Protein tetratricopeptide repeat (TPR) and the companion non-TPR helices: Bioinformatic analysis of interhelical interactions

Supplementary material

Figure S1: Theoretically predicted TCH sequences. Curated, nonredundant TPR protein sequences were collected from GenBank and UniProt sites, and the TPR and TCH sequences were delineated as described in Methods. The helices of the last TPR and the downstream TCH, and the signature amino acids of the TPR are highlighted and marked as in Table 1. A few sequences that did not show any secondary structure in the TCH region by prediction are so indicated (No TCH).

```
                    A-helix        B-helix             dTCH
                    (1-13)         (17-30)
                     W LG  Y        A   Y  A    P
                     4  78  11      20  24  27   32
A0AVF1(9)         YIYLSWLARCYIMNKKPRLAWELYLKMETSGESFSLLQLIANDCYKMGQFYYSAKAFDVLER
A6NLP5(3)         RQSFVQRGLLARLQGRDDDARRDFERAARLGSPFARRQLVLLNPYAALCNRMLADMMGQLRRPRDSR
O00170(3)         VKAYFKRGKAHAAVWNAQEAQADFAKVLELDPALAPVVSRELQALEARIRQKDEEDKARFRGIF
O15050(5)         LGKIRDAAYFYKRSQCYKDAFRCFEQIQEFDLALKMYCQEELFEEAAIAVEKYEEMLK
O15294(15)        PSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKPPYEHPKDLKLSDGR(No TCH)
O15460(2)         QLDAGEEATTTKSQVLDYLSYAVFQLGDLHRALELTRRLLSLDPSH
O75344(2)         AKALFRCGQACLLLTEYQKARDFLVRAQKEQPFNHDINNELKKLASCYRDYVDKEKEMWHRMF
O76094(12)        LSLIREAANFKLKYGRKKEAISDLQQLWKQNPKDIHTLAQLISAYS
O94906(12)        PHVLLAVAKLFWSQRKITKAREWFHRTVKIDSDLGDAWAFFYKFELQ
O95801(30         LKAIIRGALCHLELKHFAEAVNWCDEGLQIDAKEKKLLEMRAKADKLKRIEQRDVRKANLKEKKERNQ.
P09913(10)        SEALHVLAFLQELNEKMQQADEDSERGLESGSLIPSASSWNGE (No TCH)
P13674(2)         VSVLDYLSYAVYQQGDLDKALLLTKKLLELDPEHQRANGNLKYFEYIMAK
P30260(12)        SLVYFLIGKVYKKLGQTHLALMNFSWAMDLDPKGANNQIKEAIDKRY
P31948(9)         IKGYTRKAAALEAMKDYTKAMDVYQKALDLDSSCKEAADGYQRCMMAQYNR
P49321(5)         AETHYQLGLAYGYNSQYDEAVAQFSKSIEVIENRMAVLNEQVKE
P49792(3)         VHVNIRLVEVYRSTKRLKDAVAHCHEAERNIALRSSLEWNSCVVQTLKEYLES
P50502(3)         AQPYKWRGKAHRLLGHWEEAAHDLALACKLDYDEDASAMLKEVQPRAQKIAEHRRKYERKREEREIKE.
