Protein sub-family analysis

Geoff Barton
Division of Computational Biology
School of Life Sciences
University of Dundee, UK

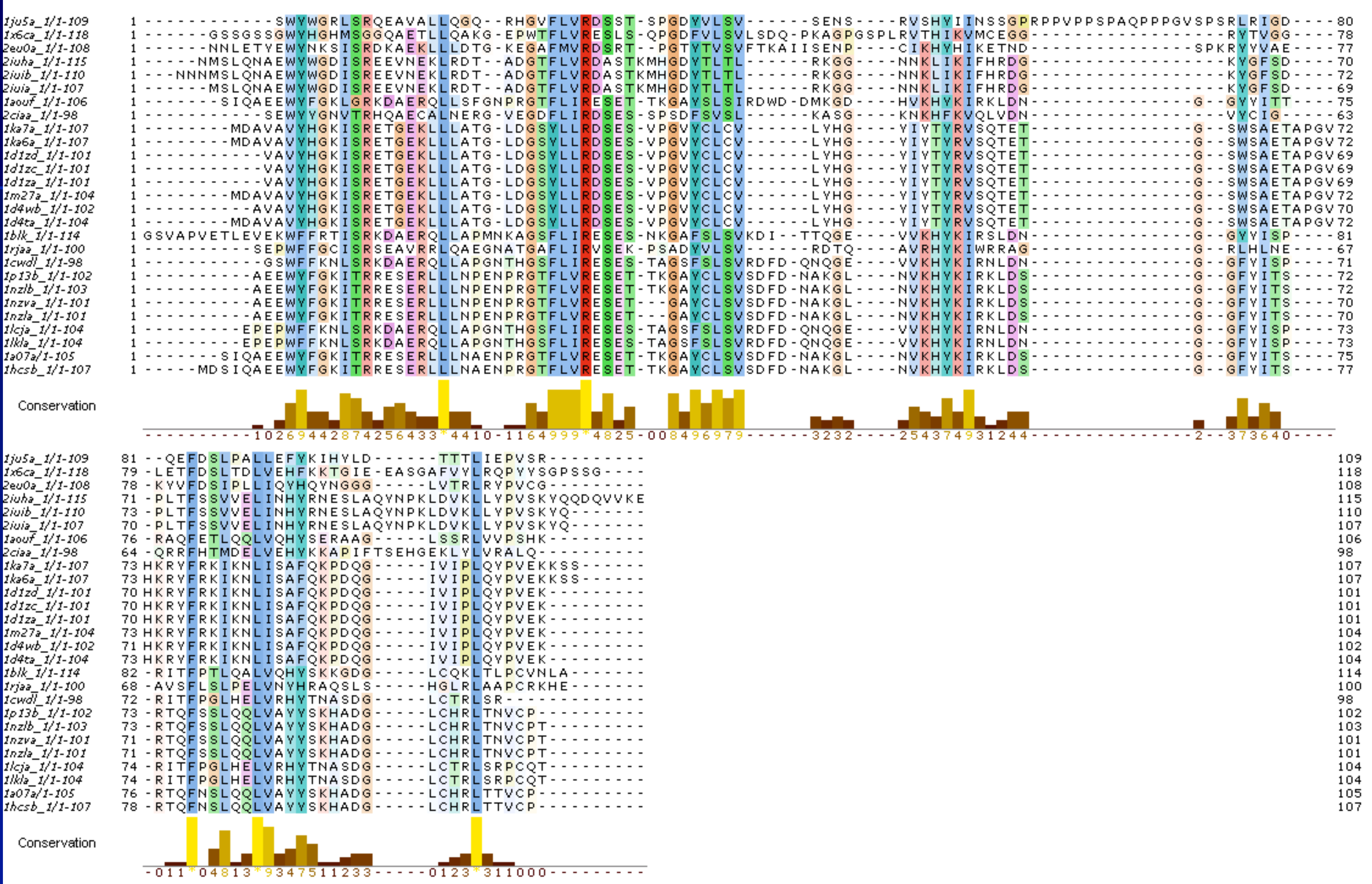
twitter:@gjbarton
blog: geoffbarton.wordpress.com

www.compbio.dundee.ac.uk
www.jalview.org

Identification of functional sites

- Whole alignment methods
  - Simple visualisation
  - Calculation of “conservation values”
- Sub-family analysis
  - AMAS analysis
  - Tree determinant positions
  - “Evolutionary trace”

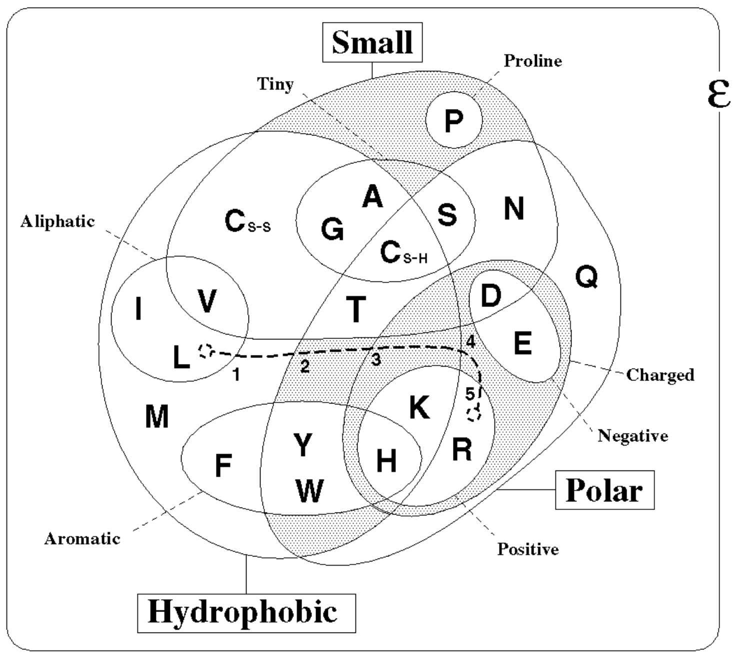
Example Multiple Sequence alignment of 27 SH2 domains



```
1ju5a_1/1-109   1 ------------SWYWGRLSRQEAVALLQGQ--RHGVFLVRDSST-SPGDYVLSV------SENS----RVSHYIINSSGPRPPVPPSPAQPPPGVSPSRLRIGD----80
1x6ca_1/1-118   1 ------GSSGSSGWYHGHMSGGQAETLLQAKG-EPWTFLVRESLS-QPGDFVLSVLSDQ-PKAGPGSPLRVTHIKVMCEGG------------------RYTVGG----78
2eu0a_1/1-108   1 ------NNLETYEWYNKSISRDKAEKLLLDTG-KEGAFMVRDSRT--PGTYTVSVFTKAIISENP----CIKHYHIKETND---------------SPKRYYVAE----77
2iuha_1/1-115   1 -----NMSLQNAEWYWGDISREEVNEKLRDT--ADGTFLVRDASTKMHGDYTLTL------RKGG----NNKLIKIFHRDG------------------KYGFSD----70
2iuib_1/1-110   1 ---NNNMSLQNAEWYWGDISREEVNEKLRDT--ADGTFLVRDASTKMHGDYTLTL------RKGG----NNKLIKIFHRDG------------------KYGFSD----72
2iuia_1/1-107   1 ------MSLQNAEWYWGDISREEVNEKLRDT--ADGTFLVRDASTKMHGDYTLTL------RKGG----NNKLIKIFHRDG------------------KYGFSD----69
1aouf_1/1-106   1 -------SIQAEEWYFGKLGRKDAERQLLSFGNPRGTFLIRESET-TKGAYSLSIRDWD-DMKGD----HVKHYKIRKLDN---------------G--GYYITT----75
2ciaa_1/1-98    1 -----------SEWYYGNVTRHQAECALNERG-VEGDFLIRDSES-SPSDFSVSL------KASG----KNKHFKVQLVDN------------------VYCIG-----63
1ka7a_1/1-107   1 --------MDAVAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV72
1ka6a_1/1-107   1 --------MDAVAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV72
1d1zd_1/1-101   1 -----------VAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV69
1d1zc_1/1-101   1 -----------VAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV69
1d1za_1/1-101   1 -----------VAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV69
1m27a_1/1-104   1 --------MDAVAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV72
1d4wb_1/1-102   1 ----------AVAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV70
1d4ta_1/1-104   1 --------MDAVAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV72
1blk_1/1-114    1 GSVAPVETLEVEKWFFRTISRKDAERQLLAPMNKAGSFLIRESES-NKGAFSLSVKDI--TTQGE----VVKHYKIRSLDN---------------G--GYYISP----81
1rjaa_1/1-100   1 ----------SEPWFFGCISRSEAVRRLQAEGNATGAFLIRVSEK-PSADYVLSV------RDTQ----AVRHYKIWRRAG---------------G--RLHLNE----67
1cwdl_1/1-98    1 -----------GSWFFKNLSRKDAERQLLAPGNTHGSFLIRESES-TAGSFSLSVRDFD-QNQGE----VVKHYKIRNLDN---------------G--GFYISP----71
1p13b_1/1-102   1 ----------AEEWYFGKITRRESERLLLNPENPRGTFLVRESET-TKGAYCLSVSDFD-NAKGL----NVKHYKIRKLDS---------------G--GFYITS----72
1nzlb_1/1-103   1 ----------AEEWYFGKITRRESERLLLNPENPRGTFLVRESET-TKGAYCLSVSDFD-NAKGL----NVKHYKIRKLDS---------------G--GFYITS----72
1nzva_1/1-101   1 ----------AEEWYFGKITRRESERLLLNPENPRGTFLVRESET---GAYCLSVSDFD-NAKGL----NVKHYKIRKLDS---------------G--GFYITS----70
1nzla_1/1-101   1 ----------AEEWYFGKITRRESERLLLNPENPRGTFLVRESET---GAYCLSVSDFD-NAKGL----NVKHYKIRKLDS---------------G--GFYITS----70
1lcja_1/1-104   1 ---------EPEPWFFKNLSRKDAERQLLAPGNTHGSFLIRESES-TAGSFSLSVRDFD-QNQGE----VVKHYKIRNLDN---------------G--GFYISP----73
1lkla_1/1-104   1 ---------EPEPWFFKNLSRKDAERQLLAPGNTHGSFLIRESES-TAGSFSLSVRDFD-QNQGE----VVKHYKIRNLDN---------------G--GFYISP----73
1a07a/1-105     1 -------SIQAEEWYFGKITRRESERLLLNAENPRGTFLVRESET-TKGAYCLSVSDFD-NAKGL----NVKHYKIRKLDS---------------G--GFYITS----75
1hcsb_1/1-107   1 -----MDSIQAEEWYFGKITRRESERLLLNAENPRGTFLVRESET-TKGAYCLSVSDFD-NAKGL----NVKHYKIRKLDS---------------G--GFYITS----77

Conservation
                  ----------10269442874256433*4410-1164999*4825-008496979------3232----254374931244---------------2--373640----

1ju5a_1/1-109  81 --QEFDSLPALLEFYKIHYLD------TTTLIEPVSR---------   109
1x6ca_1/1-118  79 -LETFDSLTDLVEHFKKTGIE-EASGAFVYLRQPYYSGPSSG----   118
2eu0a_1/1-108  78 -KYVFDSIPLLIQYHQYNGGG-----LVTRLRYPVCG---------   108
2iuha_1/1-115  71 -PLTFSSVVELINHYRNESLAQYNPKLDVKLLYPVSKYQQDQVVKE   115
2iuib_1/1-110  73 -PLTFSSVVELINHYRNESLAQYNPKLDVKLLYPVSKYQ-------   110
2iuia_1/1-107  70 -PLTFSSVVELINHYRNESLAQYNPKLDVKLLYPVSKYQ-------   107
1aouf_1/1-106  76 -RAQFETLQQLVQHYSERAAG-----LSSRLVVPSHK---------   106
2ciaa_1/1-98   64 -QRRFHTMDELVEHYKKAPIFTSEHGEKLYLVRALQ----------   98
1ka7a_1/1-107  73 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEKKSS------   107
1ka6a_1/1-107  73 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEKKSS------   107
1d1zd_1/1-101  70 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEK---------   101
1d1zc_1/1-101  70 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEK---------   101
1d1za_1/1-101  70 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEK---------   101
1m27a_1/1-104  73 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEK---------   104
1d4wb_1/1-102  71 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEK---------   102
1d4ta_1/1-104  73 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEK---------   104
1blk_1/1-114   82 -RITFPTLQALVQHYSKKGDG-----LCQKLTLPCVNLA-------   114
1rjaa_1/1-100  68 -AVSFLSLPELVNYHRAQSLS-----HGLRLAAPCRKHE-------   100
1cwdl_1/1-98   72 -RITFPGLHELVRHYTNASDG-----LCTRLSR-------------   98
1p13b_1/1-102  73 -RTQFSSLQQLVAYYSKHADG-----LCHRLTNVCP----------   102
1nzlb_1/1-103  73 -RTQFSSLQQLVAYYSKHADG-----LCHRLTNVCPT---------   103
1nzva_1/1-101  71 -RTQFSSLQQLVAYYSKHADG-----LCHRLTNVCPT---------   101
1nzla_1/1-101  71 -RTQFSSLQQLVAYYSKHADG-----LCHRLTNVCPT---------   101
1lcja_1/1-104  74 -RITFPGLHELVRHYTNASDG-----LCTRLSRPCQT---------   104
1lkla_1/1-104  74 -RITFPGLHELVRHYTNASDG-----LCTRLSRPCQT---------   104
1a07a/1-105    76 -RTQFNSLQQLVAYYSKHADG-----LCHRLTTVCP----------   105
1hcsb_1/1-107  78 -RTQFNSLQQLVAYYSKHADG-----LCHRLTTVCP----------   107

Conservation
                  -011*04813*9347511233-----0123*311000---------
```

