

# Acta Crystallographica Section D

Volume 70 (2014)

Supporting information for article:

**Structure of sulfamidase provides insight into the molecular pathology of mucopolysaccharidosis IIIA**

**Navdeep S. Sidhu, Kathrin Schreiber, Kevin Pröpper, Stefan Becker, Isabel Usón, George M. Sheldrick, Jutta Gärtner, Ralph Kräzner and Robert Steinfeld**

```

Cf MSWNNRLKFNISLCT-L-IISLALCI THGDADTAPKNVLLLLADAGFEMRSYLNKICQ
Cg -----MDKLF-TFLKCLTLLSH--FODSCKNVLTVQDDAGFETOVYNNISCK
Dr -----MAVFQAWTLLCLLLCFD--VGGCRSRNVLLIADGGFETOVYNNITVVO
Ss -----MAYGNHNVLLLLAVIC--VGECKRNVLLIADGGFETSVYNNITVVR
Xt -----MISMGSRWYVWI--YSLLL--LSKAWGRNVLLIADGGFESVYNNITAIH
Bt -----MRSSTGGCC--WVLLISVGLCC--VHRARPRNVLLIADGGFESGAYNNSAIS
Cl -----MRRRGRACGLLLLLLALGLCR--PLGARPRNVLLIADGGFESGAYNNSAIS
Mm -----MHCPLACCTI--LLVLGVC--GAHSRNVLLIADGGFESGAYNNSAIA
Hs 1 -----MSCPVPACAL--LLVLGLC--RARPRNVLLIADGGFESGAYNNSAIA
Tc -----MRRRPACCAL--LLVLGLC--RARPRNVLLIADGGFESGAYNNSAIA
      * * * * *

Cf TPNLDLAKESLLFNNAFTSVSSSPSASLLTGLPQHONGMYGLHHGHHFNSFDDVQS
Cg -----TPFLNKLGSLSLVTHAYASVSSSPSSTILSGLPQHONGMYGLHODVHHFMSFDGVR
Dr -----TPHLRALSKRSLIFKNAFTSVSSSPSSTILTGLPQHONGMYGLHOGVHHFNSFDGVQS
Ss -----TPHLAALGRSLVFNQNAFTSVSSSPSSTILTGLPQHONGMYGLHOGVHHFNSFDGVQS
Xt -----TPNLRLSKRSLIFKNAFTSVSSSPSMAIIMTGLPQHONGMYGLHODVHHFNSFDDVRS
Bt -----TPHLDALARRSLVFRNAFTSVSSSPSASLLTGLPQHONGMYGLHODVHHFNSFDRVQS
Cl -----TPHLDALARRSLVFRNAFTSVSSSPSASLLTGLPQHONGMYGLHODVHHFNSFDDVRS
Mm -----TPHLDALARRSLVFRNAFTSVSSSPSASLLTGLPQHONGMYGLHODVHHFNSFDDVRS
Hs 47 -----TPHLDALARRSLVFRNAFTSVSSSPSASLLTGLPQHONGMYGLHODVHHFNSFDDVRS
Tc -----TPHLDALARRSLVFRNAFTSVSSSPSASLLTGLPQHONGMYGLHODVHHFNSFDDVRS
      * * * * *

Cf LPRILKKNIRTIIGIIGKAVGPNKYVYDFFAHTEENMSILQVGRNITIKLLVRFELSVQ
Cg -----LPILLHQANIHTGIIGKAVGPKSVYDFFEETEENMSILQVGRNITIKLLVRFELSVQ
Dr -----LPILLKRANIHTGIIGKAVGPGVYDFFAHTEENMSVQLQVGRNITIKLLVRFKFLSH
Ss -----LPILLGQANIHTGIIGKAVGPGVYDFFAHTEENMSVQLQVGRNITIKLLVRFKFFQTQ
Xt -----LPILLRQAGIRTIIGIIGKAVGPEVYDFFAHTEENMSVQLQVGRNITIKLLVRFKFLQT
Bt -----LPILLRQAGIRTIIGIIGKAVGPEVYDFFAHTEENMSVQLQVGRNITIKLLVRFKFLQT
Cl -----LPILLRQAGIRTIIGIIGKAVGPEVYDFFAHTEENMSVQLQVGRNITIKLLVRFKFLQT
Mm -----LPILLRQAGIRTIIGIIGKAVGPEVYDFFAHTEENMSVQLQVGRNITIKLLVRFKFLQT
Hs 107 -----LPILLRQAGIRTIIGIIGKAVGPEVYDFFAHTEENMSVQLQVGRNITIKLLVRFKFLQT
Tc -----LPILLRQAGIRTIIGIIGKAVGPEVYDFFAHTEENMSVQLQVGRNITIKLLVRFKFLQT
      * * * * *

Cf -----NKTQPFLLYAFHDFRCGHSQPYGAFCEKFGNGD-VGMGTPDWNPIYQW