P53041(3)         IKGYYRRAASNMALGKFRAALRDYETVVKVKPHDKDAKMKYQECNKIVKQKAFERA
P53804(8)         IEFHMNCWKKLKTTTFNDKIDKDFLQGICLTPDCEGVISKIIIF
Q02790(3)         EKGLFRRGEAHLAVNDFELARADFQKVLQLYPNNKAAKTQLAVCQQRIRRQLAREKKLYANMFE
Q07617(9)         VKAFYRRALAHKGLKNYQKSLIDLNKVILLDPSIIEAKMELEEVTRLLNLK
Q08752(3)         TKALYRRAQGWQGLKEYDQALADLKKAQGIAPEDKAIQAELLKVKQKIKAQKDKEKAVYAK
Q12797(5)         GFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQ
Q13042(9)         ASTYSAIGYIHSLMGNFENAVDYFHTALGLRRDDTFSVTMLGHCIEMY
Q13099(12)        FLVRLCTDLGLKDAQEYARKLKRLEKMKEIREQRIKSGRDGSGGSRGKREGSASGDSGQ (No TCH)
Q13217(9)         VNALKDRAEAYLIEEMYDEAIQDYETAQEHNENDQQIREGLEKAQRLL
NP_001295302(3)   IKALFRKGKVLAQQGEYSEAIPILRAALKLEPSNKTIHAELSKLVKK
Q15785(6)         VKAFYRRAQAHKALKDYKSSFADISNLLQIEPRNGPAQKLRQEVKQNLH
Q17RB8(4)         PEVYFRKGKVLCDAGFLGDALQLFLQCLALDEDFAPAKLQVQKILC
Q1L5Z9(10)        ALCMRLLFEPVTTPCGHTFCLKCLERCLDHAPHCPLCKDKLSELLASRNFNIT (No TCH)
Q496Y0(4)         FKAHFRKAQALATLGKVEEALREFLYCVSLDGKNKRARCEAQRDNLE
Q49AM3(3)         PRGLFRLGKALMGLQRFREAAAVFQETLRGGSQPDAARELRSCLLHLTLQGQR
Q5I0X7(3)         EVPYYNRGLILYRLGYFDDALEDFKKVLDLNPGFQDATLSLKQTILDKEEKQRRNVAKNY
Q5SRH9(5)         PNALLELALLLMEQDRNEEAIKLLESAKQNYKNYSMESRTHFRIQAATLQAK
Q5T4D3(8)         ASYHGNLAVLYHRWGHLDLAKKHYEISLQLDPTASGTKENYGLLRRK
Q5TAA0(4)         PELQLLRGKCLRIKGEDANAAACFKRAVELDDAGSSHTDGFGCLLEALLAQW
Q5VTQ0(4)         PFTLFELASLYKSQGEIDKAIKFLETARNNYKDYSLESRLHFRIQAALHL
Q5W5X9(7)         AETYRLLGGADLAQGNHSGARKKLKKCLQIQTLLYGPQDKRTLATQQAMGMLSTAPKVASKPRQA
```

```
Q6P597(5)       AKTKNNLASAYLKQNKYQQAEELYKEILHKEDLPAPLGAPNTGTAGDAEQALRRSSSLSKIRESIRRGS
Q6PD62(19)      TEVVLYLARALFKCGKLQECKQTLLKARHVAPSDTVLMFNVALVLQRLATSVLKDEKSNLKEVL
Q6PGP7(40)      STARWYLLRHLYAKDDYELIDVLVNNAKTHGDTRALELNQRLSSQ
Q6PID6(4)       WESWQTLGRAQLGLGEIILAIRSFQVALHIYPMNPEIWKEDLSWARTLQEQQKVAQRIKKSEAPAEVTH
Q6ZXV5(11)      VQGKHNLCVVYFEEKDLLKAERCLLETLALAPHEEYIQRHLNIVRDKISSSSFIEPIF
Q7Z494(8)       GETLKNLAVLSYEGGDFEKAAELYKRAMEIKEAETSLLGGKAPSRHSSSGDTFSLK
Q7Z4L5(33)      PAVGYKLAFNYLKAKRYVDSIDICHQVLEAHPTYPKIRKDILDKARASLRP
