G		D	F	N	V	H	A		E	T	G	K	L	P	M	S	…	I	Q	Y	G	…	F	A	R	Y	P	D	G	I	S	H
2RH7		E	G		T	Q	E	V	K	I		-	G	A	P	L	P	F	A	…	F	G	Y	G	…	Y	T	G	Y	P	E	E	I	S	D
1G7K		E	G		H	N	T	V	K	L		-	G	G	P	L	P	F	A	…	F	Q	Y	G	…	Y	V	K	H	P	A	D	I	P	D
2G6X		Q	G		R	M	T	N	K	M		-	K	G	A	-	-	L	T	…	M	G	Y	G	…	F	G	T	Y	P	S	G	Y	E	-
2P4M		E	G		E	Q	T	V	K	L		-	G	G	P	L	P	F	A	…	-	Q	Y	G	…	F	T	K	Y	P	E	D	I	P	D
3CFA		-	-		-	-	-	-	-	-		-	-	-	-	-	-	-	-	...	C	M	Y	G	…	F	I	K	Y	V	S	G	I	P	D

PDB-ID		Loop						Strand4											Loop						Strand5									
1EMA		…	P	E	G	-		-	Y	V	Q	E	R	T	I	F	F		K	D	D	G	-		N	Y	K	T	R	A	E	V	K	…
3U0M		…	-	T	G	G		E	F	T	E	E	A	V	T	R	Y		…	E	P	G	G		C	L	V	Y	H	A	Q	V	R	-
2DD7		…	T	N	G	G		-	Y	T	N	T	R	R	E	I	Y		-	E	D	G	G		I	L	E	V	N	F	R	Y	T	…
1XQM		…	P	E	G	-		-	F	T	W	E	R	T	T	T	Y		-	E	D	G	G		F	L	T	A	H	Q	D	T	S	…
2WIS		…	P	E	G	-		-	L	S	I	D	R	T	V	R	F		E	N	D	G	-		T	M	T	S	H	H	T	Y	E	…
2RH7		…	P	E	G	-		-	F	T	Y	E	R	N	I	R	Y		-	Q	D	G	G		T	A	I	V	K	S	D	I	S	…
1G7K		…	P	E	G	-		-	F	K	W	E	R	V	M	N	F		-	E	D	G	G		V	V	T	V	T	Q	D	S	S	…
2G6X		…	N	N	G	G			Y	T	N	T	R	I	E	K	Y		-	E	D	G	G		V	L	H	V	S	F	S	Y	R	…
2P4M		…	P	E	G	-		-	F	T	W	E	R	I	M	N	F		-	E	D	G	A		V	C	T	V	S	N	D	S	S	…
3CFA		…	P	E	G	-		-	F	Y	W	E	R	T	T	T	Y		-	E	D	G	G		F	L	T	A	H	Q	D	T	S	-

PDB-ID		Loop				Strand6										Loop																		
1EMA		G	D	-		…	R	I	E	L	K	G	I	-		D	F	K	E	D	G	N	I	L	G	H	K	L	E	Y	N	Y	N	S
3U0M		…	D	…		-	N	T	E	M	M	Y	P	A		-	-	-	-	-	-	P	Q	I	P	Y	G	A	S	V	H	I	E	P
2DD7		F	N	-		…	D	V	E	C	I	G	H	-		G	F	P	S	Q	S	P	I	F	K	D	T	I	-	V	K	S	C	G
1XQM		G	D	-		…	K	V	K	I	L	G	N	-		N	F	P	A	D	G	P	V	M	Q	N	K	V	-	G	R	W	E	N
2WIS		-	D	T		…	R	I	T	V	N	C	D	-		G	F	Q	P	D	G	P	I	M	R	D	Q	L	-	V	D	I	L	P
2RH7		E	D	…		…	N	V	D	F	K	A	K	-		D	L	R	R	M	G	P	V	M	Q	Q	D	I	-	V	G	M	Q	D
1G7K		-	D	G		…	K	V	K	F	I	G	V	-		N	F	P	S	D	G	P	V	M	Q	K	K	T	-	M	G	W	E	A
2G6X		-	-	A		…	D	F	K	V	M	G	T	-		G	F	P	E	D	S	V	I	F	T	D	K	I	I	R	S	N	A	-
2P4M		G	N	-		…	H	V	K	F	S	G	L	-		N	F	P	P	N	G	P	V	M	Q	K	K	T	-	Q	G	W	E	P
3CFA		L	D	G		…	K	V	K	I	L	G	N	-		N	F	P	A	D	G	P	V	M	Q	N	K	A	-	G	G	W	E	N