AMAS method of calculating conservation

Taylor Venn diagram of amino acid properties

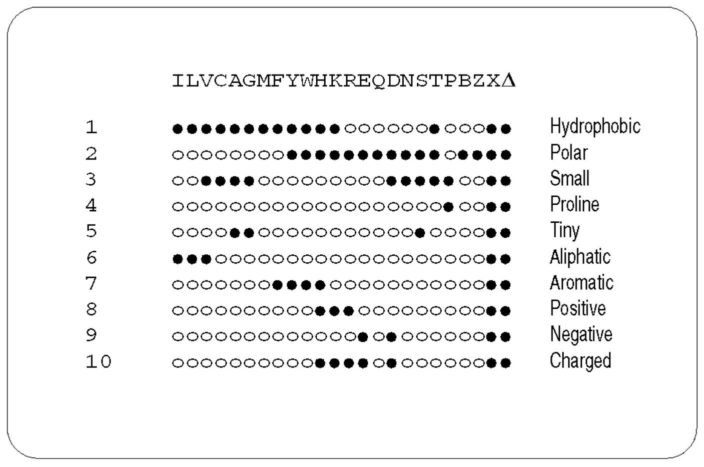


Count maximum number of set boundaries that must be crossed to include all amino acids at an alignment position

This gives a measure of the physico-chemical property variability at the alignment position

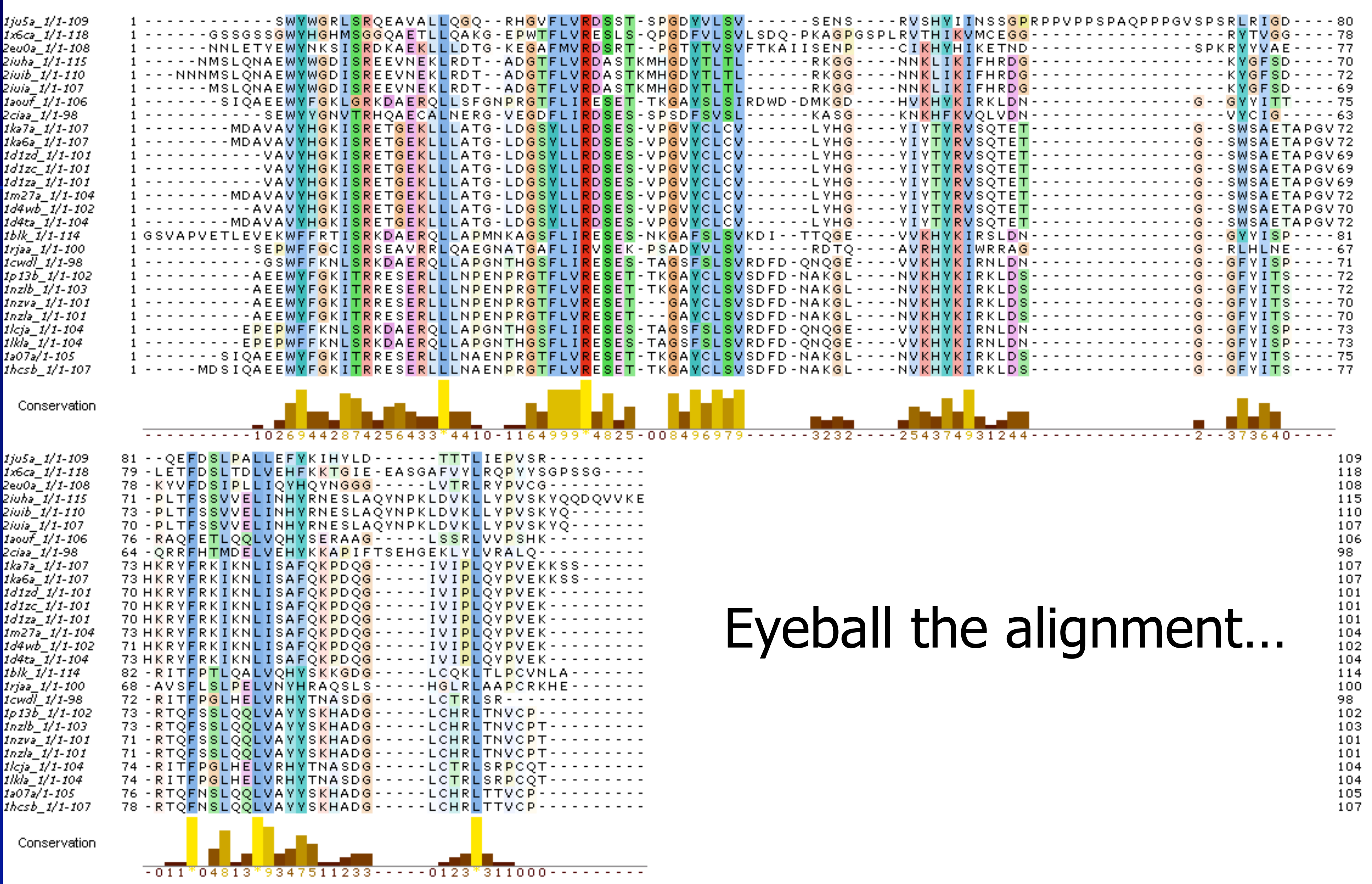
For comprehensive review of methods see:

Valdar, WS (2002) Proteins, 48, 227-41



	I	L	V	C	A	G	M	F	Y	W	H	K	R	E	Q	D	N	S	T	P	B	Z	X	Δ	
1	●	●	●	●	●	●	●	●	●	●	●	●	○	○	○	○	○	○	●	○	○	○	●	●	Hydrophobic
2	○	○	○	○	○	○	○	○	●	●	●	●	●	●	●	●	●	●	●	○	●	●	●	●	Polar
3	○	○	●	●	●	●	○	○	○	○	○	○	○	○	○	●	●	●	●	●	○	○	●	●	Small
4	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	●	●	Proline
5	○	○	○	○	●	●	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	●	●	Tiny
6	●	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	●	Aliphatic
7	○	○	○	○	○	○	○	●	●	●	●	○	○	○	○	○	○	○	○	○	○	○	●	●	Aromatic
8	○	○	○	○	○	○	○	○	○	○	●	●	●	○	○	○	○	○	○	○	○	○	●	●	Positive
9	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	●	○	○	○	○	○	○	●	●	Negative
10	○	○	○	○	○	○	○	○	○	○	●	●	●	●	○	●	○	○	○	○	○	○	●	●	Charged

Example Multiple Sequence alignment of 27 SH2 domains



```
1ju5a_1/1-109   1 ------------SWYWGRLSRQEAVALLQGQ--RHGVFLVRDSST-SPGDYVLSV------SENS----RVSHYIINSSGPRPPVPPSPAQPPPGVSPSRLRIGD---- 80
1x6ca_1/1-118   1 ------GSSGSSGWYHGHMSGGQAETLLQAKG-EPWTFLVRESLS-QPGDFVLSVLSDQ-PKAGPGSPLRVTHIKVMCEGG------------------RYTVGG---- 78
2eu0a_1/1-108   1 ------NNLETYEWYNKSISRDKAEKLLLDTG-KEGAFMVRDSRT--PGTYTVSVFTKAIISENP----CIKHYHIKETND---------------SPKRYYVAE---- 77
2iuha_1/1-115   1 -----NMSLQNAEWYWGDISREEVNEKLRDT--ADGTFLVRDASTKMHGDYTLTL------RKGG----NNKLIKIFHRDG------------------KYGFSD---- 70
2iuib_1/1-110   1 ---NNNMSLQNAEWYWGDISREEVNEKLRDT--ADGTFLVRDASTKMHGDYTLTL------RKGG----NNKLIKIFHRDG------------------KYGFSD---- 72
2iuia_1/1-107   1 ------MSLQNAEWYWGDISREEVNEKLRDT--ADGTFLVRDASTKMHGDYTLTL------RKGG----NNKLIKIFHRDG------------------KYGFSD---- 69
1aouf_1/1-106   1 -------SIQAEEWYFGKLGRKDAERQLLSFGNPRGTFLIRESET-TKGAYSLSIRDWD-DMKGD----HVKHYKIRKLDN---------------G--GYYITT---- 75
2ciaa_1/1-98    1 -----------SEWYYGNVTRHQAECALNERG-VEGDFLIRDSES-SPSDFSVSL------KASG----KNKHFKVQLVDN------------------VYCIG----- 63
1ka7a_1/1-107   1 --------MDAVAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV 72
1ka6a_1/1-107   1 --------MDAVAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV 72
1d1zd_1/1-101   1 -----------VAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV 69
1d1zc_1/1-101   1 -----------VAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV 69
1d1za_1/1-101   1 -----------VAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV 69
1m27a_1/1-104   1 --------MDAVAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV 72
1d4wb_1/1-102   1 ----------AVAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV 70
1d4ta_1/1-104   1 --------MDAVAVYHGKISRETGEKLLLATG-LDGSYLLRDSES-VPGVYCLCV------LYHG----YIYTYRVSQTET---------------G--SWSAETAPGV 72
1blk_1/1-114    1 GSVAPVETLEVEKWFFRTISRKDAERQLLAPMNKAGSFLIRESES-NKGAFSLSVKDI--TTQGE----VVKHYKIRSLDN---------------G--GYYISP---- 81
1rjaa_1/1-100   1 ----------SEPWFFGCISRSEAVRRLQAEGNATGAFLIRVSEK-PSADYVLSV------RDTQ----AVRHYKIWRRAG---------------G--RLHLNE---- 67
1cwdl_1/1-98    1 -----------GSWFFKNLSRKDAERQLLAPGNTHGSFLIRESES-TAGSFSLSVRDFD-QNQGE----VVKHYKIRNLDN---------------G--GFYISP---- 71
1p13b_1/1-102   1 ----------AEEWYFGKITRRESERLLLNPENPRGTFLVRESET-TKGAYCLSVSDFD-NAKGL----NVKHYKIRKLDS---------------G--GFYITS---- 72
1nzlb_1/1-103   1 ----------AEEWYFGKITRRESERLLLNPENPRGTFLVRESET-TKGAYCLSVSDFD-NAKGL----NVKHYKIRKLDS---------------G--GFYITS---- 72
1nzva_1/1-101   1 ----------AEEWYFGKITRRESERLLLNPENPRGTFLVRESET---GAYCLSVSDFD-NAKGL----NVKHYKIRKLDS---------------G--GFYITS---- 70
1nzla_1/1-101   1 ----------AEEWYFGKITRRESERLLLNPENPRGTFLVRESET---GAYCLSVSDFD-NAKGL----NVKHYKIRKLDS---------------G--GFYITS---- 70
1lcja_1/1-104   1 ---------EPEPWFFKNLSRKDAERQLLAPGNTHGSFLIRESES-TAGSFSLSVRDFD-QNQGE----VVKHYKIRNLDN---------------G--GFYISP---- 73
1lkla_1/1-104   1 ---------EPEPWFFKNLSRKDAERQLLAPGNTHGSFLIRESES-TAGSFSLSVRDFD-QNQGE----VVKHYKIRNLDN---------------G--GFYISP---- 73
1a07a/1-105     1 -------SIQAEEWYFGKITRRESERLLLNAENPRGTFLVRESET-TKGAYCLSVSDFD-NAKGL----NVKHYKIRKLDS---------------G--GFYITS---- 75
1hcsb_1/1-107   1 -----MDSIQAEEWYFGKITRRESERLLLNAENPRGTFLVRESET-TKGAYCLSVSDFD-NAKGL----NVKHYKIRKLDS---------------G--GFYITS---- 77

Conservation
                  ----------10269442874256433*4410-1164999*4825-008496979------3232----254374931244---------------2--373640----

1ju5a_1/1-109  81 --QEFDSLPALLEFYKIHYLD------TTTLIEPVSR--------- 109
1x6ca_1/1-118  79 -LETFDSLTDLVEHFKKTGIE-EASGAFVYLRQPYYSGPSSG---- 118
2eu0a_1/1-108  78 -KYVFDSIPLLIQYHQYNGGG-----LVTRLRYPVCG--------- 108
2iuha_1/1-115  71 -PLTFSSVVELINHYRNESLAQYNPKLDVKLLYPVSKYQQDQVVKE 115
2iuib_1/1-110  73 -PLTFSSVVELINHYRNESLAQYNPKLDVKLLYPVSKYQ------- 110
2iuia_1/1-107  70 -PLTFSSVVELINHYRNESLAQYNPKLDVKLLYPVSKYQ------- 107
1aouf_1/1-106  76 -RAQFETLQQLVQHYSERAAG-----LSSRLVVPSHK--------- 106
2ciaa_1/1-98   64 -QRRFHTMDELVEHYKKAPIFTSEHGEKLYLVRALQ---------- 98
1ka7a_1/1-107  73 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEKKSS------ 107
1ka6a_1/1-107  73 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEKKSS------ 107
1d1zd_1/1-101  70 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEK--------- 101
1d1zc_1/1-101  70 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEK--------- 101
1d1za_1/1-101  70 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEK--------- 101
1m27a_1/1-104  73 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEK--------- 104
1d4wb_1/1-102  71 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEK--------- 102
1d4ta_1/1-104  73 HKRYFRKIKNLISAFQKPDQG-----IVIPLQYPVEK--------- 104
1blk_1/1-114   82 -RITFPTLQALVQHYSKKGDG-----LCQKLTLPCVNLA------- 114
1rjaa_1/1-100  68 -AVSFLSLPELVNYHRAQSLS-----HGLRLAAPCRKHE------- 100
1cwdl_1/1-98   72 -RITFPGLHELVRHYTNASDG-----LCTRLSR------------- 98
1p13b_1/1-102  73 -RTQFSSLQQLVAYYSKHADG-----LCHRLTNVCP---------- 102
1nzlb_1/1-103  73 -RTQFSSLQQLVAYYSKHADG-----LCHRLTNVCPT--------- 103
1nzva_1/1-101  71 -RTQFSSLQQLVAYYSKHADG-----LCHRLTNVCPT--------- 101
1nzla_1/1-101  71 -RTQFSSLQQLVAYYSKHADG-----LCHRLTNVCPT--------- 101
1lcja_1/1-104  74 -RITFPGLHELVRHYTNASDG-----LCTRLSRPCQT--------- 104
1lkla_1/1-104  74 -RITFPGLHELVRHYTNASDG-----LCTRLSRPCQT--------- 104
1a07a/1-105    76 -RTQFNSLQQLVAYYSKHADG-----LCHRLTTVCP---------- 105
1hcsb_1/1-107  78 -RTQFNSLQQLVAYYSKHADG-----LCHRLTTVCP---------- 107

Conservation
                  -011*04813*93475112 33-----0123*311000---------
```