Q86TV6(13)      HEVWNGLGEVLQAQGNDAAATECFLTALELEASSPAVPFTIIPRVL
Q86TZ1(12)      ALVYNFRAKVRGKIGLIEEAMADYNQALDLEDYASVI (end of sequence)
Q86UR1(3)       WEVLHNVASAQCQLGLWTEAASSLREAMSKWPEGSLNGLDSALDQVQRRGSLPPRQVPR
Q86WT1(11)      CIVNLVIGTLYCAKGNYEFGISRVIKSLEPYNKKLGTDTWYYAKRCFLSLLENMSKH
Q8IUR5(9)       AQAWMNMGGIQHIKGKYVSARAYYERALQLVPDSKLLKENLAKLDRLEKRLQEVREKDQT
Q8IWX7(5)       GKALIPLALEGTDVGKVKAAHALAKIAAVSNPDIAFPGERVYEVVRPLVRLLDTQRDGLQNYEALLGLT
Q8IYB4.2(7)     IRSRYNLGISCINLGAYREAVSNFLTALSLQRKSRNQQQVPHPAISGNIWAALRIALSLMDQPELFQAA
Q8N4P2(15)      DKKMYHLCIVNLVIGTLYCAKGNYDFGISRVIKSLEPYNKKLGTDTWYYAKRCFLSLLENMSKHTIMLR
Q8N5D0(3)       LKAHFRLARCLFELKYVAEALECLDDFKGKFPEQAHSSACDALGRDITAALFSKNDGEEKKGPGGGAP
Q8N5M4(3)       AKALYRAGVAFFHLQDYDQARHYLLAAVNRQPKDANVRRYLQLTQSELSSYHRKEKQLYLGMFG
Q8N6N2(3)       FKATYRAGIAFYHLGDYARALRYLQEARSREPTDTNVLRYIQLTQLKMNRCSLQREDSGAGSQTRD
Q8NBP0(7)       EVCQYMKGLSHVAMGQFYEGIKAQTKVMLNDPLPGQKASPEYLKVKYLREYSRYLHAHLDTPLTEYNID
Q8NDW8(32)      PAIGFKLAFNYLKDKKFVEAIEICNDVLREHPDYPKIREEILEKARRSLRP
Q8NEE8(11)      AQYYLYRAKSRQLLQNIFGARQDVATVLLLNPKQPKLSLLMTNLFPGMSVEEVLSTQIA
Q8NHQ1(3)       NGVYPRMNEVYTRLGEMNNAVRNLQELLELDSSSSLCVLVSTVGKLCRLINEDVN
Q8TAM2.2(8)     YEPHFNFATISDKIGDLQRSYVAAQKSEAAFPDHVDTQHLIKQLRQHFAML
Q8TE82(16)      RVAYHRLAALQHRLGHGELAEHFYLKALSLCNSPLEFDEETLYYVKVYLVLGDIIFYDLKDPFDAAGYY
Q8WXU2(3)       MKAHVRRGTAFCQLELYVEGLQDYEAALKIDPSNKIVQIDAEKIRNVIQGTELKS
Q92623(3)       FKALYRSGVAFYHLGDYDKALYYLKEARTQQPTDTNVIRYIQLTEMKLSRCSQREKEAM
Q96AE7(8)       AVNHFTLGNVYVAMEEFEKALVWYESTLKLQPEFVPAKNRIQTIQCHLMLKKGRRSP
Q96AY4(28)      AKIRHGLGLSLWASGNLEEAQHQLYRASALFETIRHEAQLSTDYKLSLFDLQTSSYQALQRVLVSLGHH
Q96HA7(10)      AKTWLNIALSREEAGDAYELLAPCFQKALSCAQQAQRPQLQRQVLQHLHTVQLRLQPQ
Q96N46(3)       CQTLVERGGQLEEEEKFLNAESYYKKALALDETFKDAEDALQKLHKYMQKSLELREKQAEKEEKQKTK
Q9BVA6(2)       VDALTEFGIFSEEDKDIIQADYLYTRALTISPYHEKALVNRDRTLPLVEEIDQRYFSIIDSKVKKVMSI
Q9BXJ9(11)      IETHLFAFEIYFRKEKFLLMLQSVKRAFAIDSSHPWLHECMIRLFNTAVCESKDL
Q9H3U1.1(3)     VKALYRRSQALEKLGRLDQAVLDLQRCVSLEPKNKVFQEALRNIGGQIQEKVRYMSSTDAKVEQMFQIL