PDB-ID		Strand7									Loop				Strand8													Loop				
1EMA		H	N	V	Y	I	M	A	D		K	Q	…		G	I	K	V	N	F	K	I	R	H	N	-		…	D	G	S	-
3U0M		Y	T	E	L	Y	H	Y	S		D	G	-		G	L	R	G	Y	T	H	M	A	L	K	…		…	D	G	G	…
2DD7		T	V	D	L	M	L	P	M		S	G	-		I	I	A	S	S	Y	A	R	A	F	Q	L		K	D	G	S	-
1XQM		G	T	E	I	V	Y	E	V		D	G	-		V	L	R	G	Q	S	L	M	A	L	K	C		-	P	G	G	-
2WIS		N	E	T	H	M	F	P	H		-	G	…		A	V	R	Q	L	A	F	I	G	F	T	T		A	D	G	G	-
2RH7		S	Y	E	S	M	Y	T	N		V	T	-		S	V	I	G	E	C	I	I	A	F	K	L		Q	T	G	K	-
1G7K		S	T	E	R	L	Y	P	R		D	G	-		V	L	K	G	E	I	H	K	A	L	K	L		K	D	G	G	-
2G6X		T	V	E	H	L	H	P	-		M	G	…		D	L	D	G	S	F	T	R	T	F	S	L		R	D	-	-	-
2P4M		S	S	E	R	L	F	A	R		-	G	G		M	L	I	G	N	N	F	M	A	L	K	L		E	-	G	G	G
3CFA		S	C	E	I	L	Y	E	V		D	G	-		V	L	C	G	Q	S	L	M	A	L	K	C		-	P	G	G	-

PDB-ID		Strand9														Loop										Strand10									
1EMA		-	V	Q	L	A	D	H	Y	Q	Q	N	T	P		…	G	P	V	…	L	P	D	N		H	Y	L	S	T	Q	S	A	L	S
3U0M		G	H	L	S	C	S	F	V	T	T	Y	R	-		…	G	N	I	-	M	P	G	-		…	H	A	V	S	H	R	L	E	…
2DD7		-	F	Y	T	A	E	V	K	N	N	I	D	F		K	N	P	I	…	K	S	G	P		M	F	T	H	R	R	V	E	E	…
1XQM		R	H	L	T	C	H	L	H	T	T	Y	R	S		K	K	P	A	…	M	P	G	H		H	F	E	D	H	R	I	E	I	…
2WIS		-	L	M	M	G	H	F	D	S	K	M	T	F		…	R	A	I	…	I	P	G	P		H	F	V	T	I	I	T	K	Q	M
2RH7		-	H	F	T	Y	H	M	R	T	V	Y	K	S		K	K	P	V	…	M	P	L	Y		H	F	I	Q	H	R	L	V	K	…
1G7K		-	H	Y	L	V	E	F	K	S	I	Y	M	A		K	K	P	V	…	L	P	G	Y		Y	Y	V	D	S	K	L	D	I	…
2G6X		-	G	G	Y	Y	S	S	V	V	D	S	H	…		…	H	P	S	…	G	P	-	-		M	F	A	F	R	R	V	E	E	…
2P4M		-	H	Y	L	C	E	F	K	T	T	Y	K	A		K	K	P	V	…	M	P	G	Y		H	Y	V	D	R	K	L	D	V	…
3CFA		R	H	L	N	C	R	L	H	T	T	Y	R	S		K	K	P	A	…	M	P	E	F		H	F	E	D	H	R	I	E	V	…

