UC Riverside

UC Riverside Previously Published Works

Title

The crystal structure of a bacterial Sufu-like protein defines a novel group of bacterial proteins that are similar to the N-terminal domain of human Sufu

Permalink

https://escholarship.org/uc/item/9515f0hz

Journal Protein Science, 19(11)

ISSN 0961-8368

Authors

Das, Debanu Finn, Robert D Abdubek, Polat <u>et al.</u>

Publication Date

2010-11-01

DOI

10.1002/pro.497

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <u>https://creativecommons.org/licenses/by/4.0/</u>

Peer reviewed

The crystal structure of a bacterial Sufu-like protein defines a novel group of bacterial proteins that are similar to the N-terminal domain of human Sufu

Debanu Das,^{1,2} Robert D. Finn,³ Polat Abdubek,^{1,4} Tamara Astakhova,^{1,5} Herbert L. Axelrod,^{1,2} Constantina Bakolitsa,^{1,6} Xiaohui Cai,^{1,5} Dennis Carlton,^{1,7} Connie Chen,^{1,4} Hsiu-Ju Chiu,^{1,2} Michelle Chiu,^{1,4} Thomas Clayton,^{1,7} Marc C. Deller,^{1,7} Lian Duan,^{1,5} Kyle Ellrott,^{1,6} Carol L. Farr,^{1,7} Julie Feuerhelm,^{1,4} Joanna C. Grant,^{1,4} Anna Grzechnik,^{1,7} Gye Won Han,^{1,7} Lukasz Jaroszewski,^{1,5,6} Kevin K. Jin,^{1,2} Heath E. Klock,^{1,4} Mark W. Knuth,^{1,4} Piotr Kozbial,^{1,6} S. Sri Krishna,^{1,5,6} Abhinav Kumar,^{1,2} Winnie W. Lam,^{1,2} David Marciano,^{1,7} Mitchell D. Miller,^{1,2} Andrew T. Morse,^{1,5} Edward Nigoghossian,^{1,4} Amanda Nopakun,^{1,7} Linda Okach,^{1,4} Christina Puckett,^{1,4} Ron Reyes,^{1,2} Henry J. Tien,^{1,7} Christine B. Trame,^{1,2} Henry van den Bedem,^{1,2} Dana Weekes,^{1,6} Tiffany Wooten,^{1,4} Qingping Xu,^{1,2} Andrew Yeh,^{1,2} Jiadong Zhou,^{1,4} Keith O. Hodgson,^{1,8} John Wooley,^{1,5} Marc-André Elsliger,^{1,7} Ashley M. Deacon,^{1,2}

¹Joint Center for Structural Genomics, http://www.jcsg.org

²Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025
 ³Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, United Kingdom
 ⁴Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, California 92121
 ⁵Center for Research in Biological Systems, University of California, San Diego, La Jolla, California 92093
 ⁶Program on Bioinformatics and Systems Biology, Sanford-Burnham Medical Research Institute, La Jolla, California 92037
 ⁷Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037
 ⁸Photon Science, SLAC National Accelerator Laboratory, Menlo Park, California 94025

Received 19 July 2010; Accepted 26 August 2010 DOI: 10.1002/pro.497 Published online 10 September 2010 proteinscience.org

Abstract: Sufu (Suppressor of Fused), a two-domain protein, plays a critical role in regulating Hedgehog signaling and is conserved from flies to humans. A few bacterial Sufu-like proteins have previously been identified based on sequence similarity to the N-terminal domain of eukaryotic Sufu proteins, but none have been structurally or biochemically characterized and their function in bacteria is unknown. We have determined the crystal structure of a more distantly related Sufu-like homolog, NGO1391 from *Neisseria gonorrhoeae*, at 1.4 Å resolution, which provides the first biophysical characterization of a bacterial Sufu-like protein. The structure revealed a striking similarity to the N-terminal domain of human Sufu (r.m.s.d. of 2.6 Å over 93% of the NGO1391 protein), despite an extremely low sequence identity of ~15%. Subsequent sequence analysis

Grant sponsors: NIH, National Institute of General Medical Sciences, and Protein Structure Initiative; Grant number: U54 GM074898; Grant sponsor: Wellcome Trust; Grant number: WT077044/Z/05/Z.

^{*}Correspondence to: Dr. Ian A. Wilson, JCSG, The Scripps Research Institute, BCC206, 10550 North Torrey Pines Road, La Jolla, CA 92037. E-mail: wilson@scripps.edu

revealed that NGO1391 defines a new subset of smaller, Sufu-like proteins that are present in ~200 bacterial species and has resulted in expansion of the SUFU (PF05076) family in Pfam.

Keywords: *Neisseria* gonorrhoeae; NGO1391; UniProt Q5F6Z8; Pfam PF05076; suppressor of fused; sufu-like; structural genomics

Introduction

The Hedgehog (Hh) and Wingless (Wnt) signaling pathways are critical for animal development.¹⁻⁴ They are important in embryonic growth in vertebrates and stem cell differentiation. Disruption of these pathways can lead to disruption of stem cell function and cancer, for example.^{5,6} Recent studies have addressed how these pathways may have evolved.⁷⁻⁹ The human Sufu (Suppressor of Fused) protein is a key transcriptional regulator in the Hh pathway. Until recently, 20 eukaryotic, two-domain proteins of ~500 residues [N-terminal domain: Pfam SUFU family (PF05076) and C-terminal domain: Pfam SUFU_C family (PF12470)] and 13 bacterial Sufu-like single-domain proteins of unknown function (~ 250 residues, with $\sim 25-40\%$ sequence identity to the N-terminal Sufu domain of eukaryotic proteins) were classified in Pfam (version 23.0).¹⁰ The Sufu_C domain interacts with the N-terminal domain of Gli transcription factors, whereas the Sufu domain is believed to interact with the C-terminal tail of Gli.^{11,12} Apart from general classification of these bacterial proteins as Sufu-like, none had been investigated experimentally and their roles were consequently unclear, since bacteria do not use Hedgehog signaling.

As part of our efforts to characterize novel protein sequence space, we identified a set of proteins in Neisseria species that are distantly related to members of PF05076. This similarity was only detectable using profile–profile sequence comparison methods.¹³ NGO1391 (UniProt id Q5F6Z8) from *Neisseria gonorrhea* FA1090 was selected as a representative of this set for structure determination, using the semiautomated, high-throughput pipeline of the Joint Center for Structural Genomics (JCSG; http://www.jcsg.org) as part of the National Institute of General Medical Sciences' Protein Structure Initiative.

Results and Discussion

Overall structure

NGO1391 from *Neisseria gonorrhoeae* FA 1090 was cloned, expressed, purified, and crystallized according to JCSG protocols as described in Materials and Methods. The crystal structure (Fig. 1) was determined by single-wavelength anomalous diffraction (SAD) phasing to a resolution of 1.40 Å. Data collection, model, and refinement statistics are summarized in Table I.¹⁴ The final model contains one monomer consisting of Gly0 (from the purification tag) and residues 1–181 of NGO1391 (the full-length protein is 182 residues), 265 waters, 4 sulfates, and 1 glycerol molecule in the asymmetric unit (ASU). Residue 182 was disordered and was not modeled. A monomer is the likely oligomeric form in solution as judged from crystal lattice packing and assembly analysis using PISA, and the monomeric form is supported by analytical size-exclusion chromatography. The Matthews' coefficient $(V_M)^{15}$ is 2.5 Å³/Da, with an estimated solvent content of 50%. The Ramachandran plot produced by Molprobity¹⁶ shows that 97.8% of the amino acids are in the favored regions with one outlier, Asp132, which is in a region of good density, but with alternate main-chain and side-chain conformations.

NGO1391 is a compact molecule with approximate dimensions of 50 Å \times 40 Å \times 30 Å and has a tapered appearance due to a bulkier N-terminal region. It consists of a central, slightly S-shaped, antiparallel β -sheet (β 1- β 7) flanked by α - and 3_{10} -helices (H1-H8) (Fig. 1). Some of the loops are long and extend out from the protein core. The C-terminus consists of a long tail with no secondary structure (residues 165-181). A glycerol molecule from the cryoprotectant binds in a relatively hydrophobic cleft formed by Ser119, Tyr122, Trp138, Leu140, Glu159, Phe162, Asp163, Ile167, and Tyr169 near the C-terminus. Most of these residues are conserved in the newly identified Sufu-like proteins (see Sequence Analysis) and three are also conserved in human Sufu (Ser119, Leu140 and Ile167 correspond to human Sufu Thr180, Val201, and Ile228, respectively).