Eyeball the alignment…

Identification of functional sites

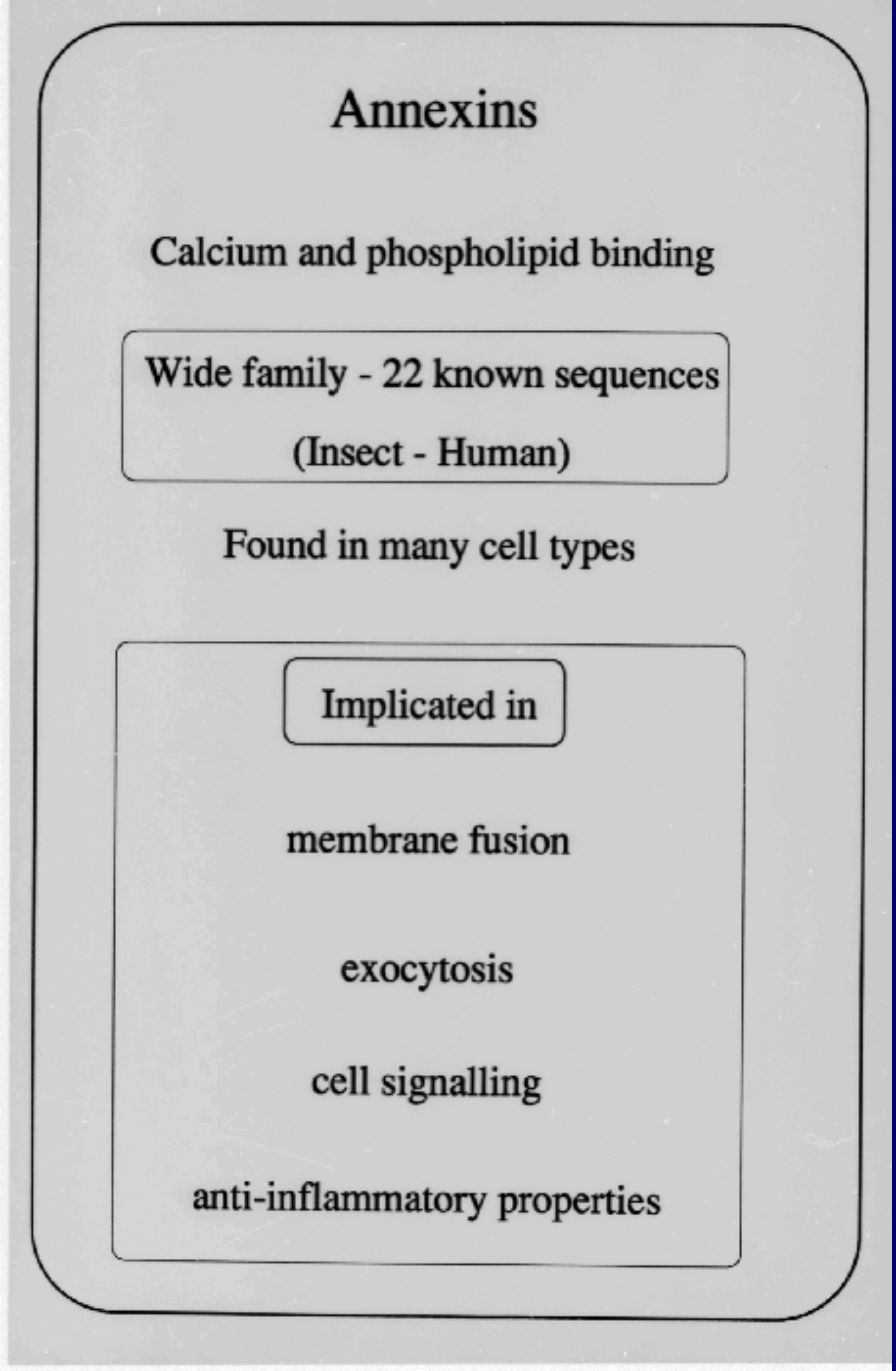
- Whole alignment methods
  - Simple visualisation
  - Calculation of “conservation values”

- Sub-family analysis
  - AMAS analysis
  - “Tree determinant” positions
  - “Evolutionary trace”

Sequence analysis of the Annexins: An example of sub-family analysis

- “Large” number of sequences (for 1990)
- Possess multiple domains
- Unknown tertiary structure at the time of analysis

- Barton, G. J., Freemont, P. F., Newman, R. & Crumpton, M. (1991), "Sequence Analysis of the Annexin Super Gene Family of Proteins” *Eur. J. Biochem*, **198,** 749-760.



Annexins

Calcium and phospholipid binding

Wide family - 22 known sequences

(Insect - Human)