PDB-ID		Loop					Strand11										
1EMA		…	K	R	D		H	M	V	L	L	E	F	V	T	A	A
3U0M		-	N	E	…		F	V	V	Q	R	L	H	A	V	A	K
2DD7		T	K	E	-		N	L	A	M	V	E	Y	Q	Q	V	F
1XQM		-	K	G	K		C	Y	K	Q	Y	E	A	A	V	G	R
2WIS		…	K	R	D		H	V	C	Q	R	E	V	A	Y	A	H
2RH7		…	G	Y	-		-	-	V	Q	H	E	T	A	I	A	A
1G7K		N	E	D	…		I	V	E	Q	Y	E	R	T	E	G	R
2G6X		…	T	-	-		E	L	G	I	V	E	Y	Q	H	A	F
2P4M		N	K	D	…		S	V	E	Q	C	E	I	S	I	A	R
3CFA		-	K	G	K		H	Y	E	Q	Y	E	A	A	V	A	R

Here, PDB code of proteins is provided in the first column, chain name – in the 2nd column (‘chain’), and the beginning and end position of each sequence in a domain – in the 3rd and 4th (‘start’ and ‘end’) columns, respectively. The sequences of the longer loops are not shown in their entirety, and the ‘skipped’ segments of loops are indicated by ellipses (“…”). The spaces within the sequences separate between secondary-structure units (strands and loops). The conserved hydrophobic and hydrophilic positions are shown in blue and green respectively.

Table S10: Fold detection results using OCR^S fingerprints for GFP-like protein against the PDB