Comparison to human Sufu

A search for other proteins of similar structure using sequence and structure-based methods was carried out using FFAS,13 DALI,17 FATCAT,18 and SSM.19 In all cases, the only significant hit (for example, the e-value of the top FATCAT hit is 3.95e-10 and the next hit is 4.35e-03; the top DALI hit has a Z-score of 16.7 and the next hit has 5.2, the top FFAS hit has a score of -49.4 and the next hit is -5.7) was to the N-terminal domain (NTD, 236 out of 484 residues) of the human Sufu protein (PDB accession code 1m1l,¹² UniProt Q9UMX1). NGO1391 can be superimposed onto the NTD of Sufu with an r.m.s.d. of 2.6 Å over 170 Ca atoms (93% of the NGO1391 protein) with structure-based (DALI) and sequencebased (FFAS) sequence identities of 15% (Fig. 2). The notable structural differences between NGO1391 and the NTD of Sufu include: short helices H2 and H3 in NGO1391 correspond to long loops in Sufu; longer



Figure 1. Crystal structure of NGO1391 from *Neisseria gonorrhoea* FA1090. (a) Stereo ribbon diagram of the NGO1391 monomer color-coded from N-terminus (yellow) to C-terminus (magenta). Helices H1–H8 (helix H5 is a 3_{10} helix) and β -strands β 1- β 7 are indicated. (b) Diagram showing the secondary structural elements of NGO1391 superimposed on its primary sequence. The α -helices, 3_{10} helix, and β -strands are indicated.

loops in Sufu occur between strands $\beta 2$ and $\beta 3$, $\beta 3$ and H3, and $\beta 4$ and H4; and additional helices are present near the C-terminus.

Human full-length Sufu was found to be a monomer in solution (n.b. the NTD of Sufu forms a tetramer in the crystal structure and only the NTD 3D structure of human Sufu is available). In its inactive signaling state, the NTD of Sufu interacts with the C-terminal region of Gli transcription factors and prevents their transport into the nucleus. In the active form, which is phosphorylation dependent, Sufu forms a complex with Fused, thereby releasing Gli. Asp159 of Sufu was found to be important in the interaction with Gli.¹² This residue corresponds to Lys99 in NGO1391 (Fig. 3), which in itself is not highly conserved across the bacterial Sufu-like

Table I. Summary of Crystal Parameters, Data Collection, and Refinement Statistics for PDB 3k5j

Space group	$P3_{2}21$					
Unit cell parameters	$a = 50.72 ext{ \AA}, b = 50.72 ext{ \AA}, c = 143.60 ext{ \AA}$					
Data collection	λ_1 SAD-Se					
Wavelength (Å)	0.9789					
Resolution range (Å)	27.7 - 1.40					
Number of observations	365,215					
Number of unique reflections	43,173					
Completeness (%)	$99.7 (98.6)^{\rm a}$					
Mean I/σ (I)	$17.2 (1.8)^{\rm a}$					
$R_{ m sym}$ on I (%)	$6.0 (76.1)^{a}$					
Highest resolution shell (Å)	1.45 - 1.40					
Model and refinement statistics						
Resolution range (Å)	27.7 - 1.40	Data set used in refinement	λ_1 SAD-Se			
Number of reflections (total)	$43,107^{\rm b}$	Cutoff criteria	$ \mathbf{F} > 0$			
Number of reflections (test)	2169	$R_{ m cryst}$	0.134			
Completeness (% total)	99.8	$R_{ m free}$	0.163			
Stereochemical parameters						
Restraints (RMSD observed)						
Bond angle (°)	1.60					
Bond length (Å)	0.015					
Average protein isotropic B-value (Å ²)	20.3°					
ESU based on R_{free} (Å)	0.05					
No. of protein residues/atoms	182/1593					
No. of water/sulfate/glycerol molecules	265/4/1					

ESU = Estimated overall coordinate error.¹⁴

 $R_{sym} = \Sigma |I_i - \langle I_i \rangle | / \Sigma |I_i|$, where I_i is the scaled intensity of the ith measurement and $\langle I_i \rangle$ is the mean intensity for that reflection.

 $R_{\text{cryst}} = \Sigma ||F_{\text{obs}}| - |F_{\text{calc}}||/\Sigma |F_{\text{obs}}|$, where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.

 $R_{\rm free} =$ as for $R_{\rm cryst}$, but for 5.0% of the total reflections chosen at random and omitted from refinement.

^a Highest resolution shell.

^b Typically, the number of unique reflections used in refinement is slightly less that the total number that were integrated and scaled. Reflections are excluded due to negative intensities and rounding errors in the resolution limits and cell parameters. ^c This value represents the total B.

proteins. Seven residues contribute to a more negatively charged region on one side of Sufu: Glu106, Asp111, Glu152, Asp159, Glu181, Glu221, and Asp262 (Fig. 4). These residues are not conserved across bacterial Sufulike proteins. Both sides of NGO1391 appear to have a similar number of negatively charged residues (Fig. 4). Some other residues are more strongly conserved between the eukaryotic and bacterial forms of the protein with several on the protein surface (NGO1391 Thr24, Glu49, Trp51, Tyr53, Pro106, Ser111, and His115) that could be functionally important.

Sequence analysis and genomic context

At the time of the NGO1391 structure determination, Pfam (version 23.0) SUFU family (PF05076) contained 33 proteins from 27 species, 20 of which were eukaryotic including the N-terminal domain of human Sufu and 13 were bacterial. The eukaryotic Sufu proteins (primarily from chordates and insects) all contained ~500 residues, and consisted of two domains of approximately equal size (Sufu and Sufu_C). The bacterial proteins, on the other hand, were uncharacterized proteins of only around 250 residues.

A search against the UniProt database using PSI-BLAST revealed that NGO1391 (Q5F6Z8) had

several sequence homologs of ~92% sequence identity in other Neisseria species: N. gonorrhoeae 1291 (C1HYY6), N. gonorrhoeae NCCP11945 (B4RNC6), N. flavescens NRL30031/H210 (C0EME3), N. meningitidis serogroup C/serotype 2a (A1KSR8), N. cinerea ATCC14685 (C0DN13), and N. lactamica ATCC23970 (C0F5L2). The next nearest hits were to proteins with ~33–39% sequence identity (B5HWZ3 from Streptomyces sviceus ATCC 29083, C2A7G8 from Thermomonospora curvata DSM 43183, C1YUI8 from Nocardiopsis dassonvillei subsp. dassonvillei DSM 43111 and A6A6Q5 from Vibrio cholerae MZO-2). However, none of these proteins (Fig. 3) were included in PF05076.

Three rounds of iterative PSI-BLAST, including sequence similarities down to a threshold of 0.1, identified \sim 300 additional bacterial proteins, in addition to the 33 vertebrate and bacterial Sufu proteins already present in PF05076. Pairwise comparisons revealed that the bacterial proteins already included in PF05076 had sequence identities of \sim 25– 40% to the eukaryotic Sufu proteins, but less than 16% identity to the additional Sufu-like proteins identified here. A multiple sequence alignment of a representative set of these new Sufu-like proteins identified residues that are invariant or similar in these proteins (Fig. 3) and their distribution in the structure (Fig. 5).



Figure 2. Structural comparison of NGO1391 and human Sufu protein. The structures of NGO1391 and the N-terminal domain (NTD) of human Sufu protein (gray) are highly similar and superimpose with an r.m.s.d. of 2.6 Å over 170 C α atoms.

Based on this analysis, Pfam PF05076 was independently revised to include these proteins using the HMMER3 software (http://hmmer.janelia.org). Thus, PF05076 in the current release of Pfam (v. 24.0, October 2009) now contains 341 proteins from 31 eukaryotic and 220 bacterial species. Inspection of the residue conservation scores in the multiple sequence alignment for this revised protein family on the Pfam website (http://pfam.sanger.ac.uk/family/PF05076), reveals that the only residues that that are conserved across the entire protein family (341 proteins) are Val134, Phe136, Leu139, Ile142, Glu146, and Leu158 in NGO1391. Of these residues, only Val134 is surface exposed whereas the rest are in the protein core.