Cg -----YKTSFFLYVGFHDFRCGHSQPYGAFCEKFGNSAEMGHDPDWPVPHYDP
Dr -----KEERSETKKEERPFLLYAFHDFRCGHSQPYGAFCEKFGNGE-SGMGRIPDWEKPYSSP
Ss -----RKE-----EERPFLLYAFHDFRCGHSQPYGAFCEKFGNGD-SGMGRIPDWEKPYSSP
Xt -----QDQRPFFLYVAFHDFRCGHSQPYGAFCEKFGNGD-PDMGIMPDSQYIYTP
Bt -----RQDQRPFFLYVAFHDFRCGHSQPYGAFCEKFGNGE-SGMGRIPDWPQIYNP
Cl -----QDDQRPFFLYVAFHDFRCGHSQPYGAFCEKFGNGE-SGMGRIPDWPQIYDP
Mm -----QDDQRPFFLYVAFHDFRCGHSQPYGAFCEKFGNGE-SGMGRIPDWPQIYDP
Hs 166 -----QDDQRPFFLYVAFHDFRCGHSQPYGAFCEKFGNGE-SGMGRIPDWPQIYDP
Tc -----QDDQRPFFLYVAFHDFRCGHSQPYGAFCEKFGNGE-SGMGRIPDWPQIYDP
      * * * * *

Cf EQVKVPYVONTAARRDIAAQYTTISRLDQGVGLVEELKNAGFKDNTLVIYTSDNIGIP
Cg -----QSVKVPYFIPDTPAARADIAAQYTTISRLDQIGLVLEELKXGVLQDITLVMYTSDNIGIP
Dr -----DQVKVPYFIPDTPAARADIAAQYTTISRLDQIGLVLEELKXGAFENDTLVIYTSDNIGIP
Ss -----DQVKVPYFVPDTPAARADIAAQYTTISRLDQIGLVLAELREAGYENDTLVIYTSDNIGIP
Xt -----EQVQVPYFIPDTPAARADIAAQYTTISGRMDQIGLVLELSEYNAGHENDTLVIYTSDNIGIP
Bt -----KDQVQVPYFVPDTPAARADIAAQYTTISGRMDQIGLVLELQELRGAGVNDTLVIYTSDNIGIP
Cl -----LDVLVQVPYFVPDTPAARADIAAQYTTISGRMDQIGLVLELQELRGAGVNDTLVIYTSDNIGIP
Mm -----QDVMVQVPYFVPDTPAARADIAAQYTTISGRMDQIGLVLELQELRGAGVNDTLVIYTSDNIGIP
Hs 218 -----LDVLVQVPYFVPDTPAARADIAAQYTTISGRMDQIGLVLELQELRDAGVNDTLVIYTSDNIGIP
Tc -----QDVLVQVPYFIPDTPAARADIAAQYTTISGRMDQIGLVLELQELRGAGVNDTLVIYTSDNIGIP
      * * * * *

Cf FPNCFITNLYDPGIAEPMIISPIHGRKNSVNTSMSTLDDITPTILDWFNITMDYPMFN
Cg -----FPGCFITNLYDPGMSPEFLVSSPYHKASWGRKSSAMTSLDDITPTILDWFKVYKYLK
Dr -----FPGCFITNLYGSGVKEPMLLSSPEHQORWGLKSOAYVSLDDITPTILDWFSLPYPSYLSM
Ss -----FPGCFITNLYGSGTAEPMLVSSPEHRRRWGTSOAYVSLDDITPTILDWFSISYPSYCLPG
Xt -----FPGCFITNLYWSGRAEPLLVSSPYHOKRWGTSOAYVSLDDITPTILDWFSIPYNYKIFG
Bt -----FPGCFITNLYWPGTAEPMLVSSPEHPRRWGTSOAYVSLDDITPTILDWFSIPYPSYAFG
Cl -----FPGCFITNLYWPGTAEPMLVSSPEHPRRWGTSOAYVSLDDITPTILDWFSIPYPSYAFG
Mm -----FPGCFITNLYWPGTAEPMLVSSPEHPRRWGTSOAYVSLDDITPTILDWFSIPYPSYAFG
Hs 278 -----FPGCFITNLYWPGTAEPMLVSSPEHPRRWGTSOAYVSLDDITPTILDWFSIPYPSYAFG
Tc -----FPGCFITNLYWPGTAEPMLVSSPEHRRRWGTSOAYVSLDDITPTILDWFSIPYPSYAFG
      * * * * *

Cf --EVSSPTGKSLPLLLKEPVE-NNTAIFASQTHHEVTMYPMRTI--RTKRHKLIHNI
Cg -----HPVQLTGKSVLPLLEVEAAESRDVYASHNLHEITMYPMRVLLRTKQYKLIHNI
Dr -----SOPVELTGRSLPALISEP--SWDTVFSQSLEHVTMYPMRSI--HKGPYRLLHNI
Ss -----TPTTPVSLTGRSLPALVSESPAAWHVYTSQSLEHVTMYPMRSI--HOGAYRLLHNI
Xt -----K-SVOLTGKSLPALQSEQ--DWTTFGQSQSHHEVTMYPMRSV--QNLQYLLHNI
Bt -----TKTVQLTGKSLPLVLEAEP--LWTVFGQSQSHHEVTMYPMRSV--HHQAYRLLHNI