2WISA	3SS0A	3OGOB	3DQ3A	3NEDA	2AWKA	2B3PA	1YJFA	3SSHA	4L12A
2AWJA	3CIYA	3OGOC	2V4EA	2YE1A	1H6RA	3V3DA	3LVAA	1B9CA	3VHTA
3DQ8A	4KA9A	3OGOD	2V4EB	3LVCA	1HUYA	2WURA	3LVAB	1Z1QA	3DQ6A
3ST3A	2JADA	3GL4A	3A8SA	3ST2A	4AS8A	1OXDA	1QXTA	1RM9A	
1XQMA	4J8AA	3GL4B	3A8SB	3ST2B	3NEZA	3DQKA	4GESB	4BDUA	
3LVDA	3DQUA	1EMKA	1H4UA	3ST2C	2HGDA	4JFGA	2G6YA	4BDUB	
4H47A	3DQCA	1S6ZA	1QYQA	1ZGPA	3SVEA	3DQ5A	3ST4A	4BDUC	
3DQ4A	3CIGA	3AI4A	4AR7A	1JBZA	2FZUA	3DQOA	3ST4B	4BDUD	
3SSLA	3SSKA	3DQNA	1Q4AA	1F0BA	1EMGA	2YE0A	3ST4C	3GB3A	
4LQTA	3DQ1A	1HCJA	2Q57A	2QU1A	2VAEA	4OJ0A	4J88A	3GB3B	
1QYOA	2YFPA	1HCJB	3SV5A	4L1IA	2AWLA	4OJ0B	3GEXA	2QLEA	
2G6EA	4LQUA	1HCJC	3DQHA	1GFLA	1Q4DA	4OJ0C	3SSPA	4OGSA	
1W7UA	1KYSA	1HCJD	1ZGQA	1GFLB	3I19A	4OJ0D	2HCGA	4B30A	
3UG0A	3DPZA	1YFPA	3KCSA	1EMMA	3SSTA	1W7SA	1Q4CA	3SSVA	
2H6VA	1W7TA	1YFPB	1BFPA	2EMDA	1Z1PA	3DQJA	1EMFA	4ACQA	
2H9WA	1GGXA	2DD7A	3DQDA	4ANJA	1GL4A	4H48A	2G6XA	2EMNA	
4L13A	3WCKA	2DD7B	2DUEA	4P7HA	4KW9A	3CGLA	2AWMA	4EULA	
4KW8A	1EMCA	2DUHA	2H8QA	3K1KA	1JC0A	2OKYA	4EN1A	1F09A	
2AHAA	1C4FA	2DUGA	2QRFA	3DQ9A	1JC0B	2YDZA	4EN1B	3SRYA	
1RMOA	3SSYA	1CV7A	2RH7A	2O29A	1JC0C	2WIQA	2OKWA	2Y0GA	
3AI5A	2HJOA	2Q6PA	1QYFA	3OSRA	3UFZA	2WIQB	3DQMA	1YHHA	
1KP5A	3DQAA	1YJ2A	2H5QA	2WSOA	1EMAA	1OXEA	1JC1A	3DQFA	
2H5PA	3SVCA	3DPXA	1QY3A	3LF3A	2G3OA	4KF4A	1JC1B	3DQ2A	
3SVBA	1Q4EA	1KYPA	3DPWA	1Q73A	4JRBA	2HFCA	1JC1C	4B5YA	
2QZ0A	1YHGA	3GJ1A	3GJ2A	1Q4BA	2AH8A	2H5RA	1KYRA	2HRSA	
1MYWA	1YHGB	3GJ1B	3GJ2B	2HGYA	2DUFA	3DQEA	2DD9A	1RMPA	
2O24A	1XMZA	3GJ1C	1YHIA	3SVDA	2EMOA	3LA1A	1EMEA	1YZWA	
4KF5C	1XMZB	3GJ1D	1JBYA	4J89A	1EMLA	3KCTA	2O2BA	1EMBA	
2HQZA	3G9AA	1OXFA	1ZGOA	4J89B	2AWKA	3DQ7A	3ST0A	4KW4A	
3DQLA	3OGOA	2WSNA	2B3QA	3ZTFA	1H6RA	3DQIA	1RMMA	2QT2A	

Table listed the PDB ID of the 273 GFP homologous protein identified by the fold detection using the OCR^S fingerprints against the Protein Data Bank. The PDB ID of the identified nidogen-1 domain protein is highlighted in yellow color.

Table S11: Fold detection results using PSI-BLAST search for GFP-like protein against the PDB