The genomic neighbors of NGO1391 are putative MafB-like (NGO1392) and MafA2 (NGO1393) proteins that are predicted to be functional partners of NGO1391 with high scores of 0.87 and 0.86, respectively.²² The Maf proteins are believed to have adhesion capabilities.^{23,24} Interestingly, this predicted association is similar to that of the β -catenin in the Wnt signaling pathway, where β -catenin can form a complex with the cadherin cell-adhesion molecules. β -catenin also functions as a transcriptional regulator. However, no sequence or structural similarity is

found between NGO1391 and β -catenin and this genomic context does not appear to be conserved across other members of this Pfam.

The crystal structure and sequence analysis presented here provides the first molecular characterization of a bacterial Sufu-like protein that may be a prokaryotic homolog of modern Sufu proteins involved in embryonic development in vertebrates. This study has also led to identification of a large number of other bacterial Sufu-like proteins. Mechanistic and mutagenesis experiments can help to elucidate which residues are important for molecular function and substrate specificity and aid in understanding the function of NGO1391 and related proteins. Additional information about the proteins described in this study is available $TOPSAN^{25,26}$ at http://www.topsan.org/ from explore?PDBid=3k5j.

Materials and Methods

Protein expression, purification, and crystallization

Clones were generated using the Polymerase Incomplete Primer Extension (PIPE) cloning method.²⁷ The gene encoding NGO1391 (GenBank: YP 208451, gi | 59801739, UniProt: Q5F6Z8) was amplified by polymerase chain reaction (PCR) from Neisseria gonorrhoeae genomic DNA using PfuTurbo DNA polymerase (Stratagene) and I-PIPE (Insert) primers (forward primer: 5'-ctgtacttccagggc ATGGACTATAA CCAAACTGTTTTATCTC3', reverse primer: 5'-aat taagtcgcgtta TCCTGCCTGCCAGACAGTACTCGC ACG-3', target sequence in upper case) that included sequences for the predicted 5' and 3' ends. The expression vector, pSpeedET, which encodes an amino-terminal tobacco etch virus (TEV) proteasecleavable expression and purification tag (MGSDK IHHHHHHENLYFQ/G), was PCR amplified with V-PIPE (Vector) primers (forward primer: 5'-taacgc gacttaattaactcgtttaaacggtctccagc-3', reverse primer: 5'-gccctggaagtacaggttttcgtgatgatgatgatgatgatg-3'). V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments together. Escherichia coli GeneHogs (Invitrogen) competent cells were transformed with the I-PIPE/V-PIPE mixture and dispensed on selective LB-agar plates. The cloning junctions were confirmed by DNA sequencing. Expression was performed in a selenomethionine-containing medium at 37°C. Selenomethionine was incorporated via inhibition of methionine biosynthesis, which does not require a methionine-auxotrophic strain.²⁸ At the end of fermentation, lysozyme was added to the culture to a final concentration of 250 µg/mL, and the cells were harvested and frozen. After one freeze/thaw cycle the cells were sonicated in lysis buffer [50 mM HEPES pH 8.0, 50 mM NaCl, 10 mM