Cl -----SKTVQLTGKSLPALAEP--LWTVFGQSQSHHEVTMYPMRSV--HHQAYRLLHNI
Mm -----SKTIQLTGKSLPALAEP--LWATVFGQSQSHHEVTMYPMRSV--YHNFRLHNI
Hs 338 -----SKTIQLTGKSLPALAEP--LWATVFGQSQSHHEVTMYPMRSV--QHHRFLHNI
Tc -----SKTVRLTGKSLPLVLEAEP--PWTVFGQSQSHHEVTMYPMRSV--HHQAYRLLHNI
      * * * * *

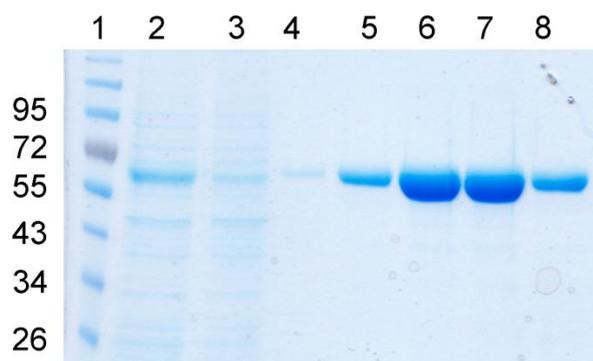
Cf NYRMPFPIQDQFYISPTFODLLNRTQTHOSLPWYKTLKYYQRPEWELYDLKYDPEELN
Cg -----NYLMPFPIQDQFYISPTFODLLNRSKSGEPTHSKSLSYYYRDRWELYDLMNDPQEKLN
Dr -----HYRMPFPIQDQFYISPTFODLLNRTQSGRPTGFKTLEYYRDRWELYDVRDPMKYN
Ss -----HYRMPFPIQDQFYISPTFODLLNRTLAERPTGFKTLQYYRDRWELYDVRDPEETVN
Xt -----NFKMPFPIQDQFYISPTFODLLNRTVSGQPTSMFKTLHYYRDRWELYDVRDISEIKN
Bt -----HFKMPFPIQDQFYISPTFODLLNRTTAGQPTGKYDLHYYRDRWELYDVRDPHETHN
Cl -----NFKMPFPIQDQFYISPTFODLLNRTVAGHTGKYDLHYYRDRWELYDVRDPHETR
Mm -----SFKMPFPIQDQFYISPTFODLLNRTTAGQPTGKYDLHYYRDRWELYDVRDPHETR
Hs 391 -----NFKMPFPIQDQFYISPTFODLLNRTTAGQPTGKYDLHYYRDRWELYDVRDPHETR
Tc -----NFKMPFPIQDQFYISPTFODLLNRTVAGRPTGKYDLHYYRDRWELYDVRDPHETHN
      * * * * *

Cf IASKSASIKFIELKERLYNMLKVTQDPMWICAPDGVLENAGHNKHAQCMPLNFD--
Cg -----LWAEPSHRKLEALKNKLFWQNTSDPMWICSPKGVLEDKGLYSSNPQCLPLYNGLDPI
Dr -----LAGDLDYSEVLESKLDLLKQWQRTDPMWICPDVLEAK--LEPCRPYLVNG--
Ss -----LAGDPALAPVLESRLHRLKQWQRTDPMWICPDVLEAK--LEPCRPYLVNG--
Xt -----IAEDPAYQDILKSMONILKQWQRTSDPMWICPDVLEAK--LEPCRPYLVNEL--
Bt -----LAADPYTQVLELLQTLKQWQRTDPMWICPDVLEAK--LAPCRPLHNEL--
Cl -----LAADPYVAVLELLQTLKQWQRTDPMWICPDVLEAK--LAPCRPLHNEL--
Mm -----LAADPDLAQVLEMLKQWQRTDPMWICPDVLEAK--LTPQCRPLHNEL--
Hs 451 -----LATDPRFAQLLEMLRDLAKQWQRTDPMWICPDVLEAK--LSPQCRPLHNEL--
Tc -----LAGDPRFAQVLELLKARLAKQWQRTDPMWICPDVLEAK--LTPQCRPLHNEL--
      * * * * *

Cf EQVKVPYVONTAARRDIAAQYTTISRLDQGVGLVEELKNAGFKDNTLVIYTSDNIGIP
Cg -----QSVKVPYFIPDTPAARADIAAQYTTISRLDQIGLVLEELKXGVLQDITLVMYTSDNIGIP
Dr -----DQVKVPYFIPDTPAARADIAAQYTTISRLDQIGLVLEELKXGAFENDTLVIYTSDNIGIP
Ss -----DQVKVPYFVPDTPAARADIAAQYTTISRLDQIGLVLAELREAGYENDTLVIYTSDNIGIP
Xt -----EQVQVPYFIPDTPAARADIAAQYTTISGRMDQIGLVLELSEYNAGHENDTLVIYTSDNIGIP
Bt -----KDQVQVPYFVPDTPAARADIAAQYTTISGRMDQIGLVLELQELRGAGVNDTLVIYTSDNIGIP
Cl -----LDVLVQVPYFVPDTPAARADIAAQYTTISGRMDQIGLVLELQELRGAGVNDTLVIYTSDNIGIP
Mm -----QDVMVQVPYFVPDTPAARADIAAQYTTISGRMDQIGLVLELQELRGAGVNDTLVIYTSDNIGIP
Hs 218 -----LDVLVQVPYFVPDTPAARADIAAQYTTISGRMDQIGLVLELQELRDAGVNDTLVIYTSDNIGIP
Tc -----QDVLVQVPYFIPDTPAARADIAAQYTTISGRMDQIGLVLELQELRGAGVNDTLVIYTSDNIGIP
      * * * * *