1GFLA	2HGYA	2YDZA	3EKHA	3LS3A	3VK1A	3GL4A	4DKMA
1EMB	1QYQA	4JRBA	3SG2A	3ZULA	1XSSA	3GL4B	2G3OA
1BFPA	2QT2A	4AS8A	3SG4A	3ZUJA	3IP2A	4OHSH	2GW4A
1Q4CA	4KA9A	4AR7A	2HPWA	3ZUFA	2H5OA	2GW4B	2A50A
1HCJA	1YJ2A	3GJ1A	3AKOB	4EMQA	1XAEA	3H1OA	4HVFA
1S6ZA	3V3DA	4LQTA	2G16A	2A48A	3PJBA	4KGEA	1OJ7A
1QYOA	3SSPA	4KW4A	2G2SA	2A46A	4EDOA	4H3MA	
2DUGA	2FWQA	4LQUA	3WCKA	4OJ0A	3BX9A	2QLHA	
2QLEA	3ST0A	2Q6PA	3GB3A	2GX0A	3RWAA	3NF0A	
4OGSA	1H6RA	1YFPA	4DXOA	2Z1OA	3PIBA	2QLIA	
2DUEA	1EMAA	2QRFA	2VVHA	2IE2A	3RWTA	4H3LA	
1Q4EA	2HRSA	1KP5A	4LJDC	4HQ8A	2FL1A	2QLGA	
4GESB	3SRYA	3LVCA	3TMRA	4IZNA	2ICRA	4H3NA	
1EMMA	2WSNA	4ANJA	4LJDA	4HQCD	4OHSA	3LF4B	
1YHIA	3W1CA	2QZ0A	4LJCA	2Z6XA	2H5PA	4DKNA	
1EMFA	1KYPA	4EN1A	1ZUXA	4HQCC	2H5QA	3CFAA	
2WURA	3AI5A	3UFZA	2GW3A	2DDCA	3NEDA	2VAEA	
2OKWA	3W1DA	4LW5A	4B30B	3U0KA	3SVNA	3CFFA	
1YHHA	3VHTA	1RRXA	3A8SA	3M22A	3PJ7A	2A50B	
2HFCA	4JFGA	3U8PA	3P8UA	3IR8A	3SVUA	2A53B	
2HQZA	2B3PA	4KF5A	2OTBA	3LF3A	3SVOA	3NEZA	
2HJOA	4BDUA	3P28A	2HQKA	2OJKA	3PJ5A	2H8QA	
1EMCA	2JADA	3EVRA	2BTJA	1MOUA	3M24A	1GGXA	
2YFPA	3VHTB	3WLCA	3S05A	2WHSA	3U8AA	1ZGPA	
3G9AA	1QY3A	3SG3A	3S05B	3VICA	3NT9A	4I2YA	
1HUYA	3K1KA	3EK8A	4DXIA	3E5VA	3NT3A	4KF4A	
2B3QA	3ST2A	4IK1A	4DXMA	2RH7A	3OGOA	2V4EC	
3OGOA	3LA1A	3SG6A	4Q9WA	3E5WA	3OGOB	2G6XA	
1YJFA	2QU1A	3EVUA	3KCSA	2P4MA	3OGOC	1G7KA	
1B9CA	2WSOA	3EKJA	3U0LA	1MOVA	3OGOD	3CFAL	

Table listed the PDB ID of the 216 GFP homologous protein identified by the fold detection using PSI-BLAST search against the Protein Data Bank.