CSF628_NEIGI DUPOTOUCUOU DUPOTOUCUOU DUPOTOUCUOU CSF628_NEIGI	Q5F6Z8_NEIGI Q5F6Z8_NEIGI					000000		0.0000	
QSF628_NEIG1 1	Q5F6Z8_NEIG1				•		•		•
Q5F628_NEIGI → → → ↓ <	C1HYY6_NEIGO B4RNC6_NEIG2 C0EME3_NEIFL AlKSR8_NEIMF C0DN13_NEICI C0F5L2_NEILA A6A6Q5_VIBCH C1YU18_NOCDA C2A7G8_THECU B5HWZ3_9ACTO	111111111111111111111111111111111111111		LQATFENRLKS FQATFENRLKS FQATFENRLKS FQATFENRLKS LRRALCIKGIM MATGE L	MDYNQTVL MDYNQTVL MDYNQTVL MDYNQTVL MDYNQTVL MDYNQTVL KSTKEKII DARIAALE	SHLQKFWK SHLQKFWK SHLQKFWK SHLQKFWK SHLQNFWK SHLQNFWK SHLQNFWK KHYAEFFE KHYAEFFE SHVRAFFM	HHDIKGFTW HHDIKGFTW HHDIKGFTW HHDIKGFTW HHDIKGFTW SHDFEILKGFTW GHDVEVVVV GHDVEVVVV	TLGR.IVEE TLGR.IVEE TLGR.IVEE ALGR.IVEE ALGR.IVEE ELGP.IKKI EDNP.VLDL 	LPDFQVF LPDFQVF LPDFQVF LPDFQVF LPDFQVF LPDFQVF VPDFQVV APGFATY MR VPDLRVL
Q5F628_NEIGI COO EXAMPLE COO			83	H3	84		НИ	H5	85
OSF628_NEIG1 41 QVIPNHEDEPWVVVSSGIGQFLGQEFFIISPFETPEHIETLAMLASASMHYPDOFOLGKTVNIGR C1HYY6_NEIG0 41 QVIPNHEDEPWVVVSSGIGQFLGQEFFIISPFETPEHIETLAMLASASMHYPDOFOLGKTVNIGR B4RNC6_NEIG2 41 QVIPNHEDEPWVVVSSGIGQFLGQEFFIISPFETPEHIETLAMLASASMHYPDOFOLGKTVNIGR COEME3_NEIFL 58 QVIPNHEDEPWVVVSSGIGQFLGQEFFIISPFETPEHIETLAMLASASMHYPDOFOLGKTVNIGR AIKSR8_NEIMF 70 QVIPNHEDEPWVVVSSGIGQFLEQEFFIVSPFETPEHIETLAMLASASMHYPDOFOLGKTVNIGR CODN13_NEICI 61 QVIPNHEDEPWVVVSSGIGPFLGQEFFIVSPFETPEHIETLAMLASASMHYPDOFOLGKTVNIGR CODS13_NEICI 61 QVIPNHEDEPWVVVSGIGSFLGQEFFIVSPFETPEHIETLAMLASASMHYPDOFOLGKTVNIGR A6AQ5_VIBCH 59 RFAPGPQINMWVCSVGASEITHSDSVLHEVVVSPFETPEHIETLAMLASASMHYPDOFOLGKTVNIGR C2A768_THECU 3 RITPASAQDAWVYVKSGIGSFLGQEFFIVSPFETPEHIETLAMLASASMHYPDOFOLGKTVNIGR C2A768_THECU 3 RITPASAQDAWVYVKSGIGSFLGQQRLEFIVAPTAERAHAETLGMVSHLHSFELVLDVGSVIDLGG C2HU18_NOCDA 46 RVAPRNEGEGWTVVKSGIGQQR.LEFIVTAPVRDQRFIDLMAMIAHYHCGG_HQLALEHSMPIGE Q5F628_NEIG1 106 PWVEQSSFRHFIISIPYYGQELEYMDNVRFFWLDPITVESLAMMIAHYHCGG_HQLALEHSMPIGE Q5F628_NEIG1 106 PWVEQSSFRHFISIPYYGQELEYMDNVRFFWLDPITVESLETKFDEAGIDYL C0EME3_NEIFL 123 PRVGQSSFWHFUSIPYYGQELEYMDNVRFFWLDP	O5F6Z8 NEIG1		→ <u>→</u>	0000 -	>	00000	000000000	0 0 0 0	→ →
Yesse_Neiging 41 ViewheberwyvvssGigOFLGOEFFIISPFETPEHIETLAMLASASMHYDOFOLGKTVNIGR B4RNC6_NEIG2 41 QVIEWhEDEPWYVVSSGIGOFLGOEFFIISPFETPEHIETLAMLASASMHYDDOFOLGKTVNIGR C0EME3_NEIFL 58 QVIEWHEDEPWYVVSSGIGOFLGOEFFIISPFETPEHIETLAMLASASMHYDDFOLGKTVNIGR A1KSR8_NEIMF 70 QVIEWHEDEPWYVVSSGIGOFLEOEFFIISPFETPEHIETLAMLASASMHYDDFOLGKTVNIGR C0DN13_NEICI 61 QVIEWHADEPWYVSSGIGOFLGOEFFIVSPFETPEHIETLAMLASASMHYDDFOLGKTVNIGR C0F5L2_NEILA 61 QVIEWHADEPWYVSSGIGOFLGOEFFIVSPFETPEHIETLAMLASASMHYDDFOLGKTVNIGR C0A6605_VIBCH 59 RFAPGPQINMWVYCSVGASEITHSDSVLHEFVVVSPIESNRLVEMLAMVTYYHSNHNLGFGHTLPIGE C1YUI8_NOCDA 46 RVAPRNEGEGWTYVTVGSSVHEAGGGTEFFLVAPTAERAHAETLGMVSHLHSFEEYRLDVGSVIDLGG C2A768_THECU 3 RITPASAQDAWVVTKGAFEADGGQRLEFIIEAPAEDPIHVESLAMVTYRH.LTGPPLDLQQTVAIGR B5HWZ3_9ACTO 47 VVEPGPRSDSWAFVTAGCWASMEGNGHGLEFVMTAPVRDQRFIDLMAMIAHYHCGG.HQLALEHSMPIGE Q5F628_NEIG1 06 H6 B7 H7 H8 Q5F628_NEIG1 06 PWVEQSSFRHFLISLPYPYGOELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL Q5F628_NEIG1 06 PWVEQSSFRHFLISLPYPYGOELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL Q5F628_NEIG1 06 PWVEQSSFRHFLISLPYPYGOELEYMDN	05E679 NETC1	41	ONTRALEDEDRUVVCCT		PPPTTOPP	ETDEUTET.	TAMTACACM	· VEDOFOT C	K T VINT CD
B4RNC6_NEIG2 41 QVIPNHEDEPWVYVSSGIGQFLGQEFFISPFETPEHIETLAMLASASMHYPDQFQLGKTVNIGR COEME3_NEIFL 58 QVIPNHEDEPWVYVSSGIGPFLGQEFFISPFETPEHIETLAMLASASMHYPDRFQLGKTVNIGR A1KSR8_NEIMF 70 QVIPNHEDEPWVYVSSGIGOFLGQEFFISPFETPEHIETLAMLASASMHYPDRFQLGKTVNIGR CODNI3_NEICI 61 QVIPNHEDEPWVYVSSGIGOFLGQEFFIVSPFETPEHIETLAMLASASMHYPDQFQLGKTVNIGR COF5L2_NEILA 61 QVIPNHEDEPWVYVSSGIGSFLGQEFFIVSPFETPEHIETLAMLASASMHYPDQFQLGKTVNIGR A6A6Q5_VIBCH 59 RFAPGPQINMWVYCSVGASEITHSDSVLHEFVVSPFETPEHIETLAMLASASMHYPDQFQLGKTVNIGR C1YUI8_NOCDA 46 RVAPRNEGGWTVVVGSSVHEAGGQEFFIVSPFETPEHIETLAMLASASMHYPDQFQLGKTVNIGR C2A768_THECU 3 RITPASAQDAWVYVKGAFEADGGQREFFIVSPFETPEHIETLAMLASASMHYPDQFQLGVVAIGR B5HWZ3_9ACTO 47 VVEPGPRSDSWAFVTAGCWASMEGNGHGLEFVMTAPVRDQRFIDLMAMIAHYPCGG_HQLALEHSMPIGE Q5F628_NEIGI 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL A1KSR8_NEIGI 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL C2HYG_NSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL 00000 C1HYG_NCA 123 PRVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFIDTHSVEELETKFDEAGIDYL C0EM3_NEIFL 123 PRVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLF	C1HYY6 NEIGO	41	OVIENHEDEEWVYVSSGI	GOFLGO	EFFIISPF	ETPEHIET	LAMLASASM	HYPDOFOLG	KTVNIGR
COEME3_NEIFL 58 QVIENHEDEPWVVVSSGIGPFLGQEFFIVSEFETPEHIETLAMLASASMHYPDRFQLGKIVNIGR AIKS8_NEIMF 70 QVIENHEDEPWVVVSSGIGOFLGQEFFIVSEFETPEHIETLAMLASASMHYPDRFQLGKIVNIGR CODNI3_NEICI 61 QVIENHEDEPWVVVSSGIGOFLGQEFFIVSEFETPEHIETLAMLASASMHYPDRFQLGKIVNIGR A6A6Q5_VIBCH 59 RFAPGPQINMWVCSVGASEITHSDVLHEFVVSFFETPEHIETLAMLASASMHYPDRFQLGKIVNIGR C1YU18_NOCDA 46 RVAPRNEGGWTVVVGSVGASEITHSDVLHEFVVSPFETPEHIETLAMLASASMHYPDRFQLGKIVNIGR C2A768_THECU 3 RITPASAQDAWVVVKGASEITHSDVLHEFVVSPFETPEHIETLAMLASASMHYPDRFQLGKIVNIGR C2A768_THECU 59 RFAPGPQINMWVCSVGASEITHSDVLHEFVVSPFETPEHIETLAMLASASMHYPDRFQLGKIVNIGR C2A768_THECU 3 RITPASAQDAWVVVKGAFEADGGQRLEFIVAPTAERAHAETIGWVSHLHSFEEYRLDVGSVIDLG C2A768_THECU 55 HWZ3_9ACTO 47 VVEPGPRSDSWAFVTACCWASMEGNGHGLEFVMTAPVRDQRFIDLMAMIAHYHCGG.HQLALEHSMPIGE 0 000000000 0 000000000 0 00000000 0 00000000	B4RNC6_NEIG2	41	QVIPNHEDEPWVYVSSGI	GQFLGQ	EFFIISPF	ETPEHIET	LAMLASASM	HYPDQFQLG	KTVNIGR
AIKSRG_NEIMF 70 QV IPNHEDEPWVYVSSGIGGPLG QEFFISEFETPEHIETLAMLASASMHYPDRFQLGKTVNIGR CODNI3.NEICI 61 QVIPNHEDEPWVYVSSGIGSFLGQEFFISEFETPEHIETLAMLASASMHYPDQFQLGKTVNIGR A6A6Q5_VIBCH 61 QVIPNHEDEPWVYVSSGIGSFLGQEFFISEFETPEHIETLAMLASASMHYPDQFQLGKTVNIGR A6A6Q5_VIBCH 59 RFAPGPQINMWVCSVGASEITHSDSVLHEFVSPFETPEHIETLAMLASASMHYPDQFQLGKTVNIGR C1YUI8_NOCDA 46 RVAPRNEGEGWTYVTVGSSVHEAGGQ.IEFFISAPTENLEMLAMVTYYHSNHNLGFGHTLPIGE C2A7G5_THECU 3 RITPASAQDAWVVVKGAFEADGGQR.IEFFISAPTAERAHAETLGMVSHLHSFEEYRLDUGSVIDLGG C2A7G5_THECU 3 RITPASAQDAWVVVKGAFEADGGQR.IEFFISAPTAERAHAETLGMVSHLHSFEEYRLDUGQTVAIGR B5HWZ3_9ACTO 47 VVEPGPRSDSWAFVTAGCWASMEGNGHGLEFVMTAPVRDQRFIDLMAMIAHYHCGG.HQLALEHSMPIGE 0000 05F628_NEIGI 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL C1HYY6_NEIG2 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL C2HY6QSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL C0EME3_NEIFL 123 PRVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL C0DN13_NEICI 26 PWVQSSFQHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL C0DSL2_NEIFL 126 PWVQSSFGHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL C0FSL2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL C0FSL2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL C0FSL2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL C0FSL2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL C0FSL2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL C0FSL2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL C0FSL2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL C0FSL2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL C0FSL2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL C0FSL2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELE	COEME3_NEIFL	58	QVIPNHEDEPWVYVSSGI	GPFLGQ	EFFIVSPF	ETPEHIET	LAMLASASM	HYPDRFQLG	KIVNIGR
COF5L2_NEILA 61 QVIENHEDEPWVYVSSGIGSFLGQEFFIVSEFETPEHIETLAMLASASMHYPDRFQLGKTVNIGR A6A6Q5_VIBCH A6A6Q5_VIBCH 59 RFAPGPQINMWVYCSVGASEITHSDSVLHEFVVVSPIESNRLVEMLAMVTYYHSNHNLGFGHTLPIGE C1YUI8_NOCDA 46 RVAPRNEGEGWTYVTVGSVIGASEITHSDSVLHEFVVVSPIESNRLVEMLAMVTYYHSNHNLGFGHTLPIGE C2A7G8_THECU 3 RITPASAQDAWVYVTKGAFEADGGQRLEFIVAPTAERAHAETLGMVSHLHSFEEYRLDVGSVIDLGG B5HWZ3_9ACTO 47 VVEPGPRSDSWAFVTAGCWASMEGNGHGLEFVMTAPVRDQRFIDLMAMIAHYHCGG.HQLALEHSMPIGE Q5F6Z8_NEIGI 06 H6 B7 H7 H8 Q5F6Z8_NEIGI 06 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL 000000000000000000000000000000000000	CODN13 NETCT	61	OVIENHADEEWVYVSSGI	GPFLG	EFFILSPE	EIPERIEI. ETPERIET	LAMLASASM	HYPDOFOLG	KTVNIGR
A6A6Q5_VIBCH 59 RFAP GPQINMWVYCSVGASEITHSDSVLHEFVVVSPIESNRLVEMLAMVTYYHSN. HNLGFGHTLPIGE C1YUI8_NOCDA 46 RVAPRNEGEGWTYVTVGSSVHEAGGGTEFFLVAPTAERAHAETLGMVSHLHSFEEYRLDVGSVIDLGG C2A768_THECU 3 RITPASAQDAWVYVKGAFEADGGQRLEFILAPAEDPIHVESLAMVTYRH. LTGPPLDLGQTVAIGR B5HWZ3_9ACTO 47 VVEPGPRSDSWAFVTAGCWASMEGNGHGLEFVMTAPVRDQRFIDLMAMIAHYHCGG.HQLALEHSMPIGE Q5F628_NEIGI 106 PWVEQSSFRHFLISIPYPYQQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL C1HYY6_NEIGO 106 PWVEQSSFRHFLISIPYPYQQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL B4RNC6_NEIG2 106 PWVEQSSFRHFLISIPYPYQQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL C1HYY6_NEIGO 106 PWVEQSSFRHFLISIPYPYQQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL AKR88_NEIMFI 123 PRVGQSSFRHFLISIPYPYQQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL C0DN13_NEICI 126 PWVEQSSFWHFLISIPYPYQQELEYMDNVRFFWLLPITQPERLFINTHSVEELETKFDEAGIDYL AlkSR8_NEIMFI 135 PWVEQSSFWHFLISIPYPYQQELEYMDNVRFFWLLPITQPERLFINTHSVEELETKFDEAGIDYL C0DN13_NEICI 126 PWVEQSSFWHFLISIPYPYQQELEYMDNVRFFWLLPITQPERLFINTHSVEELETKFDEAGIDYL C0F5L2_NEILA 126 PWVEQSSFWHFLISIPYPYQQELEYMDNVRFFWLLPITQPERLFINTHSVEELETKFDEAGIDYL C0F5L2_NEILA 126 <th>COF5L2 NEILA</th> <th>61</th> <th>QVIPNHEDEPWVYVSSGI</th> <th>GSFLGQ</th> <th>EFFIVSPF</th> <th>ETPEHIET</th> <th>LAMLASASM</th> <th>HYPDRFQLG</th> <th>KTVNIGR</th>	COF5L2 NEILA	61	QVIPNHEDEPWVYVSSGI	GSFLGQ	EFFIVSPF	ETPEHIET	LAMLASASM	HYPDRFQLG	KTVNIGR