Q9H892(3)       TKAYFHMGKANLALKNYSVSRECYKKILEINPKLQTQVKGYLNQVDLQEKADLQEKEAHEL
Q9HCD6(4)       YEAYYARARAKRSSRQFAAALEDLNEAIKLCPNNREIQRLLLRVEEECRQMQQPQQPP
Q9NZN9(3)       VKAYYVRARAHAEVWNEAEAKADLQKVLELEPSMQKAVRRELRLLENRMAEKQEEERLRCRNMLSQGA
Q9UGR2(3)       STAFRYLAQYYFKCKLWDEASTCAQKCCAFNDTREEGKALLRQILQLRNQGETPTTEVPA
Q9UJX3.4(8)     IKAVVKKAELLSREQKYEDGIALLRNALANQSDCVLHRILGDFLVAVNEYQEAMDQYSIA
Q9UJX4(6)       SETVLNLAFAQLILGIPEQALSLLHMAIEPILADGAILDKGRAMFLVAKCQVASAASYD
Q9ULT0.3(14)    HEAWQGLGEVLQAQGQNEAAVDCFLTALELEASSPVLPFSIIPREL (end of sequence)
Q9UNE7.2(3)     VKAHFFLGQCQLEMESYDEAIANLQRAYSLAKEQRLNFGDDIPSALRIAKKKRWNSIEERRIHQESE
Q9Y2Z0(3)       STAMLRKGICEYHEKNYAAALETFTEGQKLDIETGFHRVGQAGLQLLTSSDPPALDSQSAGITGADANF
Q9Y5Q9(13)      QESFYNLGRGLHQLGLIHLAIHYYQKALELPPLVVEGIELDQLDLRRDIAYNLSLIYQSSGNTG
Q9Y6J0(5)       NELYYLLADYHFKNKEQSKAIKFYMHDICICPNRFDSWAGMALARASRIQDKLNSNELKSDGPIWK
A0A087WTJ8(5)   GNLYSCIGNAQIELGQMEAALQSHRKDLEIAKEYDLPDAKSRALDNIGRVFARVGKFQQAIDT
A0A1C7CYW7(22)  ARMFLLRGQCCLEEQRHAEAWTAVESGLLVDPDHRGLKRLKARIRREASSGCWLQ
Q9UIM3(3)       LKALYRRGVAQAALGNLEKATADLKKVLAIDPKNRAAQEELGKVVIQGKNQDAGLAQGLRKMFG
A0A2P9AEM0(14)  PVLNDHLGDAYWRVGRRLEATFQWNQARDLKPDPDVLAALQQKLLKGLPPIESNTAQETPKVKP
A0A2P9AEQ5(6)   TALFLPLSSVLAQLGRREEARQMLLKFRPGVGQKGLENLADTLHLSFKWDYEHASVPERLNDGVRIA
A0A2P9AF65(3)   PTILSNPGGYQAVLGAIKAGAPWRGRLLIALGNTPVGLEFAERLLLDLRESSAPPSSNELASVISGYIR
A0A2P9AF75(4)   YPVRQTLGAVYMAMGEHEAAAAAFEHVLEQTPNNAWALWGLREVFRRTGGTADAEEMDARFKAAWVGAP
A0A2P9AGV7(5)   AWYYYDRGEALLVLGRYPEAVASFSCLPRKSAWQWARLAACHALDGNIDKARACAREGRRLNPGLTVQA
A0A2P9AGW1(6)   PYGYTNLAIVRAFFGHESQAAAAIATLRRRNPGFGVKNLRRSERYMRADDLERVINVLRAAGLRE
A0A2P9AGZ2(4)   ADAVFNLARLEFDAGNLAEARRNWSRYLQLDAGSEWARMAARGIQFVDLQLARMSAG
A0A2P9AIZ1(2)   TLVREYLGEAHLQIGDIAGAQEQLREIEKRTGKGSREYGMLSEQIDRFMKS
A0A2P9AJ95(8)   YDAARTRAEFETMKAVFDKAFFDKHRPSGIADDTPIFVVGMPRSGTTLVEQIIASHRQVYGAGELGIL
A0A2P9AJH6(3)   PEYYNNVGYSYLLRGKLQDARSNFLKAYELAPNDPTVANNLQLLSSSVKNLE
A0A2P9AMF0(1)   PYAHRAYGYLNSRLGNSEESIRWMHKAYELNPYDLSMAAAYAYGLIFVGRYREGTPMLAHAAETSSAHP