Table S12: Fold detection results using HMMER search for GFP-like protein against the PDB

1C4FA	2B3QA	3VHTB	1H6RA	2H5OA	3MGFA	3E5TA	1XAEA
1Q4CA	1EMKA	3AI4A	3CB9A	3GB3A	2Z6XA	2RH7A	2ICRA
1W7SA	1B9CA	2HFCA	4AR7A	1ZUXA	2A48A	2ARLA	2GW4B
2Q6PA	3G9AA	1YJ2A	2WSOA	2BTJA	2A46A	2P4MA	4DKNA
2QLEA	2QU1A	3W1DA	1KP5A	4B30A	4OJ0A	3IP2A	4JEOA
1Q4EA	1JC0A	1YHHA	1RMMA	2H5PA	3ADFA	3T6HA	2G6XA
1GFLA	4ANJA	2B3PA	3U8PA	4DXPA	2DDCA	1MOVA	2G3OA
1HCJA	4ANJA	4BDUA	2QZ0A	2OTBA	3ZUFA	4EDSA	3CFAA
1Q73A	2H6VA	2WSNA	2AWLA	4DXOA	3LS3A	4OQWA	3CFFA
2OKWA	3SSPA	2FWQA	1OXDA	3P8UA	2ZO6A	4EDOA	4DKMA
2DUIA	2QT2A	3ST2A	1QYFA	2GW3A	2VAEA	3PIBA	2A52B
2DUGA	4GESB	4JRBA	2AWKA	3S05A	4KF4A	3PJBA	3LF4A
1BFPA	2HGDA	4JRBA	4L12A	2VADA	1XQMA	3RWAA	2DD9A
1EMMA	2H9WA	1QYQA	1QXTA	2H5QA	2H8QA	3BX9A	4HVFA
1JBYA	3DPWA	1KYPA	1OXFA	3KCSA	2V4EC	3RWTF	2DD7A
2DUEA	3SRYA	4J8AA	1CV7A	3NEDA	1GGXA	3VK1A	3CFAL
2HJOA	1RMOA	1RRXA	3OSQA	4JC2A	1XMZA	3PJ5A	2GW4A
2HQZA	1QYOA	4KW4A	4LW5A	3NEZA	3M22A	3SVNA	2A50A
2EMDA	2FZUA	3CD1A	3EKJA	3LF3A	3E5VA	2C9JA	
2WURA	1YJFA	3LA1A	3O77A	4EMQA	4I2YA	3SVOA	
1F09A	2HGYA	2YE0A	3ED8A	3U0LA	1G7KA	3SVUA	
1HUYA	3ST0A	2Q57A	4KF5A	2GX2A	1ZGPA	2OJKA	
1Z1PA	2AH8A	3ZTFA	2G3DB	3NF0A	1ZGQA	4OHSA	
1S6ZA	2HCGA	4GF6B	3AKOA	2A47A	2ZO7A	3NT3A	
2YFPA	2AHAA	2YDZA	4HE4A	4H3MA	1UISA	2PXSA	
1EMCA	3W1CA	1YFPA	4KF5C	2GX0A	2C9IA	3U8AA	
4EULA	2G6EA	4LQUA	4DXNA	3U0KA	2V4EB	3H1OA	
1EMFA	1YHIA	2YE1A	2HQKA	2IOVA	1XSSA	3M24A	
3OGOA	1MYWA	4H47A	3WCKA	4H3NA	3E5WA	4KPIA	
2Y0GA	1YHGA	3GEXA	2VVHA	2ZMUA	1MOUA	4KGEA	