C1YUI8_NOCDA 46 RVAPRNEGEGWTYVTVGSVHEAGGGGLEFFLVAPTAERAHAETLGWVSHLHSFEEYRLDVGSVIDLGG C2A7G8_THECU 3 RITPASAQDAWYVTKGAFEADGGQGLEFIIEAPAEDPIHVESLAMVTYRH.LTGPPLDLGQTVAIGR B5HWZ3_9ACTO 47 VVEPGPRSDSWAFVTAGCWASMEGNGHGLEFVMTAPVRDQRFIDLMAMIAHYHCGG.HQLALEHSMPIGE Q5F628_NEIGI 47 VVEPGPRSDSWAFVTAGCWASMEGNGHGLEFVMTAPVRDQRFIDLMAMIAHYHCGG.HQLALEHSMPIGE Q5F628_NEIGI 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL Q5F628_NEIGI 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL B4RNC6_NEIG2 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL B4RNC6_NEIG2 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL C0EME3_NEIFLI 123 PRVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFIDTHLIELETKFDEAGIDYL C0DN13_NEICI 126 PWVQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFIDTHSVEDLETKFDEAGIDYL C0F5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFIDTHSVEDLETKFDEAGIDYL C0F5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITHQERLFIDTHSVEDLETKFDEAGIDYL C0F5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITHQERLFIDTHSVEDLETKFDEAGIDYL C0F5L2_NEILA 126 PWVGQSSFWHF	A6A6Q5_VIBCH	59	RFAPGPQINMWVYCSV <mark>G</mark> A	SEITHSDSVLH	EFVVVSPI	ESNRLVEM	LAMVTYYHS	NHNLGFG	HTLPIGE
C2A7G9_THECU 3 RITEASA0DAWVYVTRCAFEADGGORLEFIIEAEAEDFIHVESLAWVTYRH.LTGPPLDLGOTVAIGR B5HWZ3_9ACTO 47 VVEPGPRSDSWAFVTAGCWASMEGNGHGLEFVMTAPVRDQRFIDLMAMIAHYHCGG.HQLALEHSMPIGE 05F628_NEIGI 06 H6 β7 H7 H8 025F628_NEIGI 06 PWVEQSSFRHFLISLPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL 025F628_NEIGI 06 PWVEQSSFRHFLISLPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL 04RNC6_NEIG2 106 PWVEQSSFRHFLISLPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL 04RNC6_NEIG2 106 PWVEQSSFRHFLISLPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL 04RNC6_NEIG2 106 PWVEQSSFRHFLISLPYPYGQELEYMDNVRFFWLLPITQERLFLDTHLEELETKFDEAGIDYL 04RNC6_NEIG2 106 PWVEQSSFRHFLISLPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL 020000 0000000000 VERFWLLPITHQERLFLDTHSVEELETKFDEAGIDYL 020000 000000000000000000000000000000000000	C1YUI8_NOCDA	46	RVAPRNEGEGWTYVTVGS	SVHEAGGGT	EFFLVAPT	AERAHAET	LGMVSHLHS	FEEYRLDVG	SVIDLGG
β6 H6 β7 H7 H8 Q5F6Z8_NEIGI 0000 0000000000 0000000000 Q5F6Z8_NEIGI 0000 00000000000 Q5F6Z8_NEIGI 0000 00000000000 Q5F6Z8_NEIGI 0000 00000000000 Q5F6Z8_NEIGI 0000 0000000000000 Q5F6Z8_NEIGI 0000 000000000000000000000000000000000000	CZA/G8_THECU B5HWZ3 9ACTO	47	RITPASAQDAWVYVIKGA VVEPGPRSDSWAEVTAGC	FEADGGQRL	EFIIEAPA EFVMTAPV	EDPIHVES.	LAMVIYRH.	G HOLALE	UTVALGR HSMPTCE
β6 H6 β7 H7 H8 Q5F628_NEIGI 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL C1HYY6_NEIGO 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL B4RNC6_NEIG2 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL B4RNC6_NEIG2 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL C0EME3_NEIFL 123 PRVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQPERLFLDTHLIELETKFDEAGIDYL A1KSR8_NEIMF 135 PWVQQSFGUHFLISIPYPYGQELEYMDNVRFFWLLPITQPERLFLDTHSVEDLETKFDEAGIDYL C0DN13_NEICI 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLDTHSVEELETKFDEAGIDYL C0F5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFLDTHSVEELETKFDEAGIDYL C0F5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVR	2011110_011010	•							
Q5F6Z8_NEIGI Q5F6Z8_NEIGI Q5F6Z8_NEIGI 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL B4RNC6_NEIG2 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL B4RNC6_NEIG2 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFINTHSVEELETKFDEAGIDYL A1KSR8_NEIHF 135 PWVEQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFIDTHLIELETKFDEAGIDYL A1KSR8_NEIHF 135 PWVEQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFIDTHSVEELETKFDEAGIDYL CODNI3_NEICI 126 PWVQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFIDTHSVEELETKFDEAGIDYL COF5L2_NEILA COF5			86	LIG		0.7	117	110	
Q5F622_NEIGI Q5F628_NEIGI 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL CHYY6_NEIG2 106 PWVEQSSFRHFLISIPYPYGQELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL ATKSR8_NEIGT 123 PRVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQERLFLNTHSVEELETKFDEAGIDYL ATKSR8_NEIMF 135 PWVEQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQPERLFLNTHSVEELETKFDEAGIDYL CODN13_NEICI 126 PWVQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITQPERLFLDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFLDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFLDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFLDTHSVEELETKFDEAGIDYL COF5L2_NEICH 127 PWLEGSNCENNLISIPYPFGEDLEIMPLNSSH.VHVAWLLPITDDERNFKIKYDLDALEQKFDDAEFENYW	055660 NETCI		ро		1	57	н/	нв	0.0
Q5F628_NEIGI 106 PWVEQSSFRHFLISLPYPYGQELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL C1HYY6_NEIGO 106 PWVEQSSFRHFLISLPYPYGQELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL B4RNC6_NEIGI 106 PWVEQSSFRHFLISLPYPYGQELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL C0EME3_NEIFL 123 PRVGQSSFRHFLISLPYPYGQELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL C0EME3_NEIFL 125 PRVGQSSFQHFLISLPYPYGQELEYMDNVRFFWLLPITHQERLFLDTHLIELETKFDEAGIDYL C0DN13_NEICI 126 PWVQQSSFQHFLISLPYPYGQELEYMDNVRFFWLLPITQPERLFLDTHSVEELETKFDEAGIDYL C0F5L2_NEILA 126 PWVGQSSFWHFLISLPYPYGQELEYMDNVRFFWLLPITRQERLFLDTHSVEELETKFDEAGIDYL C0F5L2_NEILA 127 PWVGQSSFWHFLISLPYPYGQELEYMDNVRFFWLLPITRQERLFLDTHSVEELETKFDEAGIDYL C0F5L2_NEILA 126 PWVGQS	Q5F6Z8_NEIGI			leee.			eeee.		· ·
C1HYO_NEIGO 106 PWVEQSSFRHFLISLPYPYGQELEYMDNVRFFWLLPITQTERLFLNTHSVEELETKFDEAGIDYL B4RNC6_NEIGI 106 PWVEQSSFRHFLISLPYPYGQELEYMDNVRFFWLLPITQTERLFLDTHLIEELETKFDEAGIDYL A1KSR8_NEIMF 135 PWVEQSSFWHFLISLPYPYGQELEYMDNVRFFWLLPITQPERLFLDTHLIEELETKFDEAGIDYL CODN13_NEICI 126 PWVDQSSFWHFLISLPYPYGQELEYMDNVRFFWLLPITQPERLFLDTHSVEELETKFDEAGIDYL CODSL_NEILA 126 PWVDQSSFWHFLISLPYPYGQELEYMDNVRFFWLLPITQPERLFLDTHSVEELETKFDEAGIDYL A6A6Q5_VIBCH 127 PWLEGSNCENWLISLPYPYGQELEYMDNVRFFWLLPITRQERLFLDTHSVEELETKFDEAGIDYL C1YUIS NOCDA 114 PWTDGSRMDHLVSSPYYGPALEWAPEEAGG.ARFLWLLPITDDERNFKIKYDLDAFEELLEAEGVNVL	Q5F6Z8_NEIG1	106	PWVEQSSFRHFLISLPYP	YGQELEYMDN.	VRFF	WLLPITQT	BLFLNTHS	VEELETKFD	EAGIDYL
COEMES_NEIFL 123 PRVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITHQERLFLDTHLIEELETKFDEAGIDYL A1KSR8_NEIMF 135 PWVEQSSFQHFLISIPYPYGQELEYMDNVRFFWLLPITQPERLFLDTHSVEDLETRFDELGIDYL CODN13_NEICI 126 PWVDQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITHQERLFLDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISIPYPYGQELEYMDNVRFFWLLPITRQERLFLDTHSVEELETKFDEAGIDYL A6A6Q5_VIBCH 127 PWLEGSNCENWLISIPYPFGEDLEIMPLNSSH.VHVAWLLPITRDERNFKIKYDLDALEQKFDDAELKY C1YUI8 NOCDA 114 PWTDGSRMDHLLVSPFYGPALEWAPEEAGG.ARFLWLLPIHRSEAALIGTESLDAFFELLEAEGVNVL	BARNCE NEIGO	106	PWVEQSSFRHFLISLPYP PWVEQSSFRHFLISLPYP	YCOELEYMDN.	VRFF	WLLPITOT	BLFLNIHS	VEELEIKED	FAGIDYL
A1KSR8_NEIMF 135 PWVEQSSFQHFLISLPYPYGQELEYMDNVRFFWLLPITQPERLFLNTHSVEDLETRFDELGIDYL CODN13_NEICI 126 PWVDQSSFWHFLISLPYPYGQELEYMDNVRFFWLLPITHQERLFLDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISLPYPYGQELEYMDNVRFFWLLPITRQERLFLDIHSVEELETKFDEAGIDYL A6A6Q5_VIBCH 127 PWLGSSNCENWLISLPYPFGEDLEIMPLNSSH.VHVAWLLPITRDERNFKIKYDLDALEQKFDDAELKYL C1YUI8 NOCDA 114 PWTDGSRMDHLLVSPFYGPALEWAPEEAGG.ARFLWLLPITRSEAALTGTESLDAFEELLEAEGVNVL	COEMES NEIFL	123	PRVGOSSFWHFLISLPYP	YGOELEYMDN.	VRFF	WLLPITHO	ERLFLDTHL	IEELETKFD	EAGIDYL
CODN13_NEICI 126 PWVDQSSFWHFLISLPYPYGQELEYMDNVRFFWLLPIIHQERLFLDTHSVEELETKFDEAGIDYL COF5L2_NEILA 126 PWVGQSSFWHFLISLPYPYGQELEYMDNVRFFWLLPITRQERLFLDIHSVEELETKFDEAGIDYL A6A6Q5_VIBCH 127 PWLEGSNCENWLISLPYPFGEDLEIMPLNSSH.VHVAWLLPITRDERNFKIKYDLDALEQKFDDAELEY C1YUI8 NOCDA 114 PWTDGSRMDHLLVSPEYFYGPALEWAPEEAGG.ARFLWLLPIHRSEAALIGTESLDAFEELLEAEGVNVL	A1KSR8_NEIMF	135	PWVEQSSFQHFLISLPYP	YGQELEYMDN.	VRFF	WLLPITQPI	ERLFLNTHS	VEDLETRFD	ELGIDYL
COF5L2_NEILA 126 FWVGQSSFWHFLISLEYFYGQELEYMDNVRFFWLLFITROFRLFIDIHSVEELETKFDEAGIDYL A6A6Q5_VIBCH 127 FWLEGSNCENWLISLFYFFGEDLEIMFLNSSH.VHVAWLLFITROFRNFKIKYDLDALEQKFDDAELY C1YUIS NOCDA 114 FWTDGSRMDHLLVSPFYFGFALEWAPEEAGG.ARFLWLLFIHRSFAALIGTESLDAFEELLEAEGVNVL	CODN13_NEICI	126	PWVDQSSFWHFLISLPYP	YGQELEYMDN.	VRFF	WLLPIIHQ	ERLFLDTHS	VEELETKFD	EAGIDYL
CIVUIS NOCDA 114 PWILESCHOPHILISSPEVEVGPALEWAPERAGE, ARFLWLLPIHRSBAALIGTESLDAAFEELDAAGEVNVL	COF5L2_NEILA	126	PWVGQSSFWHFLISLPYP	YGQELEYMDN.	VRFF	WLLPITRQ.	BRLFLDIHS	VEELETKFD	EAGIDYL
	C1YUI8 NOCDA	114	PWTDGSRMDHLLVSPPYP	YGPALEWAPEE	AGG. ARFL	WLLPIHRS	PAALIGTES	LDAFEELLE	AEGVNVL
C2A7G8_THECU 70 PWITGSPQDHFMVDLPYPFGPKLEWAELPNGEPVRFLWLVPITASEAEYARTHGADALGDLLEQNQADVL	C2A7G8_THECU	70	PWITGSPQDHFMVDLPYP	FGPKLEWAELP	NGEPVRFL	WLVPITAS	AEYARTHG	ADALGDLLE	QNQADVL
B5HWZ3_9ACTO 116 <u>Pwv</u> pgStceHllvslPYLHGpDLEHCPlpggH. <u>Ar</u> ILWILPVTTABIEFRRRHGHEALEQLFDEAEIIPT	B5HWZ3_9ACTO	116	PWVPGSTCEHLLVSLPYL	HGPDLEHCPLP	GGH.ARIL	WILPVTTA	IEFRRRHG	HEALEQLFD	EAEIIPT
Q5F628_NEIG1	Q5F6Z8_NEIG1								
OFF628 NEIGI 171 DINRASITVOAG	Q5F6Z8 NEIG1	171	DINRASTVWQAG						
CIHYY6_NEIGO 171 DINRASTVWQAG	C1HYY6_NEIGO	171	DINRASTVWQAG						
BABNC6 NETG2 171 DINBASTVWOAG	B4RNC6_NEIG2	171	DINRASTVWQAG						
	COEME3_NEIFL	188	DINRASTV						
COEMES_NEIFL 188 DINRASTV	CODN13 NETCT	200	DINBASTYWOAG						
COEMES_NEIFL 188 DINRASTV AIKSR8_NEIMF 200 DINRASVI CODNIS NEICI 191 DINRASTVW0AG	COF5L2 NEILA	191	DINRASTV						
COEMES_NEIFL 188 DINRASTVV ALKSR8_NEIMF 200 DINRASVV CODN13_NEICI 191 DINRASTVWQAG COF5L2 NEILA 191 DINRASTVV	A6A6Q5_VIBCH	196	EVKRSSVV						
COEMES_NEIFL 188 DINRASTVV AlksR8_NEIMF 200 DINRASVI CODN13_NEICI 191 DINRASTVWQAG COF5L2_NEILA 191 DINRASTVV AGAGQ5_VIBCH 196 EVKRSSVV	C1YUI8_NOCDA	183	DLDRPPLF						
COEMES_NEIFL 188 DINRASTVV AlKSR8_NEIMF 200 DINRASVI CODN13_NEICI 191 DINRASTVWQAG COF5L2_NEILA 191 DINRASTV AGAG65_VIBCH 196 EVKRSSVV CLYUI8_NOCDA 183 DLDRPPLF	CZA7G8 THECU	140	DPQRRSVV						