A0A2P9APE8(5)   QQLHHAAGKVLNDLKRYDEAFDHFNRAKQASGYKFDIYLYRQWIDSMIETFTPDFLAARAGYGNPSE
A0A2P9APL0(4)   AEGWMQLIRSYVVLGKADQARDALNRGIAVFGSDSEEAKKFTAFAVSIGLTATE
A0A2P9APL7(5)   HRMLAGIHACLGNKREAKEALAVFLKDSPGDSISKQRKQWEKNYTAPGALERWIGHMRFAGLPE
A0A2P9APZ8(6)   GSSHRVVIMALVKLNRLDEAQAAARRLMEAFPTYTLTLQRQINPWRDKVFAESYVEALGISGVPE
A0A2P9AQG7(3)   PDMLLKLGVSLVGLNQRDVACATFGEIGKRYPNVSGTLKERVKQERALAAC
A0A2P9AQK8(3)   FGALSGLAQIMAVTGHKQSALEAWQRVLGIYPMMRSAQDQVGTLSEELAGEGI
```

```
A0A2P9AT53(6)    TPTYRALACALAHLGRIDEAKGVGQRLRDLVPNLSADLERVLFRDSGKLPLILSGLRLAGLPA
A0A2P9AUX3(7)    AEAWSNQALIYERRGDKAKAAKSYREAVRLNPTYQPARDGLARTSTG
A0A2P9AXI4(4)    GSAYRLLAATYAQLGRLEDARRAASELLKLNPEFSIGRYSSRAPYRDKALLARYVEGLRLAGLPE
A0A2P9AXY3(2)    AEAFNDRGLAYQAMGENTKARADFDKAIQLKPGYAAALTNRAHLPPP
B3KST6(5)        SVLFNLTTMYELESSRSMQKKQALLEAVAGKEGDSFNTQCLKLA
B3KT23(2)        HMVLSNLAAVLMHRERYTQAKEIYQEALKQAKLKKDEISVQHIRGELAELSKKSRPLTNSVKL
B4DRC7(6)        WQIWENYILTSTDVGEFSEAIKAYHRLLDLRDKYKDVQVLKILVRAVIDGMTDRSGDVATGLKGKLQEL
B4DZ87(11)       DFAYETMGTIEVQRGNMEKAIDMFNKAINLAKSEMEMAHLYSLCDAAHAQTEVAKKYGLKPPTL
B7Z5B3(5)        FLGHLYAAEALISLDRISDAITHLNPENVTDVSLGISSNEQDQG (No TCH)
B7Z639(9)        PAALMNLGAILHLNGRLQKAEANYLRALQLKPDDVITQSNLRKLWNIMEKQGLKTSKT
E9PCN5(5)        TVMLFNLGSAYCLRSEYDKARKCLHQAQWLMPPSTLGGRGRRIMRSGV (No TCH)
F4MH46(10)       AAAWMDLGTLYESCNQPQDAIKCYLNAARSKRCSNTSTLAARIKFLQAQLCNLPQSSLQNKTKLLPSIE
Q8N9N0(2)        EQGLAEDGGMSSVTQEGRQGCSYLVASIRLWGSRLGRVMYSMANCLLLMKDYVLAVEAYHSVIKYYPEQ
```

Figure S2: Multiple alignment of TCH sequences shows lack of a signature motif. TCH sequences from experimentally determined as well as predicted structures are shown, following their alignment by ClustalW. The alignment was done to generate the output in order of similarity of the sequences as determined by ClustalW (as opposed to the input order), with the goal to maximize the possibility of finding a consensus if any; however, as shown, no consensus was found. Use of several other alignment programs (Kalign, MAFFT, MUSCLE, T-Coffee) also revealed no significant similarity (data not shown).