Table listed the PDB ID of the 228 GFP homologous protein identified by the fold detection using HMMER against the Protein Data Bank.

Table S13: Fold detection results using HHpred for GFP-like protein against the PDB

4KW9A	3LS3A	3IR8A	1QY3A	2VADA	3G9AA	1EMMA	4EN1A	2WURA
2H5OA	3ZUFA	2WSOA	2G6EA	3SG6A	2G16B	1YHIA	3UFZA	3VK1A
2A47A	2IB5A	2A46A	3DQ7A	4L13A	2AWLA	1HUYA	2QZ0A	4J89A
2IE2A	4IK8A	2DUEA	1OXDA	3U0LA	2Y0GA	2B3QA	2HCGA	2YE1A
3SVNA	4P7HA	3ST4A	1F0BA	2OTBA	2O24A	3GL4A	1QYFA	3AKOB
3RWAA	3IP2A	3U0KA	1F09A	3U8AA	2QU1A	1UISA	1G7KA	4J8AA
3SVUA	3KCSA	2HQKA	3GJ1A	4LW5A	1YFPA	1HCJB	2RH7A	3I19A
4H48A	1ZGOA	2VVIA	2EMDA	4HE4A	2AWJA	3H1OA	4DKMA	2ZO6A
3CD1A	3ST4C	3NEZA	1QXTA	1XMZB	1JBYA	2OKWA	3A8SB	3LVAA
1RRXA	4J88A	2ZO6A	1YHGA	4HQ8A	1ZUXA	1YHHA	4HQCD	1KYPA
3GEXA	3EVUA	2WSNA	4OQWA	4IZNA	4HQCC	2HFCA	2HPWA	1YJ2A
1QYOA	3SG2A	3W1CA	4EDOA	3VHTB	2DDCA	3LVCA	4JRBA	3U0KA
3ST2B	2DD7A	4I2YA	3PIBA	2JADA	4LJCA	3E5VA	3GB3A	1KP5A
2ZO7A	2G2SA	3OSQA	3PJBA	3AI4A	1ZGQA	3P28A	3E5TA	3WLDA
4OHSA	2G16A	2H8QA	3RWAA	2GW3A	2HPWA	3AKOA	2HRSA	1MOVA
2C9IA	3CGLA	4KGEA	3BX9A	3WCKA	2EMOA	1XMZA	3AI5A	1RM9A
2C9JA	2GX0A	2Z1OA	3RWTF	3NEDA	3ST0A	2ICRA	4B30B	1EMLA
3EVPA	2POXA	3NT3A	3PJ5A	3ST4B	2HGYA	2C9IA	3A8SA	3AI5A
1YZWA	2YFPA	1MOUA	2C9JA	4P7HA	2IOVA	2C9JA	4BDUA	3SVRA
4OHSA	2QLEA	4LJDC	4OHSA	3SRYA	3M24A	1QYQA	4DXOA	4Q9WA
2H5QA	3AKOA	3TMRA	4L12A	2WURA	2OTEA	1CV7A	2HQZA	2GW3A
3GJ1D	3BXCA	2ZMUA	3EKJA	4EN1A	4IK8A	2QRFA	2HJOA	2IB5A
3GL4B	2G6YA	3U8PA	3O77A	2VAEA	4HQ9A	2DUGA	1XQMA	1ZGPA
1Q4EA	3EVPA	4KPIA	4KF5A	2A46A	4AR7A	2PXSA	4KF5C	
4GESB	3E5WA	3M22A	2G3DB	3EVRA	4HQCA	2V4EB	3P8UA	
3OGOA	4OGSA	4DKNA	3SSPA	4DXNA	4OJ0B	1XSSA	3EVPA	
4IK9A	2A50A	4HVFA	3CGLA	3SVOA	4ANJA	4HE4A	3GB3A	
3DQKA	1KYSA	4LJDA	3W1DA	1EMCA	2B3PA	1EMKA	1YJFA	
4BDUA	1OXFA	2A50B	3VHTA	1EMFA	1RMOA	3OSRA	2H5PA	
4L1IA	3ED8A	2Z6XA	2A48A	1B9CA	2OJKA	3SV5A	4JC2A	