Figure 3. Multiple sequence alignment of Sufu-like proteins. The Neisseria proteins with UniProt codes Q5F6Z8 (NGO1391) from *N. gonorrhoeae* FA1090, C1HYY6 from *N. gonorrhoeae* 1291, B4RNC6 from *N. gonorrhoeae* NCCP11945, C0EME3 from *N. flavescens* NRL30031/H210, A1KSR8 from *N. meningitidis* serogroup C/serotype 2a, C0DN13 from *N. cinerea* ATCC14685 and C0F5L2 from *N. lactamica* ATCC23970 share ~92% sequence identity. More distant homologs with ~33–39% sequence identity are B5HWZ3 from *Streptomyces sviceus* ATCC 29083, C2A7G8 from *Thermomonospora curvata* DSM 43183, C1YUI8 from *Nocardiopsis dassonvillei* subsp. dassonvillei DSM 43111, and A6A6Q5 from *Vibrio cholerae* MZO-2. The alignment was generated using the CLUSTALW web-server.²⁰

imidazole, 1 mM Tris(2-carboxyethyl)phosphine-HCl (TCEP)] and the lysate was clarified by centrifugation at 32,500 \times g for 30 minutes. The soluble fraction was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with lysis buffer, the resin was washed with wash buffer [50 mM HEPES pH 8.0, 300 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP], and the protein was eluted with elution buffer [20 mM HEPES pH 8.0, 300 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP]. The eluate was buffer exchanged with TEV buffer [20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP] using a PD-10 column (GE Healthcare), and incubated with 1mg of TEV protease per 15 mg of eluted protein. The protease-treated eluate was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with HEPES crystallization buffer [20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP] and the resin was washed with the same buffer. The flow-through and wash fractions were combined and concentrated for crystallization trials to 13 mg/mL by centrifugal ultrafiltration (Millipore). NGO1391 was crystallized by mixing 200 nL protein solution with 200 nL crystallization solution in a sitting drop format over a 50 μ L reservoir volume using the nanodroplet vapor