Found in many cell types

Implicated in

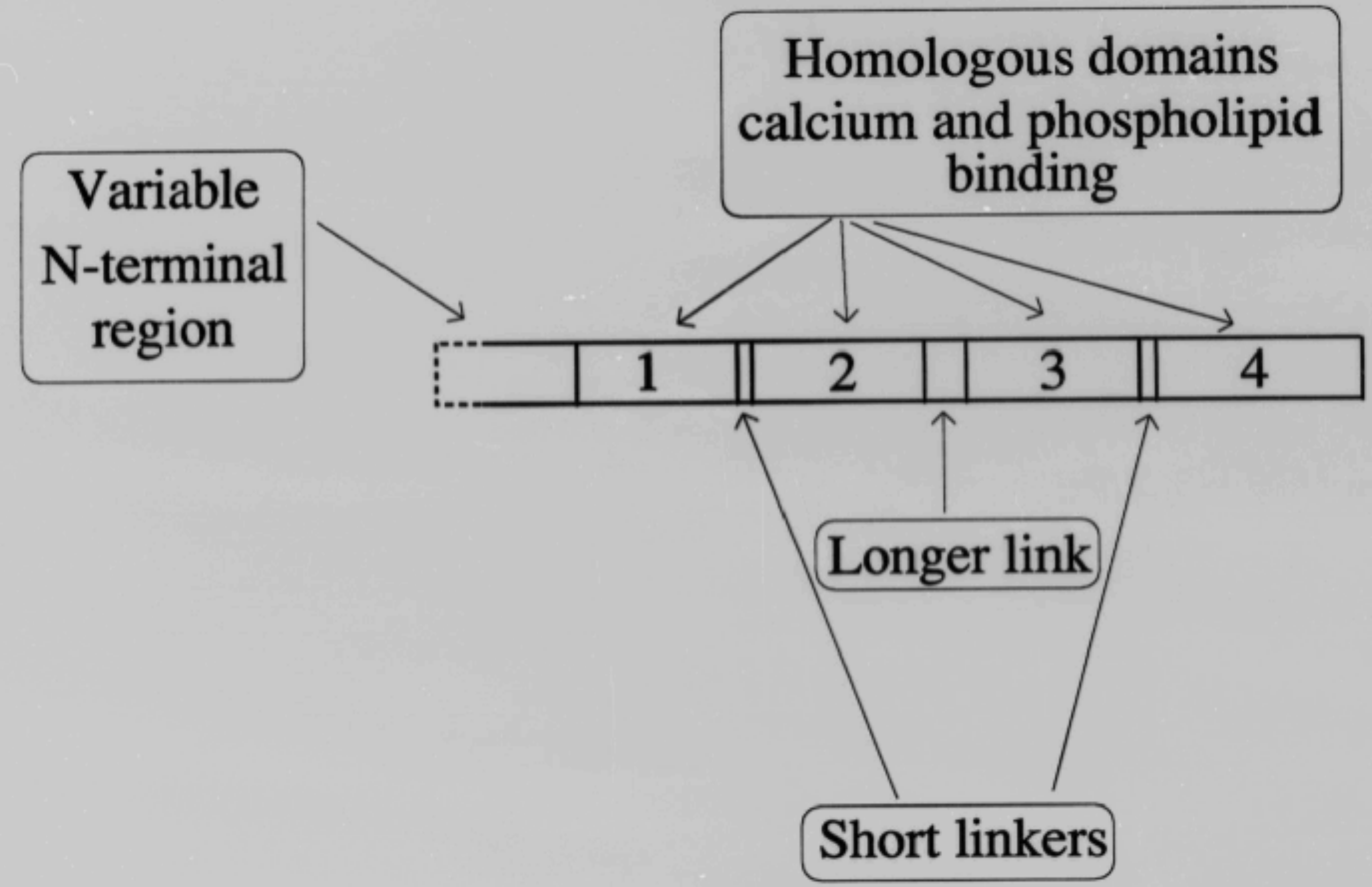
membrane fusion

exocytosis

cell signalling

anti-inflammatory properties

Annexins



Homologous domains calcium and phospholipid binding

Variable N-terminal region

Longer link

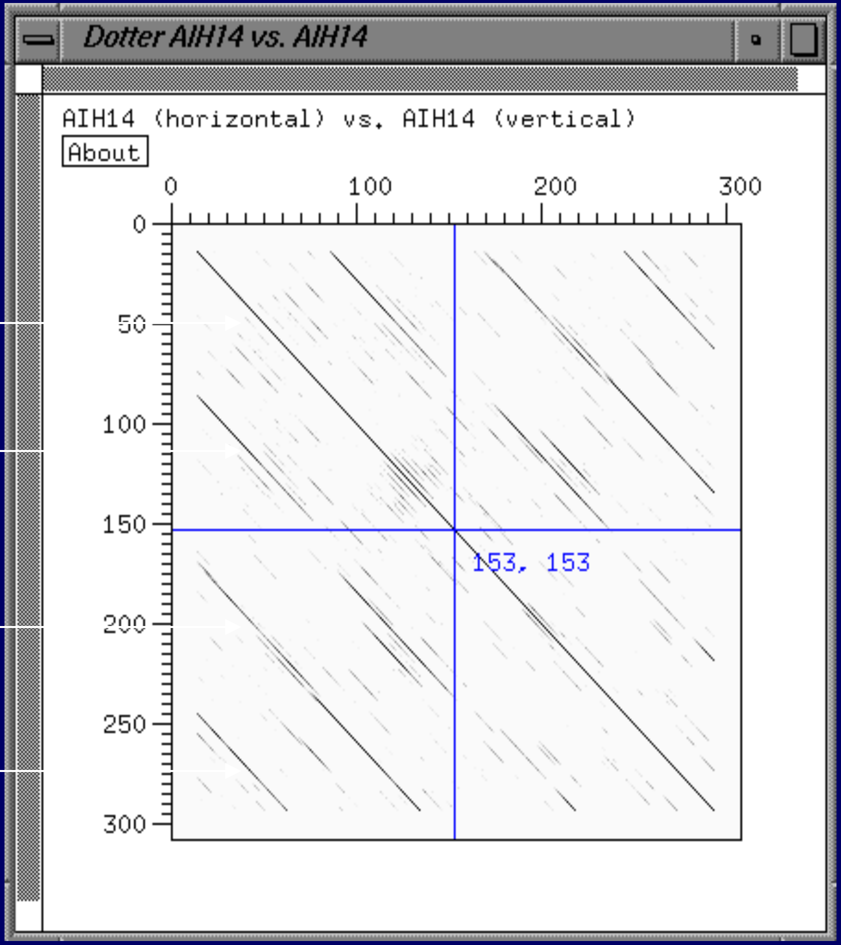
Short linkers

Annexin VI has 8 repeats

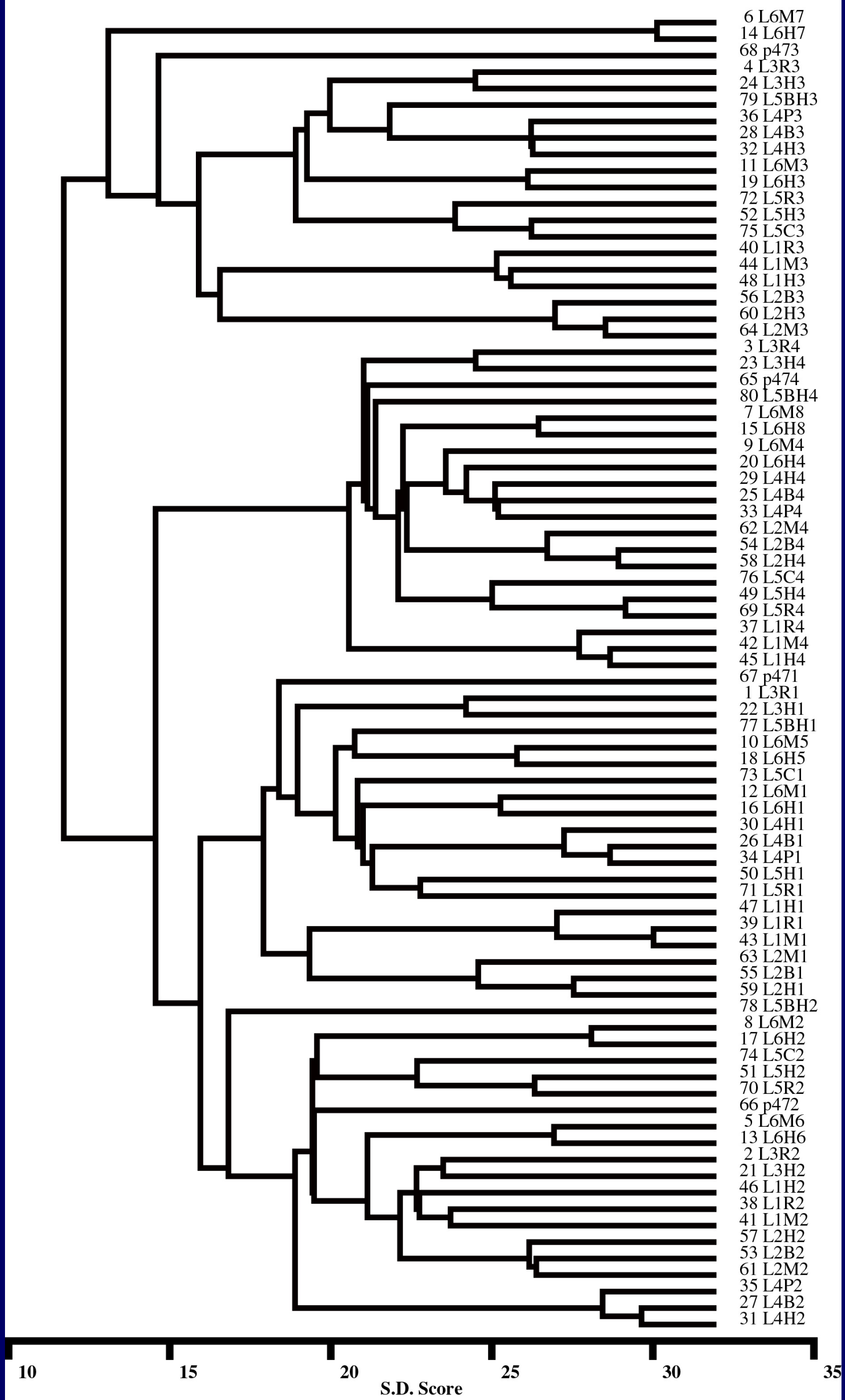
Sequence Analysis of Annexin Domains

Dot-Plot comparison of Human Annexin I with itself.

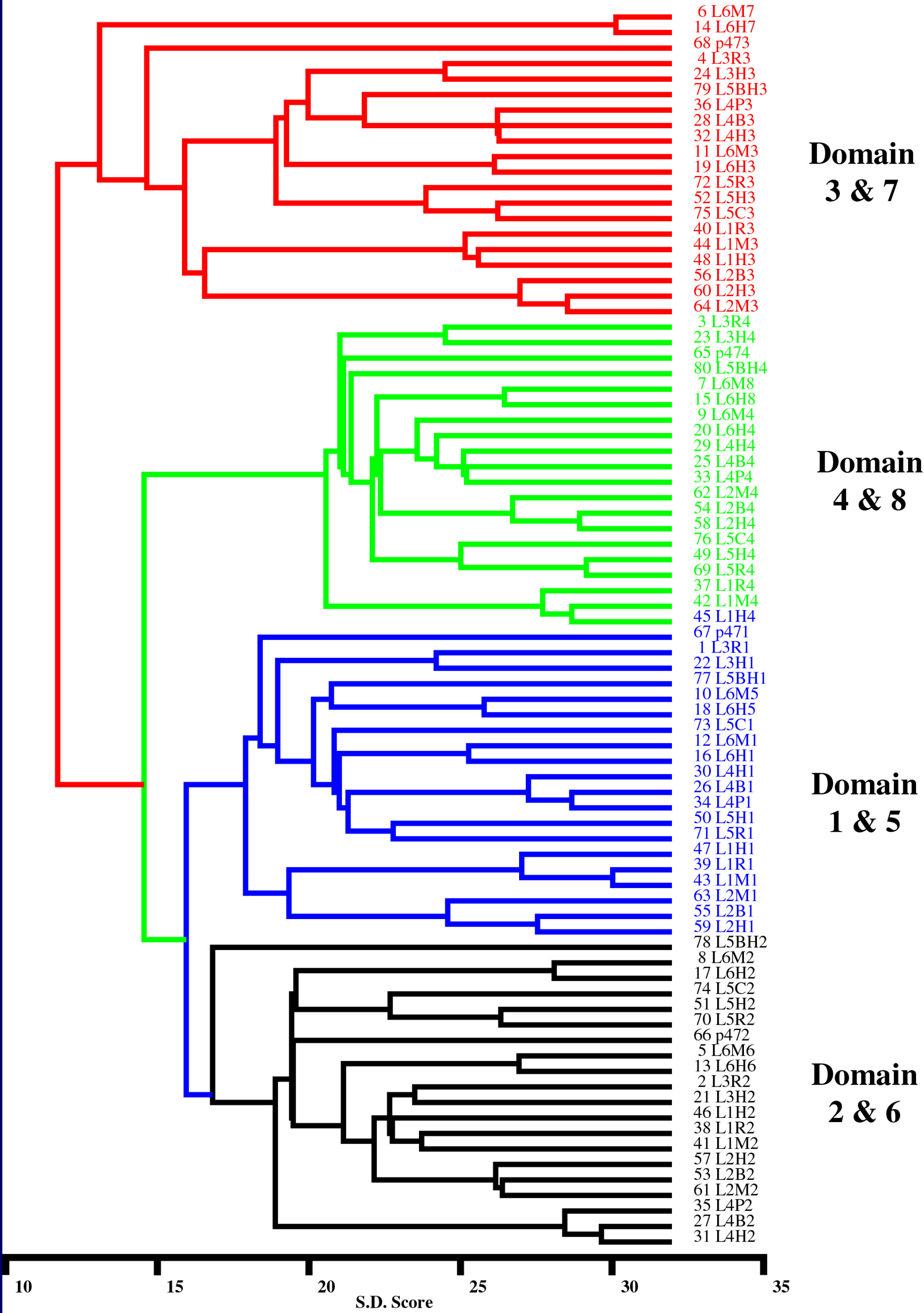
Four repeats (domains ?) are visible.