```

**Figure S1** Multiple sequence alignment of human SGSH and homologous sequences. All 10 important active site residues are highlighted. Numbering shown is for human SGSH. Sequences homologous to human SGSH were retrieved using the BLASTP server of the National Library of Medicine. Closest sequences annotated as SGSH were blindly selected from the results and sequence alignment performed using CLUSTAL Omega. Abbreviations: Cf, *Camponotus floridanus* (Florida carpenter ant), EFN70845.1; Cg, *Crassostrea gigas* (Pacific oyster), EKC39134.1; Dr, *Danio rerio* (zebrafish), NP\_001116740.1; Ss, *Salmo salar* (Atlantic salmon), NP\_001133784.1; Xt, *Xenopus tropicalis* (western clawed frog), NP\_001120065.1; Bt, *Bos taurus* (domestic cow), NP\_001095659.2; Cl, *Canis lupus familiaris* (dog), NP\_001003114.1; Mm, *Mus musculus* (house mouse), AAF29460.1; Hs, *Homo sapiens*, NP\_000190.1; Tc, *Tupaia chinensis* (Chinese tree shrew), ELW68227.1.



**Figure S2** SDS-PAGE image showing several steps of SGSH purification. Lane 1: Protein size markers, molecular weights are indicated at the left side in kDa. Lane 2: crude supernatant from cell culture containing recombinantly expressed SGSH. Lane 3: Flow through. Lane 4 and 5: Fractions eluted from the column at 100 mM imidazole. Lane 6 to 8: Fractions eluted from the column at 250 mM imidazole that were used for crystallisation trials.