Figure 4. Electrostatic surface potential representations of human Sufu and NGO1391. (a) Surface for the NTD of human Sufu shows a greater accumulation of negatively charged residues on one side (left panel) of the protein due to surface residues Glu106, Asp111, Glu152, Asp159, Glu181, Glu221, and Asp262. (b) Surface for the NGO1391 indicates that charged residues are equally distributed on both sides of the protein. The color scale is in units of kT/e from -9 to +9. The figure in the right panel represents a 180° rotation around the vertical axis compared to the left panel.

diffusion method²⁹ with standard JCSG crystallization protocols.³⁰ The crystallization reagent consisted of 1.6M ammonium sulfate, and 0.1M MES pH 6.0. Glycerol was diluted using the reservoir solution and then added 1:1 to the drop to a final concentration 20% (v/v) as a cryoprotectant prior to harvesting. A cubic-shaped crystal of approximate size 100

 $\mu m \times 100 \ \mu m \times 100 \ \mu m$ was harvested after 15 days at 277 K for data collection. Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM)^{31} at the Stanford Synchrotron Radiation Lightsource (SSRL, SLAC National Accelerator Laboratory, Menlo Park, CA). The diffraction data were indexed in trigonal space



Figure 5. Sequence conservation plot of NGO1391. The Sufu-like proteins from Figure 3 have been used for the calculation of conservation scores²¹ (a) A stereo stick representation in which the most conserved residues are represented in dark pink, the light pink residues are less conserved and residues in cyan are least conserved. (b) and (c) Two views of some of the most conserved surface exposed residues in the same color scheme. The molecular orientations in the top and bottom panels are different to facilitate visualization using PyMOL.

group $P3_221$. Protein concentrations were determined using the Coomassie Plus assay (Pierce). To determine its oligomeric state in solution, NGO1391 was analyzed using a 1×30 cm² Superdex 200 size-exclusion column (GE Healthcare). The mobile phase consisted of 20 mM Tris pH 8.0, 150 mM NaCl, and 0.02% (w/v) sodium azide. The molecular weight was calculated using ASTRA 5.1.5 software (Wyatt Technology).

Data collection, structure solution, and refinement

SAD data were collected at 100 K using a MarMosaic 325 CCD detector (Rayonix) at the SSRL on beamline 11–1 at the selenium peak (λ_1) wavelength using the BLU-ICE³² data collection environment. The SAD data were integrated and reduced using XDS³³ and XSCALE, respectively. All other programs for data manipulation were from the CCP4 suite.³⁴ The heavy atom substructure and phasing calculations were performed using SHELXD³⁵ and autoSHARP³⁶, respectively, and ARP/wARP³⁷ was used for automatic model building to 1.40 Å resolution. Model completion and crystallographic refine-COOT³⁸ ment were performed using and REFMAC5,³⁹ respectively. The refinement protocol included the experimental phase restraints in the form of Hendrickson-Lattman coefficients from autoSHARP. Data and refinement statistics are summarized in Table I.