Annexin Domains



Annexin Domains

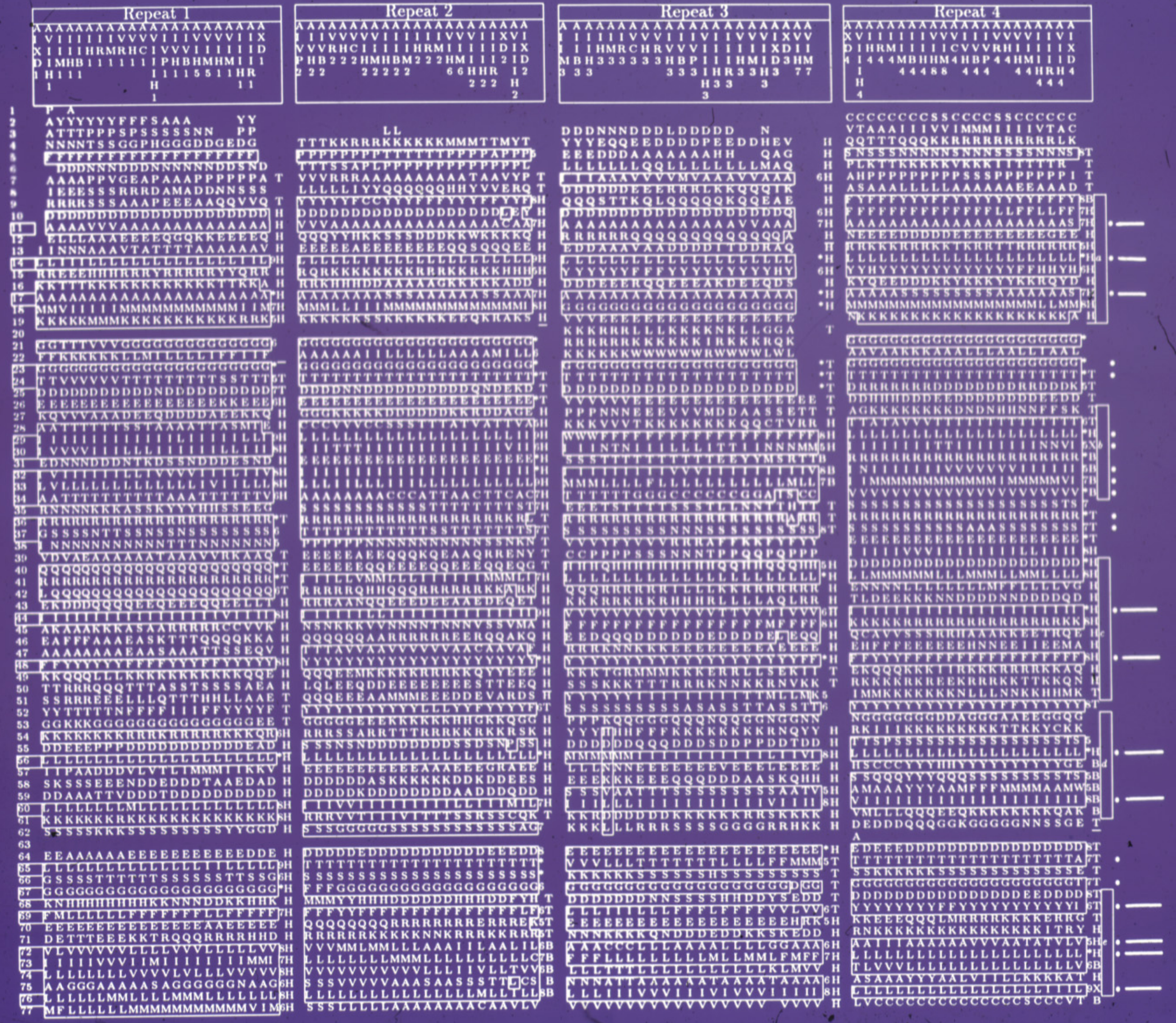


Domain
3 & 7

Domain
4 & 8

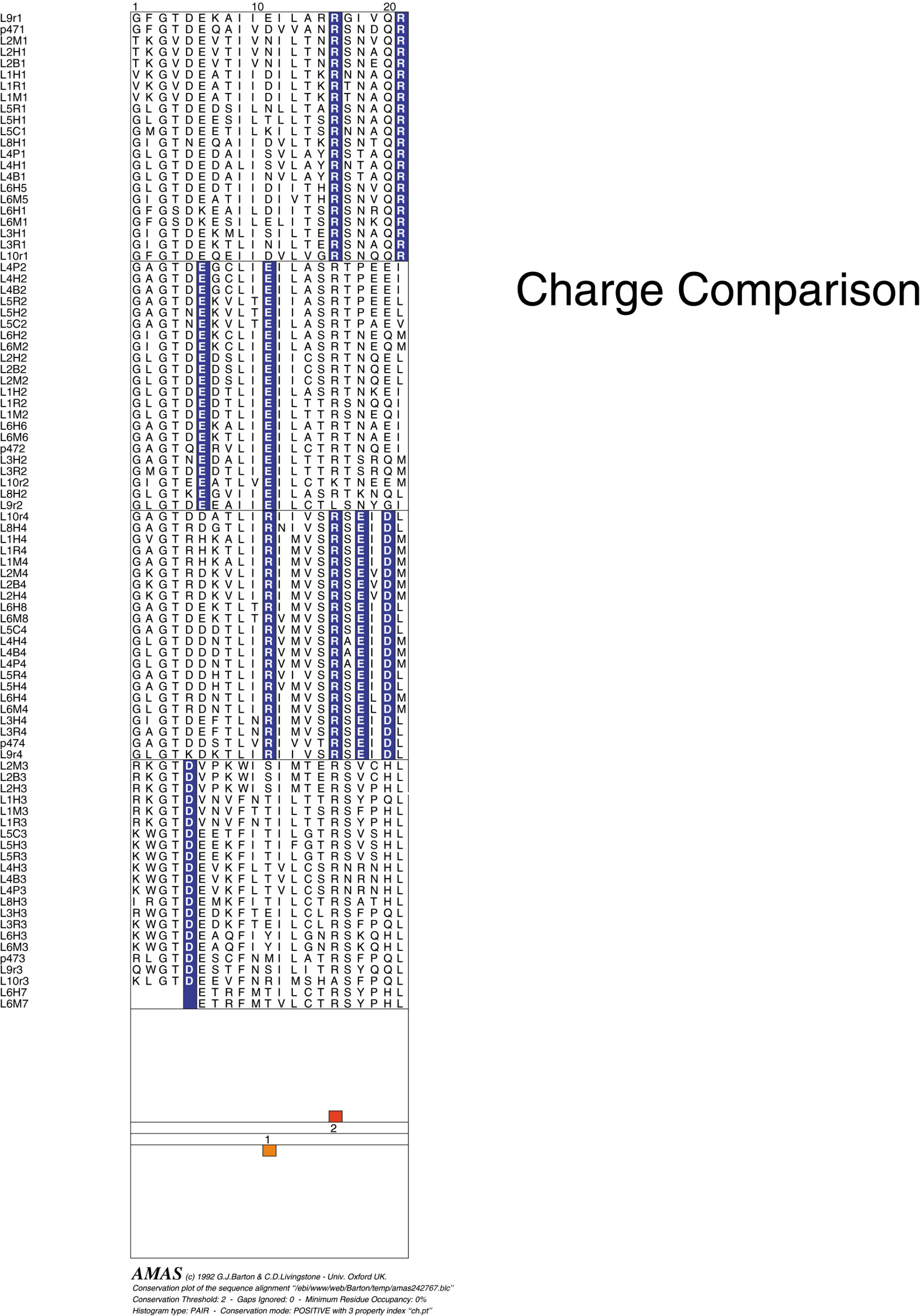
Domain
1 & 5

Domain
2 & 6



	Repeat 1		Repeat 2		Repeat 3		Repeat 4	
	AAAAAAAAAAAAAAAAAAAAAAA		AAAAAAAAAAAAAAAAAAAAAAA		AAAAAAAAAAAAAAAAAAAAAA		AAAAAAAAAAAAAAAAAAAAAA	
	IVIIIIIIVVVVIIIVVVVIIX		IIIVVVVVIIIIIIIVVVIIXVI		IIIIIIVVVIIIVIIVVVIXVV		XVIIIIIIVVVIIIVVVVIIVI	
	XIIIIHRMRHCIVVVIIIIIID		VVVRHCIIIIIIHRMIIIIIIDIX		IIIHMRCHRVVVIIIIIIIXDII		DIHRMIIIIICVVVRHIIIIIX	
	DIMHB111111IPHBHMHMII1		PHB222HMHBM222HMI112ID		MBH333333HBPIIIHMID3HM		4I444MBHHM4HBP44HMIIID	
	1H111 I11155 11HR		222 22222 66HHR I2		333 3331HR33H3 77		I 44488 444 44HRH4	
	1 H 11		222 H		H33 3		H 444	
	1		2		3		4	
1	P A							
2	AYYYYYYFFFSAAA YY						CCCCCCCCSSCCCCSSCCCCCC	
3	ATTTPPPSPSSSSSNN PP		LL		DDDNNNDDDLDDDDD N		VTAAAIIIVVIMMMIIIIVTAC	
4	NNNNTSSGGPHGGGDDGEDG		TTTKKRRRRKKKKKKMMMTTMYT		YYYEQQEEDDDDPEEDDHEV	H	QQTTTQQQKKRRRRRRRRRRLK	
5	FFFFFFFFFFFFFFFFFFFFF		PPPPPPPPPTTTTTTTPPPPPAPPP	6	EEEDDDAAAAAAAHH QAG	H	SNSSSNNNNNSNNNSSSSNNNS	8T
6	DDDNNNDDDNNNNNNNDDSND		TTTSSAPLPPPPPPPPPPPPPPL		LLLLLLLQQLLLLLLLLMAQ	H	PLKTTKKKKKVKKKIITTTTTR	T
7	AAAAPPVGEAPAAAPPPPPPA	T	VVVRRRAAAAAAAAAAATAAVYP	T	ITTAAAVVVVMVAAAVVAAA	6H	AHPPPPPPPPPSSSPPPPPPPI	T
8	IEEESSSRRRDAMADDNNSSS		LLLLLIYYQQQQQQHHYVVERQ	T	DDDDDDEEERRRLKKQQQIK	H	ASAAALLLLLAAAAAAAEEAAAD	T
9	RRRRSSSAAAPEEEAAQQVVQ	T	YYYYYFCCYYYYFFFYYYYFFYYF	8H	QQQSTTKQLQQQQQKQQEAE	H	FYFFFYYYFFYYYYYYYYYFFFY	8B
10	DDDDDDDDDDDDDDDDDDDDDD	H	DDDDDDDDDDDDDDDDDDDLEY	H	DDDDDDDDDDDDDDDDDDDQ	6H	FFFFFFFFFFFFFFLLFFLLFF	7H
11	AAAAVVVAAAAAAAAAAAAAA	H	VVVAAAAAAAAAAAAAAAAACAA	7H	AAAAAAAAAAAAAAAVVAAA	7H	AAAAAAAAAAAAAAAAAAAAAS	7H
12	ELLLAAAEEEEEQGGQKKEEEEEQ	H	QQQYYHKKSSSDDDKKWKKKKQ	H	RRRRRRQQQQQQQQQQQQQA	H	NEEEEDDDDDEEEEEEEEEEGEE	H
13	IINNNAAAVTATTTTTAAAAAAV	H	EEEEEAEEEEEEEEQQSQQQEE	H	EDDAAAVAADDDDITDDRAQ	H	RRKKKRRRKKTKRRTTRRRRRR	6H
14	LLIIIILLLLLLLLLLLLLIIL	9H	LLLLLLIILLLLLLLLLLLLLLL	9H	LLLLLLLLLLLLLLLLLLLL	*H	LLLLLLLLLLLLLLLLLLLLLL	*H a
15	RREEEEHHHHRRRRYRRRRRRYYQRR	H	RQRKKKKKKKKRRRKKRKKHHH	6H	YYYYYYFFFYYYYYYYYYHY	6H	YYHYYYYYYYYYYYYYFFHHYH	6H
16	KKTTTTKKKKKKKKKKKKKTTKKA	H	RRKHHHHDDAAAAAGKKKKKKADD	H	DDDEEERQQEEEAKDEEQDS	H	KYQEEDDDKKYKKKYYKKRQYD	H
17	AAAAAAAAAAAAAAAAAAAAAAA	*H	AAAAAAAAASSSAAAAAAASSAAA	7H	AAAAAAAAAAAAAAAAAAAA	*H	AAAAASSSSSSSSSAAAAAAAAS	7H
18	MMVIIIIIIMMMMMMMMMMMIIM	7H	MMMLLIIIIMMMMMMMMMMMMMMI	8H	GGGGGGGGGGGGGGGGGGGG	*H	MMMMMMMMMMMMMMMMMMMLLMM	8H
19	KKKKKMMMKKKKKKKKKKKKKRRK	6H	KKKKKKSSKKKKKKKKEEQKRAKS	H	VVVEEEEEEEEEEEEEEEEE		NKKKKKKKKKKKKKKKKKKKKKA	H
20					KKKRRRLLLKKKKNKLLGGA	T		
21	GGTTTVVVGGGGGGGGGGGGGGG	6	GGGGGGGGGGGGGGGGGGGGGGG	*	RRRRRRKKKKKKIRKKKKRQK		GGGGGGGGGGGGGGGGGGGGGG	*
22	FFKKKKKKLLMILLLLLIFFIIF		AAAAAAIILLLLLLAAAAAMILL	6	KKKKKKWWWWWWRWWWWWLWL		AAVAAKKKAAALLLAALLIAAL	
23	GGGGGGGGGGGGGGGGGGGGGGGG	*	GGGGGGGGGGGGGGGGGGGGGGG	*	GGGGGGGGGGGGGGGGGGGG	*T	GGGGGGGGGGGGGGGGGGGGGG	*
24	TTVVVVVVVTTTTTTTTTTSSTTT	5T	TTTTTTTTTTTTTTTTTTTTTTT	*T	TTTTTTTTTTTTTTTTTTTT	*T	TTTTTTTTTTTTTTTTTTTTTT	*T
25	DDDDDDDDDDDNDDDDDDDDDDDD	7T	DDDDNNDDDDDDDDDDDQNDEKD	T	DDDDDDDDDDDDDDDDDDDD	*T	DRRRRRRRDDDDDDDDRRDDDK	6T
26	EEEEEEEEEEEEEEEEEEEKKEEE	6H	EEEEEEEEEEEEEEEEEEEEEEE	*T	VVVVVVVEEEEEEEEEEEEEEEEE	T	DDHHHHDDDEEDDDDDDDDEEDD	T
27	KQVVVAAADEEQDDDDAEEKKQ	H	GGGKKKKKKDDDDDDDKKRDDAGE	H	PPPNNNEEEVVVMDDAASSETT	T	AGKKKKKKKKDNDNHHNNFFSK	T
28	AATTTTTTTSSTAAAATTASMTE	H	CCCVVVVCCSSSTTTATVATTVA	6H	KKKVVVTKKKKKKKKQQCTVRR	H	TTATAVVVTTTTTTTTTTTTTT	6T
29	IIIIIIIIIIIILIIIIIIILLI	9H	LLLLLLLLLLLLLLLLLLLLLII	9H	WWWFFFFFFFFFFFFFFFFFFFF	8H	LLLLLLLLLLLLLLLLLLLLLL	*H
30	IVVVVIIIILLLIIIIIIILLIII	8H	IIITTTIIIIIIIIIIIIIIIVII	6H	IIINTNIIILLLITTIINNNMM	5H	IIIIIIIIITTIIIIIIIINNVI	5X b
31	EDNNNDDDNTKDSSNDDDESND	H	EEEEEEEEEEEEEEEEEEEEEEE	*H	SSSTTTTTTTTTTTEEYYMSRTT	B	RRRRRRRRRRRRRRRRRRRRRR	*B
32	IVIIIIIIILLIVVVVIIIILIIV	8H	IIIIIIIIIIIIIIIIIIIIIIII	*H	IIIIIIIIIIVVVVIIIIIIIIIV	8B	INIIIIIIIVVVVVVVIIIIII	5B
33	LVLLLLLLLLLLLLLIVIILLL	8H	LLLIILLLIIIILLLLLLLLLLL	9H	MMMLLLLFLLLLLLLLLLLMLL	7B	IIMMMMMMMMMMMMIMMMMMVI	7B
34	AATTTTTTTTTTTAAATTTTTTTV	6H	AAAAAAAACCCATTAACTTCAC	7H	TTTTTTGGGCCCCCGGATSCC		VVVVVVVVVVVVVVVVVVVVVV	*B
35	RNNNNNKKKASSKYYYYHHSSEEG		SSSSSSSSSSSSTTTTTTTTTST	7H	EEETSTTTTTSSSTLLNNTTHTT	T	SSSSSSSSSSSSSSSSSSSTS	7
36	RRRRRRRRRRRRRRRRRRRRRRR	*T	RRRRRRRRRRRRRRRRRRRRKRL	T	RRRRRRRRRRRRRRRRRRRRARR	T	RRRRRRRRRRRRRRRRRRRRRR	*T