```
-------------------AKSRALDNIGRVFAR---------------------
-------------------WKTDALYSLGVLFYNNGADVLRKATPL---------
---------------------IHTLAQLISAYS----------------------
-------------LGPDDPNVAKTKNNLASCYLKQ--------------------
----------------------KEAADGYQRCMMAQYNR-----------------
----------------------KKAIDGL---------------------------
----------------------PARDGLARTS------------------------
--------------------HSSACDALGRDITAALFS-----------------
------------DAELMLADIIEARDGSEAAQVYITRQLQ---------------
--------------------QQAEARELLERL------------------------
ESFSLLQLIANDCYKMGQFYYSAKAFDVLER------------------------
--------------------------ADLERVLFR--------------------
-----------------HEALAARAGAMLEAVTA----------------------
-------RAAQEELGKVVIQGKNQDAGLAQGLRK----------------------
----------------------DVLAALQQKLLK----------------------
---------------------TDGFGCLLEALLAQW-------------------
-----------------ESRTHFRIQAATLQAK----------------------
-----------------ESRLHFRIQAALHL------------------------
--------------------FHRVGQAGLQLL------------------------
-------------------DAARVAQERLQSI------------------------
------------------PVVSRELQALEARIRQKDEEDKARFRGIF--------
-----------------------GLKRLKARIRR----------------------
----------------------KLLEMRAKADKLKRIEQRDVRKANLKEKKERNQ
---------------------KLLKENLAKLDRL----------------------
---------------------EKALVNRDRTLPL----------------------
----------------------KGLENLADTLHLSFKWDYEH-------------
-------------------VIENRMAVLNEQVKE----------------------
-----------KAVRRELRLLENRMAEKQEEERLRCRNM----------------
--------------------ANNQIKEAIDKRY----------------------
----------------------------ISKIIIF--------------------
------------------------RHSKIDKAMECVW------------------
-------------------NQAKELERLIDKAMKK---------------------
----------------------KIRKDILDKARA----------------------
---------------------PKIREEILEKARR----------------------
--------------KVFQEALRNIGGQIQEKVRYMSSTDAKVEQMFQIL------
```

```
---------------------HRALELTRRLLS----------------------
----------------------RALELNQRLS-----------------------
------------------ANVRRYLQLTQSELSSYHRKEKQLYLG----------
-------------------TVANNLQLLSS-------------------------
----------------LMFNVALVLQRLATSV-----------------------
--------------DAARELRSCLLHLTLQGQR----------------------
--------------------------LTLQRQI----------------------
----------------FVEHRYALLELSLLGQ-----------------------
------------------EEGKALLRQILQLR-----------------------
------------------IEAKMELEEVTRLLNLK---------------------
-------------------WALWGLREVFR-------------------------
------------------------AEFLTALCELRLSI-----------------
--------------PEPQQREQQLLEFLSRLTSLLES------------------
------------------QQIREGLEKAQRLL-----------------------
-------------------------EFAERLLLD---------------------
---------------------LEMLEIALRNLN----------------------
------------------EYIQRHLNIVRDKIS----------------------
-------------------EIQRLLLRVEEECRQMQQ------------------
--------------------TWYYAKRCFLSLLENMSKH----------------
----------------------YYAKRCFLSLLENMSK-----------------
----------------------PKLSLLMTNL-----------------------
----------------------PWLHECMIRLFNTA-------------------
-------------------NVLRYIQLTQLKMNRCSLQRED--------------
------------------TNVIRYIQLTEMKLSRCSQREKEAM------------
--------FDLALKMYCQEELFEEAAIAVEKYEEMLK------------------
------------------EEAKKFTAFAV--------------------------
--------------GTLKERVKQERALA---------------------------
----------------ATEAMKRDYESAKKKVE----------------------
----------------GSDAMKRGLETTKRKIEEANR------------------
--------------LTAKESLMSLFEETKRKAEQ----------------------
------------------------GTTLVEQII----------------------
-----------------ADAYSNMGNTLKEM-------------------------
-----------------LEWNSCVVQTLKEYLES----------------------
----------------------NGLDSALDQVQ----------------------
-------------------------RSAQDQVGTLSEE-----------------
------------------SREYGMLSEQIDR--------------------------
----------------ADSRESNFLKSLIDAVN----------------------
------------------HDINNELKKLASCYRDYVDKEKEMWHRMF--------
--------------------FNTQCLKLA---------------------------
-----------------------KLKIKAQSYLDAIQ------------------
----------------DAKMKYQECNKIVKQKAFERA------------------