Validation and deposition

The quality of the crystal structure was analyzed using the JCSG Quality Control server (http:// smb.slac.stanford.edu/jcsg/QC). This server automatically processes the coordinates and data through a variety of validation tools including Auto-DepInputTool,⁴⁰ MolProbity,¹⁶ WHATIF 5.0;⁴¹ agreement between the atomic model and the data using SFcheck 4.0⁴² and RESOLVE,⁴³ the protein sequence using CLUSTALW,⁴⁴ atomic occupancies using MOLEMAN2,45 consistency of NCS pairs and evaluates difference in $R_{\rm cryst}/R_{\rm free}$, expected $R_{\rm free}/R_{\rm cryst}$ and maximum/minimum B-factors by parsing the refinement log-file and PDB header. Protein quaternary structure analysis was performed using the PISA server,⁴⁶ Figure 1(B) was adapted from an analysis using PDBsum,⁴⁷ and all others were prepared with PyMOL.48 Figure 3 was prepared using the PDB2PQR⁴⁹ server and the APBS^{49,50} module in PyMOL. Atomic coordinates and experimental structure factors for NGO1391 have been deposited in the PDB under the accession code 3k5j.

Acknowledgment

Portions of this research were performed at SSRL. The SSRL is a national user facility operated by SLAC National Accelerator Laboratory and Stanford University on behalf of the United States Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health (National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. Genomic DNA from *Neisseria gonorrhoeae* FA 1090 (ATCC Number: 700825D) was obtained from the American Type Culture Collection (ATCC).

References

- 1. McMahon AP, Ingham PW, Tabin CJ (2003) Developmental roles and clinical significance of hedgehog signaling. Curr Top Dev Biol 53:1–114.
- 2. Jiang J, Hui CC (2008) Hedgehog signaling in development and cancer. Dev Cell 15:801–812.
- 3. Varjosalo M, Taipale J (2008) Hedgehog: functions and mechanisms. Genes Dev 22:2454–2472.
- 4. MacDonald BT, Tamai K, He X (2009) Wnt/beta-catenin signaling: components, mechanisms, and diseases. Dev Cell 17:9–26.
- 5. Pasca di Magliano M, Hebrok M (2003) Hedgehog signalling in cancer formation and maintenance. Nat Rev Cancer 3:903–911.
- Lai SL, Chien AJ, Moon RT (2009) Wnt/Fz signaling and the cytoskeleton: potential roles in tumorigenesis. Cell Res 19:532–545.
- 7. Burglin TR (2008) The Hedgehog protein family. Genome Biol 9:241.
- 8. Burglin TR (2008) Evolution of hedgehog and hedgehog-related genes, their origin from Hog proteins in ancestral eukaryotes and discovery of a novel Hint motif. BMC Genomics 9:127.
- 9. Hausmann G, von Mering C, Basler K (2009) The hedgehog signaling pathway: where did it come from? PLoS Biol 7:e1000146.
- Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A (2008) The Pfam protein families database. Nucleic Acids Res 36:D281–D288.
- Dunaeva M, Michelson P, Kogerman P, Toftgard R (2003) Characterization of the physical interaction of Gli proteins with SUFU proteins. J Biol Chem 278: 5116-5122.
- 12. Merchant M, Vajdos FF, Ultsch M, Maun HR, Wendt U, Cannon J, Desmarais W, Lazarus RA, de Vos AM, de Sauvage FJ (2004) Suppressor of fused regulates Gli activity through a dual binding mechanism. Mol Cell Biol 24:8627–8641.
- Jaroszewski L, Rychlewski L, Li Z, Li W, Godzik A (2005) FFAS03: a server for profile–profile sequence alignments. Nucleic Acids Res 33:W284–W288.
- Cruickshank DW (1999) Remarks about protein structure precision. Acta Crystallogr D Biol Crystallogr 55: 583–601.
- Matthews BW (1968) Solvent content of protein crystals. J Mol Biol 33:491–497.
- 16. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35:W375–W383.
- Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. Bioinformatics 24:2780–2781.
- Ye Y, Godzik A (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics 19(Suppl 2):246–255.
- 19. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure

alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 60:2256–2268.

- Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics Chapter 2: Unit 2.3.
- Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res 33:W299–W302.
- 22. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 37:D412–D416.
- 23. Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, Klee SR, Morelli G, Basham D, Brown D, Chillingworth T, Davies RM, Davis P, Devlin K, Feltwell T, Hamlin N, Holroyd S, Jagels K, Leather S, Moule S, Mungall K, Quail MA, Rajandream MA, Rutherford KM, Simmonds M, Skelton J, Whitehead S, Spratt BG, Barrell BG (2000) Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491. Nature 404:502–506.
- 24. Bentley SD, Vernikos GS, Snyder LA, Churcher C, Arrowsmith C, Chillingworth T, Cronin A, Davis PH, Holroyd NE, Jagels K, Maddison M, Moule S, Rabbinowitsch E, Sharp S, Unwin L, Whitehead S, Quail MA, Achtman M, Barrell B, Saunders NJ, Parkhill J (2007) Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. PLoS Genet 3:e23.
- 25. Krishna SS, Weekes D, Bakolitsa C, Elsliger MA, Wilson IA, Godzik A, Wooley J (2010) TOPSAN: Use of a Collaborative Environment for Annotating, Analyzing and Disseminating Data on JCSG and PSI structures. Acta Crystallogr F Struct Biol Cryst Commun 66: 1143–1147.
- Weekes D, Krishna SS, Bakolitsa C, Wilson IA, Godzik A, Wooley J (2010) TOPSAN: a collaborative annotation environment for structural genomics. BMC Bioinf 11:426.
- 27. Klock HE, Koesema EJ, Knuth MW, Lesley SA (2008) Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. Proteins 71:982–994.
- Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J Mol Biol 229:105–124.
- 29. Santarsiero BD, Yegian DT, Lee CC, Spraggon G, Gu J, Scheibe D, Uber DC, Cornell EW, Nordmeyer RA, Kolbe WF, Jin J, Jones AL, Jaklevic JM, Schultz PG, Stevens RC (2002) An approach to rapid protein crystallization using nanodroplets. J Appl Crystallogr 35:278–281.
- 30. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, Vincent J, Robb A, Brinen LS, Miller MD, McPhillips TM, Miller MA, Scheibe D, Canaves JM, Guda C, Jaroszewski L, Selby TL, Elsliger MA, Wooley J, Taylor SS, Hodgson KO, Wilson IA, Schultz PG, Stevens RC (2002) Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci USA 99:11664–11669.
- Cohen AE, Ellis PJ, Miller MD, Deacon AM, Phizackerley RP (2002) An automated system to mount cryocooled protein crystals on a synchrotron beamline, using compact sample cassettes and a small-scale robot. J Appl Crystallogr 2002:720–726.

- 32. McPhillips TM, McPhillips SE, Chiu HJ, Cohen AE, Deacon AM, Ellis PJ, Garman E, Gonzalez A, Sauter NK, Phizackerley RP, Soltis SM, Kuhn P (2002) Blu-ice and the distributed control system: software for data acquisition and instrument control at macromolecular crystallography beamlines. J Synchrotron Radiat 9: 401–406.
- Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Crystallog 26:795–800.
- 34. Collaborative Computing Project N(1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50:760–763.
- Sheldrick GM (2008) A short history of SHELX. Acta Crystallogr A 64:112–122.
- Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. Methods Mol Biol 364:215–230.
- 37. Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc 3:1171–1179.
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66:486-501.
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximumlikelihood method. Acta Crystallogr D Biol Crystallogr 53:240-255.
- 40. Yang H, Guranovic V, Dutta S, Feng Z, Berman HM, Westbrook JD (2004) Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. Acta Crystallogr D Biol Crystallogr 60: 1833–1839.
- 41. Vriend G (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graph 8:52–56, 29.
- 42. Vaguine AA, Richelle J, Wodak SJ (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr D Biol Crystallogr 55:191–205.
- Terwilliger TC (2000) Maximum-likelihood density modification. Acta Crystallogr D Biol Crystallogr 56: 965–972.
- 44. Thompson JD, Higgins DG, Gibson TJ (1994) CLUS-TAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680.
- 45. Kleywegt GJ (2000) Validation of protein crystal structures. Acta Crystallogr D Biol Crystallogr 56:249–265.
- Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774–797.
- 47. Laskowski RA, Chistyakov VV, Thornton JM (2005) PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. Nucleic Acids Res 33:D266–D268.
- DeLano WL (2008) The PyMOL molecular graphics system. DeLano Scientific LLC: Palo Alto, CA, USA.
- 49. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res 35:W522–W525.
- 50. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 98:10037–10041.