37	GSSSSNTTSSNSSNSSSSSSSSS	6	TTTTTTTTTTTTTSSTTTTTTTTS	7T	SSSSSSSSNNNSSSSSSSSSSS	8T	SSSSSSSSSSSAAASSSSSSSS	7T
38	INNNNNNNNNNNTTTNNNNNNNN	5	PPPPPPNNNNNNNNNNNSSNKN	T	VVVYFYVVVVRRRAFFKKFYFYY		EEEEEEEEEEEEEEEEEEEEEE	*T
39	VDVAEAAAAAAATAAAVVRKAAQ	T	EEEEEAEEQQQKQEAAQRRENY	T	CCPPPPSSSNNNTPPQQPQPPP		IIIIIVVVIIIIIIIIILLIIII	8H
40	QQQQQQQQQQQQQQQQQQQQQQQ	*T	EEEEEEQQEEEEQQEEEEQQEQG	T	HHHQHHHHHHHHHHHQQHHQQQHH	5H	DDDDDDDDDDDDDDDDDDDDDD	*H
41	RRRRRRRRRRRRRRRRRRRRRRR	*T	ITTLLVMMLLLLTTTTTTTMMMLI	7H	LLLLLLLLLLLLLLLLLLLLLL	*H	LLMMMMMMLLLMMMLLMMLLLL	8H
42	LQQQQQQQQQQQQQQQQQQQQQQ	6T	RRRRRQHHQQQRRRRRRRKKARK	H	QQQRRRRRRRILLLKKRRRRRRR	H	ENNNNLLLLLLLMFFLLLLVG	H
43	EKDDDQQQQQEEQEEEQQEELLT	H	RRRAANQQEEEEDEEAAEDEQET	H	KKKRRKRKRHHHRLLLLAQLRR	H	TLDEEKKKNNDDDDNNDDDDQD	H
44	IIIIIIIIIIIIIIIIIIVIIII	8H	IIIIIIILLIIIIIIIIIIIIIII	9H	VVVVVVVVVVVVVVVVTTVVVTIVVV	6H	IIIIIIIIIIIIIIIIIIIIII	*H
45	AKAAAKKKKASAARRRRRRCCVVK	H	NSNKKKVVNNNNNTNNNVSSVMA	H	FFFFFFFFFFFFFFFFFMFFF	8H	KKKKKRRRRRRRRRRRRRRRKK	8H
46	EAFFFAAAEASKTTTQQQQKKA	H	QQQQQQAARRRRRRREERQQAKQ	H	EEDQQQDDDDDDEDDDDELEQQ	H	QCAVVSSSRRHAAAKKEETRQE	H c
47	AAAAAAAAEAASAAATTSSEQV	H	TTTAVVAAVVVVVVVVAACAAVAF	H	RRRKNNKKKEEEEEEEEEAEEE	H	EHFFFEEEEEEHNNEEIIEEMA	H
48	FFYYYYYYFFFFYYYFFYYYYYY	8H	YYYYYYYYYYYYYYYYYYYYYYYY	*H	YYYYYYYYYYYYYYYYYYYYYYFF	*H	FFYYYFFFFFFFFFFFFFFFFFF	8H
49	KKQQQLLLKKKKKKKKKKKKQQE	H	QQQEEMKKKKKRRRRKKQYYEEE	H	KKKTGRMMMKKKERRLLSEKII	T	EKQQQKKKIIRKKKRRRRKKAQ	H
50	TTRRRQQQTTTASSTSSSSAEA	H	LQLEEQDDEEEEEEEEESTTEEQ	H	SSSKKKTTTRRRKNNKKRNVKK	T	RKKKKRKREEKRRRKKTTKKQN	H
51	SSRRREEELLLQTTTHHLLAAE	T	QQQEEEEAAMMMEEEDDEVARDS	H	YYYYYYYIIIIIIIIIITMLLMK	5	IMMKKKKKKKNLLLNNKKHHMK	T
52	YYTTTTTTNFFFFIIIFFYYYYF	T	YYYYYYYYYYYYYLLLYYFYYYYYF	6T	SSSSSSSSSASASSTTASSTT	6	YYYYYYYYYYYFYYYFFYYYYYYY	8T
53	GGKKKGGGGGGGGGGGGGGGEE	T	GGGGGGEEEKKKKKKHHGKKQGG	H	PPPKQQGGGQQQNQQGGNGGNN		NGGGGGGGDDAGGGAAEEGGQG	
54	KKKKKKKKRRRRKRRRRRRKKKQR	6H	RRRSSARRTTTRRRKKRKKRSK	H	YYYIIIHFFFKKKKKKKKRNQYY	H	RKIIIKKKKKKKKKTTKKYCKK	
55	DDEEEPPPDDDDDDDDDDDDEAD	H	SSSNSNDDDDDDDDSSDSNPSS	H	DDDDDDQQQDDDSDDPPDDTDD	H	TTSPSSSSSSSSSSSSSSSSTS	5
56	LLLLLLLLLLLLLLLLLLLLLLL	*H	LLLLLLLLLLLLLLLLLLLLLLLL	*H	MMMMMMIIIIIIIIIIILIIVI	8H	LLLLLLLLLLLLLLLLLLLLLL	*H
57	IIPAADDDVLVTLIMMMIIKKV	H	EEEEEEEEEEEEAAAEEEGRAEE	H	LLLNNNEEEEEEEEVEEELEEE	H	HSCCCYYYHHHYYYYYYYYYYGE	B d
58	SKSSSEEENDDEDDDTAAEDAD	H	DDDDDDASKKKKKKDDKDDEES	H	EEEKKKEEEQQQDDDAASKQHH	H	SSQQQYYYQQQSSSSSSSSSTS	5B
59	DDAAATTVDDDTDDDDDDDDDD	H	DDDDDKDDDDDDDDAADDDQDD	H	SSSVAATTTSSSSSSSSAATV	5H	AMAAAYYYAAMFFFMMMMAAMW	5B
60	LLLLLLLMLLLLLLLLLLLLLL	8H	IIIVVIIIIIIIIIILLIIIIMIL	7H	IIILLLIIIIIIIIIIIIVIIII	8H	VIIIIIIIIIIIIIIIIIIIII	8B
61	KKKKKKKRKKKKKKKKKKKKKK	8H	RRRVVTIIIIVITTTTSSRSSCQK	T	KKRDDDDDKKKKKKRRSKKKK	H	VMLLLQQQEEQKKKKKKKKQAK	B
62	SSSSSKKKSSSSSSSSSYYGGD	H	SSSGGGGGSSSSSSSSSSSSSAG	7	KKKLLLRRRSSSSGGGGRRHKK	H	DEDDDQQQGGKGGGGGNNSSGE	T
63							A	
64	EEAAAAAAEEEEEEEEEEEEEDDE	H	DDDDDEDDDDDDDDDDDDEEEDD	6	EEEEEEEEEEEEEEEEEEEEEEE	*H	EDEEEDDDDDDDDDDDDDDDDD	8T
65	LLLLLLLLLLLLLLLLIILLLLLL	9H	TTTTTTTTTTTTTTTTTTTTTTT	*	VVVLLLTTTTTTTLLLLFFMMM	5T	TTTTTTTTTTTTTTTTTTTTTA	7T
66	GSSSSTTTTTTSSSSSSTTSSG	6H	SSSSSSSSSSSSSSSSSSSSSSS	*	KKKKKKSSSSSSHSSSSSSSS	T	SSKKKKKKSSSSSSSSSSSSSE	T
67	GGGGGGGGGGGGGGGGGGGGGGG	*H	FFFGGGGGGGGGGGGGGGGGGGG	6	GGGGGGGGGGGGGGGGGGGDGG	T	GGGGGGGGGGGGGGGGGGGGGT	7T
68	KNHHHHHHHHHKKNNNDDKKHHK	H	MMMMYYHHHHDDDDDDHHHHDDFYH	T	DDDDDDDNNSSSSHHDDYSEDD	T	DDDDDDDDDDDDDDDEEDDDD	8T
69	FMLLLLLLFFFFFFFLLFFFFFF	7H	FFFYYFFFFFFFFFFFFFFFFLF	6T	LLLIIILLLFFFLFFFFVVLVV	6T	YYYYYYYYFFYYYYYYYYYYYI	6T
70	EEEEEEEEEEEEEEEEAAEEEEEE	H	QQQQQQQQRRRRRRRRERRREK	5T	EEEEEEEEEEEEEEEEEEEEHRK	5H	KKEEEQQQLMRRRRRKKKKERRG	T
71	DETTTEEEKKTRQQQRRRRHHD	H	RRRRRRRKKKKKNNKRRRKKRRR	5T	NNNKKKKQNDDDEDDKKSKEDD	H	RNKKKKKKKKKKKKKKKKKITRY	H
72	VLVVVVVVVLLLLVVVLLLLLLVV	8H	VVVMMLMMLLLAAAIILAALIL	6B	AAACCCLLLAAAALLLLGGAAA	6H	AATTTAAAAAAVVVAATATVLV	5H e
73	IIIIIVVVIIMIIIIIIIIIMMI	7H	LLLLLLLMMMLLLLLLLLLLC	7B	FFFLLLLLLLLLMLLMMLFMFF	7H	LLLLLLLLLLLLLLLLLLLLLL	*H
74	LLLLLLLVVVVVLVLLLVVVVVV	8H	VVVVVVVVVVVVLLLIIVLLTVV	6B	LLLTTTLLLLLLLLLLLLKLMVV	H	TLVVVLLLLLLLLLLLLLLLLV	6B
75	AAGGGAAAAASAGGGGGGNAAG	6H	SSSVVVVVVAAASAASSSTTLCS	B	NNNATTAAAAAATAAAATAAAA	6H	ASAAAYYYAALVIILLKKKKAT	H
76	LLLLLLMMLLLLMMMLLLLLLL	8H	LLLLLLLLLLLLLLLLMLLTLL	8B	LLLIIIVVVIIIVIVVVIIIII	8H	LLLLLLLLLLLLLLLLLLLIIIL	9X
77	MFLLLLLLMMMMMMMMMMMMVIM	6H	SSSLLLLLLAAAAAAAACAAVLV		VVVVVVVVVVVVVVVVVVV VVVV	H	LVCCCCCCCCCCCCCCSCCCVT	B