---------------KDAKMKYQECNKIVK--------------------------
--------------------IQAELLKVKQKIKAQKDKEKAVYAK----------
------------------KAIQAELLKVKQKIKAQ---------------------
------------------KTIHAELSKLVKK-------------------------
-------------ASAMLKEVQPRAQKIAEHRRKYERKREEREIKE---------
------------------KIVQIDAEKIRNVIQGTELKS----------------
-------------------ETYKSNLKIAELKL----------------------
---------------------KDLVLKIAELLCKN---------------------
-----------DATLSLKQTILDKEEKQRRNVAKNY-------------------
-------------------------SKQRKQWEKN---------------------
------------ITQSNLRKLWNIMEKQGLKTSKT---------------------
-----------------TGTAGDAEQALRRSS-----------------------
------------------KTAKVAEKRLRKHI-------------------------
------------------KDAEDALQKLHKYMQKSLELREKQAEKEEKQKTK---
--------------LQRQVLQHLHTVQL---------------------------
----------------TQVKGYLNQVDLQEKADLQEK------------------
------------------APAKLQVQKILC-------------------------
---------------------DVQVLKILVRAVI----------------------
--------------PAKNRIQTIQCHLMLKKGRRSP-------------------
----------------------------VKNLRRS---------------------
----------------KASPEYLKVKYLREYSRYL---------------------
------------------DAWAFFYKFELQ-------------------------
------------------KQAKFFFNKWLQFEESE---------------------
```

```
------------EIWKEDLSWARTLQEQQKVAQRIKKSEAPAEVTH---------
----------------EEKLQAITFAEAAKVQQRI--------------------
---------------------TSTLAARIKFLQ----------------------
---------PFARRQLVLLNPYAALCNRMLADMMGQLRRPRDSR-----------
-----------------------AALTNRAHLPP----------------------
---------------------PAASTNLQKALSKA--------------------
-------------------WQWARLAACHAL------------------------
---------------------WARMAARGIQFVDLQLAR----------------
----------------------RARCEAQRDNLE---------------------
----ASRYDEYVNVKDFSDKINRALAAAKKDNDF----------------------
---------------------------YASVI------------------------
------------------SWAGMALARASRIQDKL--------------------
-------------------DLRRDIAYNLSLIYQ---------------------
--------------------LSMAAAYAYGLIF----------------------
-----------------SGNIWAALRIALS-------------------------
------------------ENIWSTLRLALSML------------------------
-------------RLNFGDDIPSALRIAKKKRWNSIEE-----------------
------------------AVPFTIIPRVL--------------------------
---------------------FSIIPREL--------------------------
------------AAKTQLAVCQQRIRRQLAREKKLYANMFE--------------
-------------------ISVQHIRGELAELSK---------------------
----------------------LDIRVSLAEVYYR--------------------
-------------------RFYFHLGDAMQ-------------------------
-------------ETLYYVKVYLVLGDIIFY------------------------
-------------------FSVTMLGHCIEMY------------------------
---------------------YEVVRPLVRL------------------------
------------------CVLVSTVGKLCRL------------------------
-----------------FDLQTSSYQALQRVLVSLGHH-----------------
-------------------PAQKLRQEVKQNLH----------------------
-------------IGRYSSRAPYRDKALLARYVEGLRL-----------------
----------------EMAHLYSLCDAAHAQTEVAKKYGL---------------
----------------EFGSVNGDNKPIWMHAEEREE------------------
-----------KAARLQIFMCQKKAKEHNERDRRTYANMFKKF------------
------------------SGTKENYGLLRRK------------------------
-----------------RTLATQQAMGMLSTAPKVA-------------------
------------------TVTAQRVDSVARVL----------------------
-----------------------VLHRILGDFLV----------------------
-------------------SRHSSSGDTFSLK-----------------------
------------------VDTQHLIKQLRQH------------------------
---------------------YLYRQWIDSMIE----------------------
------------------HQRANGNLKYFEYIMAK---------------------
--------------------LGRVMYSMANCLL----------------------
```

Figure S3: Leucine-centered 11-aa region of TCH sequences of known structure. The Leu residue (approximately L7 in the full-length TCH sequence) is highlighted (as detailed in Results).

```
AKSRALDNIGR
WKTDALYSLGV
VVSRELQALEA
NQAKELERLID
NVRRYLQLTQS
TVANNLQLLSS
NVALVLQRLAT
LRSCLLHLTLQ
HRYALLELSLL
EGKALLRQILQ
EAKMELEEVTR
```

WALWGLREVFR
REQQLLEFLSR
QIREGLEKAQR
QIREGLEKAQR
YIQRHLNIVRD
EIQRLLLRVEE
AMKRGLETTKR
REYGMLSEQID
RESNFLKSLID
DINNELKKLAS
AIQAELLKVKQ
TIHAELSKLVK
TIHAELSKLVK
DAEDALQKLHK
SWARTLQEQQK
NPYAALCNRML
WQWARLAACHA
KINRALAAAKK
WAGMALARASR
NIWAALRIALS
NIWSTLRLALS
DIPSALRIAKK
DIPSALRIAKK
RFYFHLGDAMQ
KVYLVLGDIIF
FSVTMLGHCIE
FSVTMLGHCIE