Table listed the PDB ID of the 248 GFP homologous protein identified by the fold detection using HHpred against the Protein Data Bank.

Table S14: Fold detection results using FASTA search for GFP-like protein against the PDB

1QYOA	2DUHA	1JBZA	4J89A	2G16B	3GB3A	2H8QA	3VK1A	3SVRA
1GFLA	1Q73A	3K1KA	2OKWA	3AKOA	3GL4A	2Z6XA	1MOVA	3SVUA
2FZUA	4GF6B	2HGDA	1QYFA	4IK9A	2HQKA	2VAEA	1UISA	3BXCA
2HCGA	4EULA	1JC1A	1KYPA	3WLDA	2VVIA	2A46A	2A48A	3NT3A
3MIQA	2Y0GA	1RMMA	1YJ2A	3EVUA	4LJCA	4HQCA	4I2YA	2IB5A
3OGOA	2O24A	2EMNA	1RM9A	3SG2A	3TMRA	3KCSA	2PXSA	3U8AA
1W7SA	2HFCA	1KP5A	4L13A	4IK8A	2GW3A	1ZGOA	2ZO6A	3IP2A
2AH8A	2QU1A	4ANJA	2HRSA	3EVRA	3P8UA	3NEZA	1XSSA	4KGEA
1W7TA	1HUYA	2HJOA	4AS8A	3WLCA	4B30A	2QLHA	1YZWA	
2WURA	1EMCA	3SSYA	4EN1A	3SG7A	2VVJA	4H3LA	1XA9A	
1Q4AA	2DUGA	2QT2A	2YDZA	3EKJA	3A8SA	2QLGA	3E5TA	
2QLEA	1YFPA	1JC0A	2AWMA	3EVVA	4DXIA	3NF0A	2PSLA	
1HCJA	2AWJA	1B9CA	4L1IA	3OSRA	3S05A	2V4EA	1XAEA	
1BFPA	1QY3A	2EMOA	1KYSA	3SG6A	2VADA	3U0KA	3T6HA	
3VHTA	2B3QA	3ST0A	1OXFA	3EVPA	2H5PA	2C9JA	2OGRA	
2JADA	3SRYA	2HGYA	2YE1A	4HE4A	4JC2A	2DDCA	2EJIA	
3AI4A	1YHIA	1EMKA	4J8AA	2HPWA	3U0LA	3VICA	2ZO7A	
4GESB	2G6EA	2B3PA	3I19A	3OSQA	2H5OA	3LF3A	2ICRA	
1EMAA	3DQ7A	1RMOA	2DUEA	3AKOB	2A47A	1MOUA	2FL1A	
1EMGA	1OXDA	2WSNA	2WSOA	2G2SA	2IE2A	3CGLA	4EDSA	
1Q4CA	3G9AA	2QRFA	4AR7A	2G16A	2Z1OA	4H3MA	4OQWA	
1S6ZA	1F0BA	1QYQA	3UFZA	4DXNA	2GX0A	2C9IA	3PJBA	
1Z1QA	1F09A	1CV7A	4H48A	1ZUXA	2POXA	2OJKA	3BXAA	
3V3DA	1YJFA	3LVAA	3CD1A	2OTEA	3LS3A	3IR8A	3M22A	
1C4FA	3GJ1A	1EMLA	1RRXA	2OTBA	3ZUFA	2WHTA	3PIBA	
1Q4DA	2EMDA	2AWLA	3ED8A	4DXOA	4HQ9A	1XQMA	3RWAA	
3ZTFA	1QXTA	2QZ0A	2Q6PA	2BTJA	2H5QA	3E5VA	4OHSA	
4JRBA	1YHGA	1MYWA	4LW5A	2VZXA	4HQ8A	2RH7A	3PJ5A	
4BDUA	3SVCA	2AHAA	3U8PA	2WIQA	2EJOA	2P4MA	3M24A	
3VHTB	1H6RA	2Q57A	2G2SB	3NEDA	3ADFA	3E5WA	4OQWE	

Table listed the PDB ID of the 248 GFP homologous protein identified by the fold detection using FASTA search against the Protein Data Bank.

Table S15. Sequence similarity among the identified “*Potential distant Cupredoxin Homologs*” sequences.

	1AAJ	WP_010687666	WP_019903657	WP_019121393	WP_019155648	WP_015594375	WP_008636814	WP_019379850	WP_021320206	WP_021239068	WP_004263537	WP_019883210	WP_008217106	WP_020895055
1AAJ	100	17.71	23.96	16.85	22.92	17.53	19.15	20.62	23.33	23.33	15.22	17.48	16.5	15.79
WP_010687666		100	49.09	19.1	19.23	18.1	16.35	16.19	18.75	18.18	27.36	23.28	16.5	13.68
WP_019903657			100	16.28	20.62	18.37	21.65	16.33	22.68	19.1	26.73	23.68	21.9	17.02
WP_019121393				100	55.12	57.81	58.91	58.14	14.77	12.5	21.18	15.22	16.81	13.04
WP_019155648					100	55.41	53.79	62.16	15.15	16.13	14.14	12.61	21.09	15.87
WP_015594375						100	50	60.4	15	14.89	17	13.39	18.6	14.96
WP_008636814							100	55.78	11.88	11.83	22	14.78	18.6	16.94
WP_019379850								100	14.85	11.58	12	12.5	16.15	13.39
WP_021320206									100	65.67	17.21	16.91	19.47	12.77
WP_021239068										100	18.35	17.07	19.42	14.44
WP_004263537											100	44.85	21.82	13.19
WP_019883210												100	21.67	12.12
WP_008217106													100	11.38
WP_020895055														100

Table listed the amino acid sequence similarities (%) among 13 identified ***Potential distant Cupredoxin Homologs*** and the representative Cupredoxin protein (1AAJ) identified using clustalW2. No significant sequence similarity is found between the identified protein sequences and the representative Cupredoxin-like protein, with the maximum sequence identity of 23.96%.