Charge Comparison

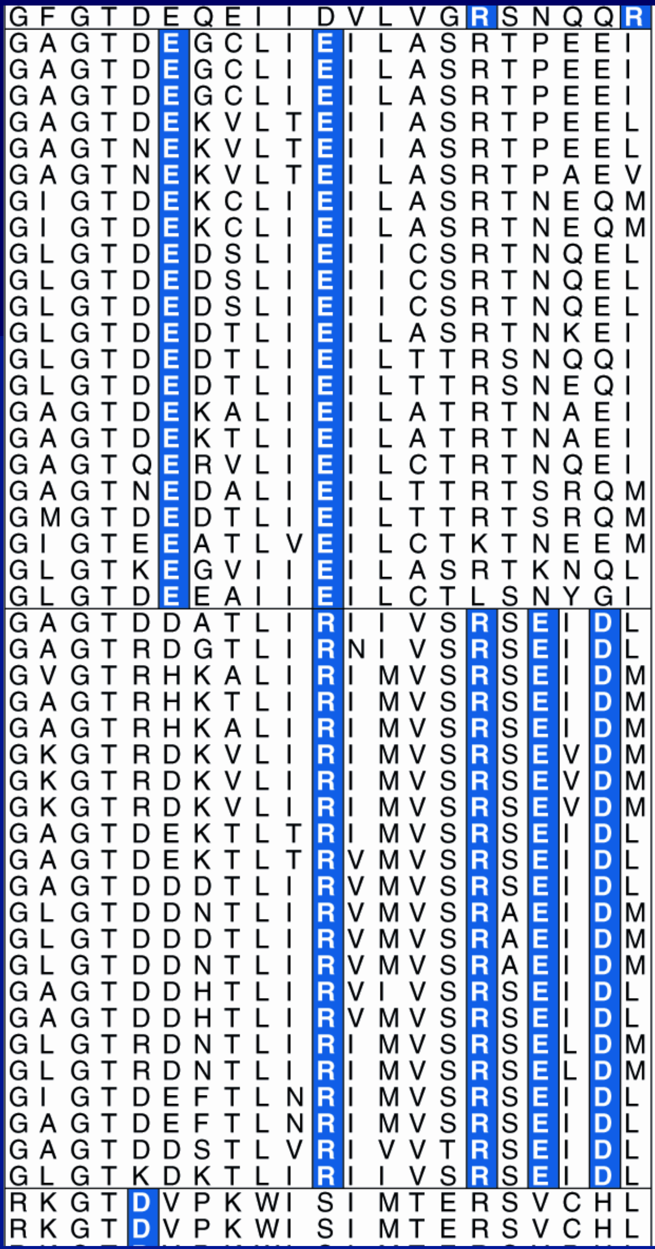


```
      1                 10                  20
L9r1  G F G T D E K A I I E I L A R R G I V Q R
p471  G F G T D E Q A I V D V V A N R S N D Q R
L2M1  T K G V D E V T I V N I L T N R S N V Q R
L2H1  T K G V D E V T I V N I L T N R S N A Q R
L2B1  T K G V D E V T I V N I L T N R S N E Q R
L1H1  V K G V D E A T I I D I L T K R N N A Q R
L1R1  V K G V D E A T I I D I L T K R T N A Q R
L1M1  V K G V D E A T I I D I L T K R T N A Q R
L5R1  G L G T D E D S I L N L L T A R S N A Q R
L5H1  G L G T D E E S I L T L L T S R S N A Q R
L5C1  G M G T D E E T I L K I L T S R N N A Q R
L8H1  G I G T N E Q A I I D V L T K R S N T Q R
L4P1  G L G T D E D A I I S V L A Y R S T A Q R
L4H1  G L G T D E D A L I S V L A Y R N T A Q R
L4B1  G L G T D E D A I I N V L A Y R S T A Q R
L6H5  G L G T D E D T I I D I I T H R S N V Q R
L6M5  G I G T D E A T I I D I V T H R S N V Q R
L6H1  G F G S D K E A I L D I I T S R S N R Q R
L6M1  G F G S D K E S I L E L I T S R S N K Q R
L3H1  G I G T D E K M L I S I L T E R S N A Q R
L3R1  G I G T D E K T L I N I L T E R S N A Q R
L10r1 G F G T D E Q E I I D V L V G R S N Q Q R
L4P2  G A G T D E G C L I E I L A S R T P E E I
L4H2  G A G T D E G C L I E I L A S R T P E E I
L4B2  G A G T D E G C L I E I L A S R T P E E I
L5R2  G A G T D E K V L T E I I A S R T P E E L
L5H2  G A G T N E K V L T E I I A S R T P E E L
L5C2  G A G T N E K V L T E I L A S R T P A E V
L6H2  G I G T D E K C L I E I L A S R T N E Q M
L6M2  G I G T D E K C L I E I L A S R T N E Q M
L2H2  G L G T D E D S L I E I I C S R T N Q E L
L2B2  G L G T D E D S L I E I I C S R T N Q E L
L2M2  G L G T D E D S L I E I I C S R T N Q E L
L1H2  G L G T D E D T L I E I L A S R T N K E I
L1R2  G L G T D E D T L I E I L T T R S N Q Q I
L1M2  G L G T D E D T L I E I L T T R S N E Q I
L6H6  G A G T D E K A L I E I L A T R T N A E I
L6M6  G A G T D E K T L I E I L A T R T N A E I
p472  G A G T Q E R V L I E I L C T R T N Q E I
L3H2  G A G T N E D A L I E I L T T R T S R Q M
L3R2  G M G T D E D T L I E I L T T R T S R Q M
L10r2 G I G T E E A T L V E I L C T K T N E E M
L8H2  G L G T K E G V I I E I L A S R T K N Q L
L9r2  G L G T D E E A I I E I L C T L S N Y G I
L10r4 G A G T D D A T L I R I I V S R S E I D L
L8H4  G A G T R D G T L I R N I V S R S E I D L
L1H4  G V G T R H K A L I R I M V S R S E I D M
L1R4  G A G T R H K T L I R I M V S R S E I D M
L1M4  G A G T R H K A L I R I M V S R S E I D M
L2M4  G K G T R D K V L I R I M V S R S E V D M
L2B4  G K G T R D K V L I R I M V S R S E V D M
L2H4  G K G T R D K V L I R I M V S R S E V D M
L6H8  G A G T D E K T L T R I M V S R S E I D L
L6M8  G A G T D E K T L T R V M V S R S E I D L
L5C4  G A G T D D D T L I R V M V S R S E I D L
L4H4  G L G T D D N T L I R V M V S R A E I D M
L4B4  G L G T D D D T L I R V M V S R A E I D M
L4P4  G L G T D D N T L I R V M V S R A E I D M
L5R4  G A G T D D H T L I R V I V S R S E I D L
L5H4  G A G T D D H T L I R V M V S R S E I D L
L6H4  G L G T R D N T L I R I M V S R S E L D M
L6M4  G L G T R D N T L I R I M V S R S E L D M
L3H4  G I G T D E F T L N R I M V S R S E I D L
L3R4  G A G T D E F T L N R I M V S R S E I D L
p474  G A G T D D S T L V R I V V T R S E I D L
L9r4  G L G T K D K T L I R I I V S R S E I D L
L2M3  R K G T D V P K W I S I M T E R S V C H L
L2B3  R K G T D V P K W I S I M T E R S V C H L
L2H3  R K G T D V P K W I S I M T E R S V P H L
L1H3  R K G T D V N V F N T I L T T R S Y P Q L
L1M3  R K G T D V N V F T T I L T S R S F P H L
L1R3  R K G T D V N V F N T I L T T R S Y P H L
L5C3  K W G T D E E T F I T I L G T R S V S H L
L5H3  K W G T D E E K F I T I F G T R S V S H L
L5R3  K W G T D E E K F I T I L G T R S V S H L
L4H3  K W G T D E V K F L T V L C S R N R N H L
L4B3  K W G T D E V K F L T V L C S R N R N H L
L4P3  K W G T D E V K F L T V L C S R N R N H L
L8H3  I R G T D E M K F I T I L C T R S A T H L
L3H3  R W G T D E D K F T E I L C L R S F P Q L
L3R3  K W G T D E D K F T E I L C L R S F P Q L
L6H3  K W G T D E A Q F I Y I L G N R S K Q H L
L6M3  K W G T D E A Q F I Y I L G N R S K Q H L
p473  R L G T D E S C F N M I L A T R S F P Q L
L9r3  Q W G T D E S T F N S I L I T R S Y Q Q L
L10r3 K L G T D E E V F N R I M S H A S F P Q L
L6H7            E T R F M T I L C T R S Y P H L
L6M7            E T R F M T V L C T R S Y P H L
```

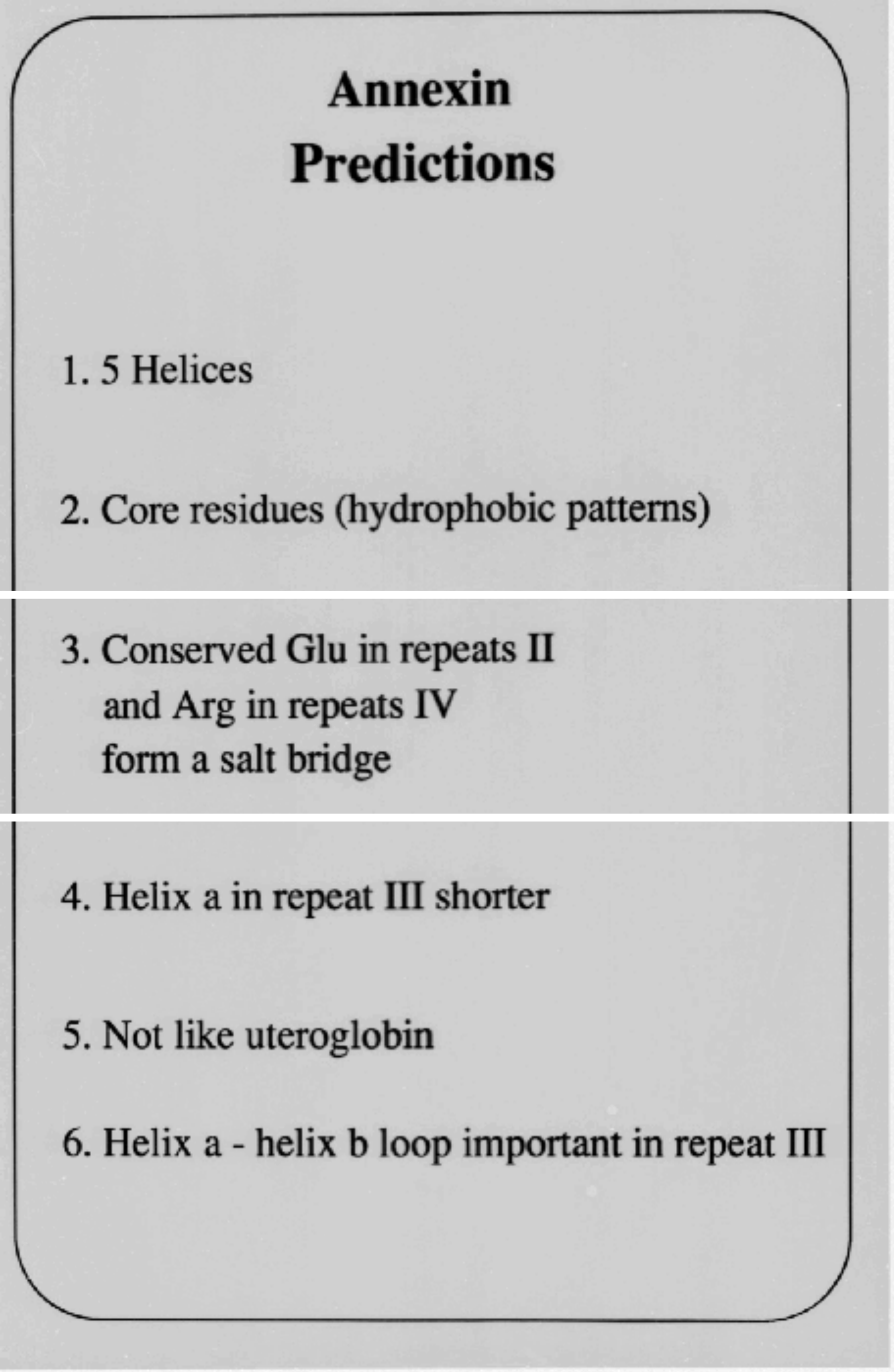
2

1

AMAS (c) 1992 G.J.Barton & C.D.Livingstone - Univ. Oxford UK.
Conservation plot of the sequence alignment "/ebi/www/web/Barton/temp/amas242767.blc"
Conservation Threshold: 2 - Gaps Ignored: 0 - Minimum Residue Occupancy: 0%
Histogram type: PAIR - Conservation mode: POSITIVE with 3 property index "ch.pt"



```
G F G T D E Q E I I D V L V G R S N Q Q R
G A G T D E G C L I E I L A S R T P E E I
G A G T D E G C L I E I L A S R T P E E I
G A G T D E G C L I E I L A S R T P E E I
G A G T D E K V L T E I I A S R T P E E L
G A G T N E K V L T E I I A S R T P E E L
G A G T N E K V L T E I L A S R T P A E V
G I G T D E K C L I E I L A S R T N E Q M
G I G T D E K C L I E I L A S R T N E Q M
G L G T D E D S L I E I I C S R T N Q E L
G L G T D E D S L I E I I C S R T N Q E L
G L G T D E D S L I E I I C S R T N Q E L
G L G T D E D T L I E I L A S R T N K E I
G L G T D E D T L I E I L T T R S N Q Q I
G L G T D E D T L I E I L T T R S N E Q I
G A G T D E K A L I E I L A T R T N A E I
G A G T D E K T L I E I L A T R T N A E I
G A G T Q E R V L I E I L C T R T N Q E I
G A G T N E D A L I E I L T T R T S R Q M
G M G T D E D T L I E I L T T R T S R Q M
G I G T E E A T L V E I L C T K T N E E M
G L G T K E G V I I E I L A S R T K N Q L
G L G T D E E A I I E I L C T L S N Y G I
G A G T D D A T L I R I I V S R S E I D L
G A G T R D G T L I R N I V S R S E I D L
G V G T R H K A L I R I M V S R S E I D M
G A G T R H K T L I R I M V S R S E I D M
G A G T R H K A L I R I M V S R S E I D M
G K G T R D K V L I R I M V S R S E V D M
G K G T R D K V L I R I M V S R S E V D M
G K G T R D K V L I R I M V S R S E V D M
G A G T D E K T L T R I M V S R S E I D L
G A G T D E K T L T R V M V S R S E I D L
G A G T D D D T L I R V M V S R S E I D L
G L G T D D N T L I R V M V S R A E I D M
G L G T D D D T L I R V M V S R A E I D M
G L G T D D N T L I R V M V S R A E I D M
G A G T D D H T L I R V I V S R S E I D L
G A G T D D H T L I R V M V S R S E I D L
G L G T R D N T L I R I M V S R S E L D M
G L G T R D N T L I R I M V S R S E L D M
G I G T D E F T L N R I M V S R S E I D L
G A G T D E F T L N R I M V S R S E I D L
G A G T D D S T L V R I V V T R S E I D L
G L G T K D K T L I R I I V S R S E I D L
R K G T D V P K W I S I M T E R S V C H L
R K G T D V P K W I S I M T E R S V C H L
```



Annexin
Predictions

1. 5 Helices

2. Core residues (hydrophobic patterns)

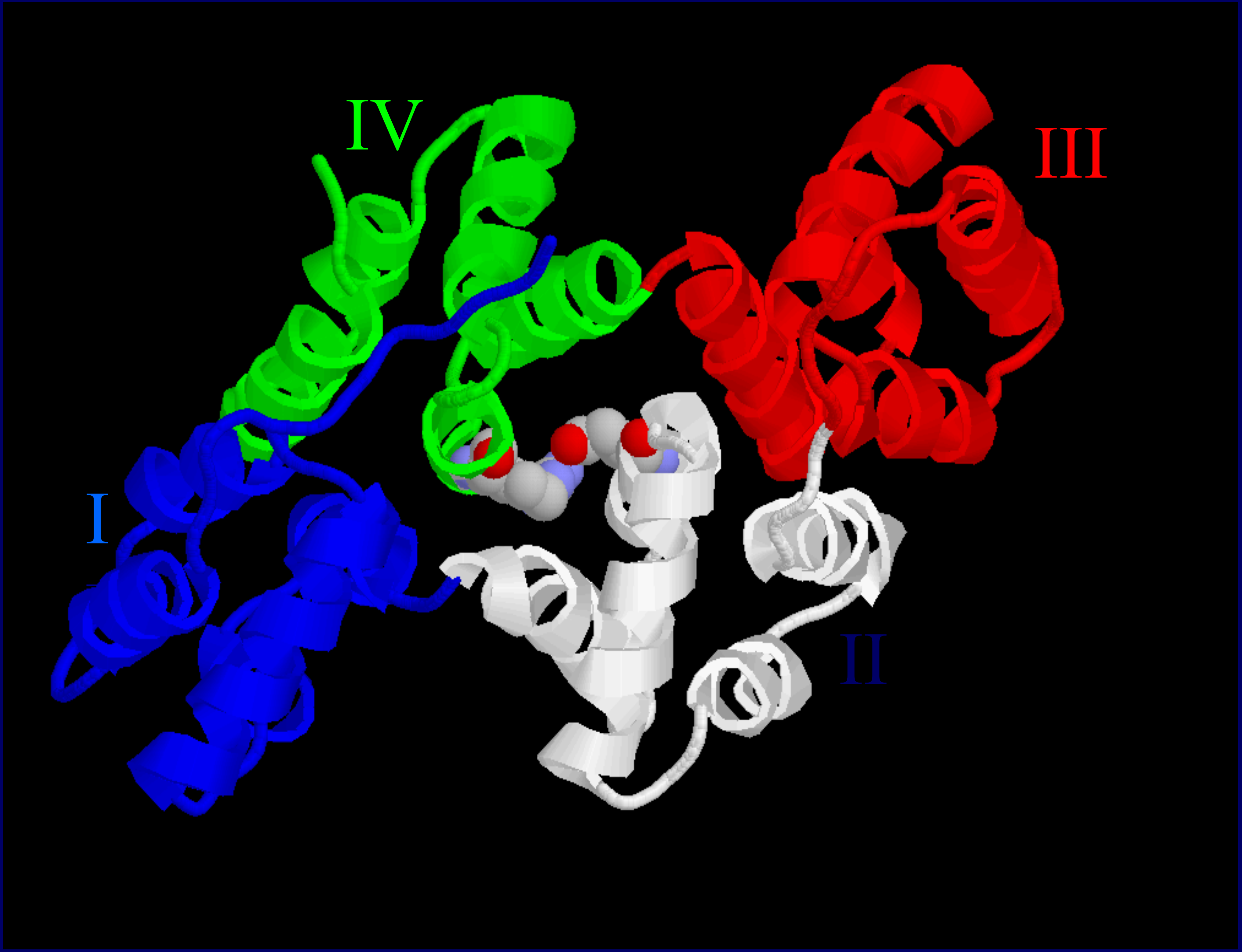
3. Conserved Glu in repeats II
   and Arg in repeats IV
   form a salt bridge

4. Helix a in repeat III shorter

5. Not like uteroglobin

6. Helix a - helix b loop important in repeat III

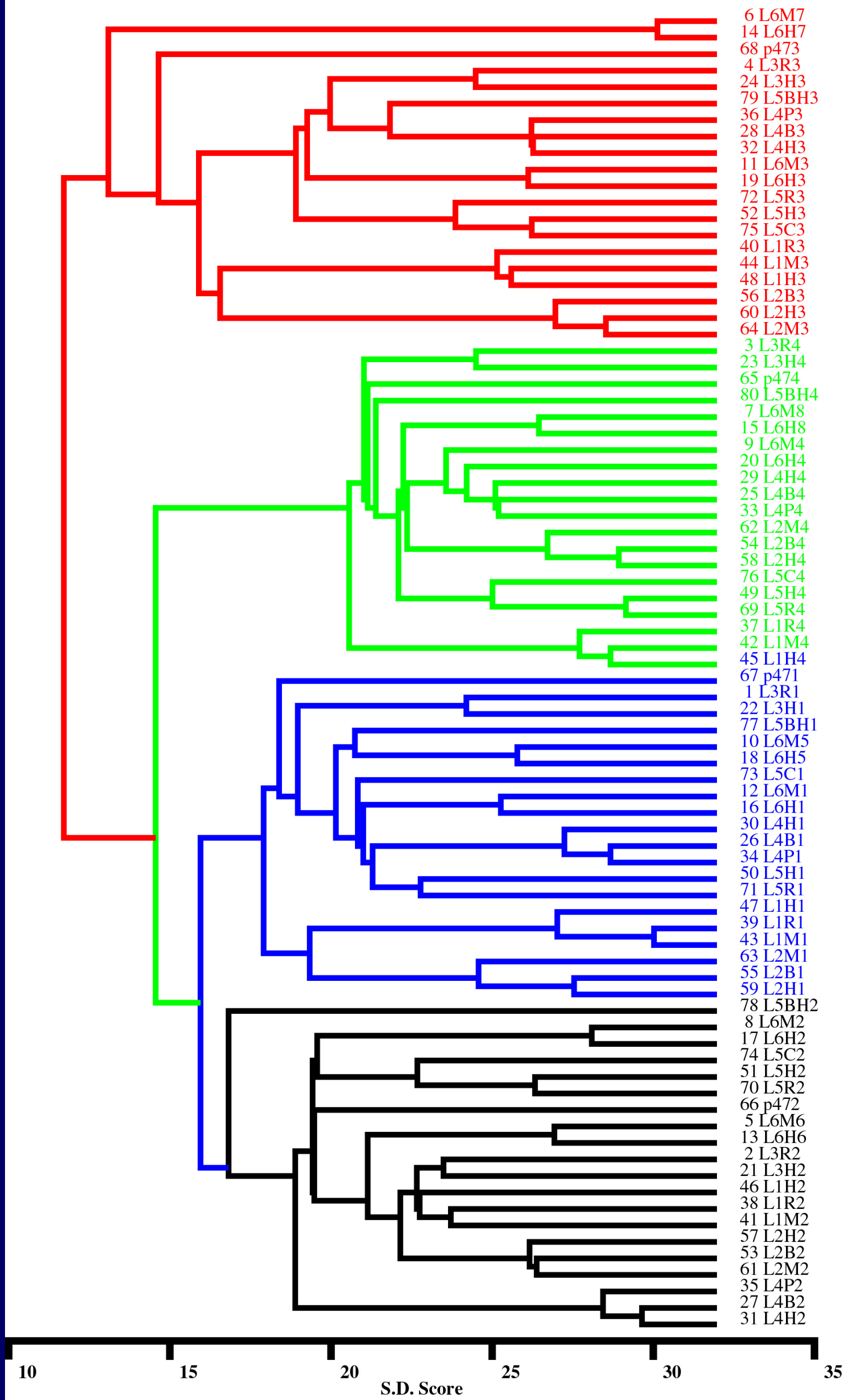
Annexin V showing Glu-Arg salt bridge between helix 2 of domain II and helix 2 of domain IV



Analysis of similarities and differences between sub-families can reveal functionally important residues

Generalise lessons learned in Annexin study

Annexin Domains

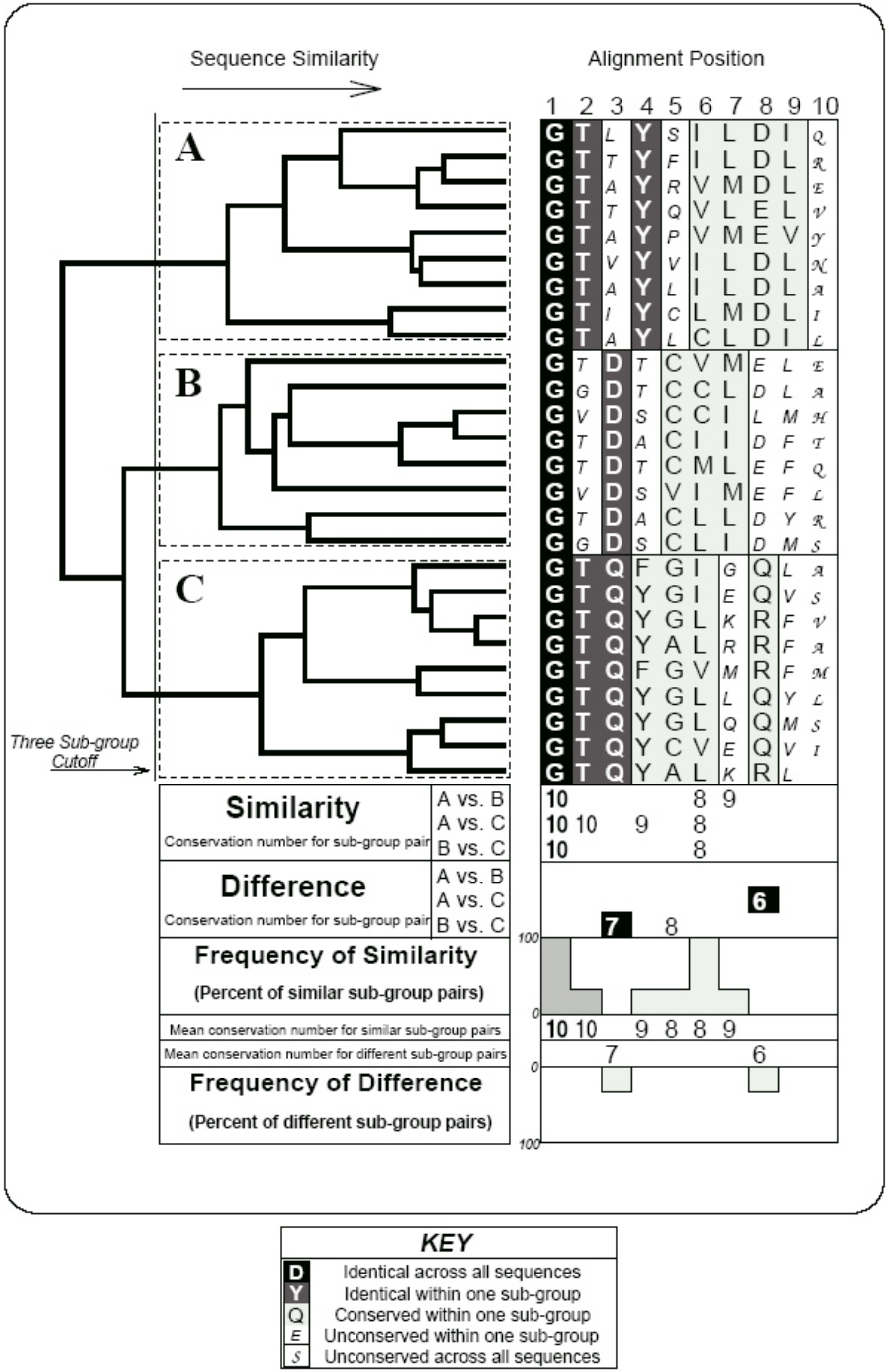


Domain
3 & 7

Domain
4 & 8

Domain
1 & 5

Domain
2 & 6



	1	2	3	4	5	6	7	8	9	10
A	G	T	L	Y	S	I	L	D	I	Q
	G	T	T	Y	F	I	L	D	L	R
	G	T	A	Y	R	V	M	D	L	E
	G	T	T	Y	Q	V	L	E	L	V
	G	T	A	Y	P	V	M	E	V	Y
	G	T	V	Y	V	I	L	D	L	N
	G	T	A	Y	L	I	L	D	L	A
	G	T	I	Y	C	L	M	D	L	I
	G	T	A	Y	L	C	L	D	I	L
B	G	T	D	T	C	V	M	E	L	E
	G	G	D	T	C	C	L	D	L	A
	G	V	D	S	C	C	I	L	M	H
	G	T	D	A	C	I	I	D	F	T
	G	T	D	T	C	M	L	E	F	Q
	G	V	D	S	V	I	M	E	F	L
	G	T	D	A	C	L	L	D	Y	R
	G	G	D	S	C	L	I	D	M	S
C	G	T	Q	F	G	I	G	Q	L	A
	G	T	Q	Y	G	I	E	Q	V	S
	G	T	Q	Y	G	L	K	R	F	V
	G	T	Q	Y	A	L	R	R	F	A
	G	T	Q	F	G	V	M	R	F	M
	G	T	Q	Y	G	L	L	Q	Y	L
	G	T	Q	Y	G	L	Q	Q	M	S
	G	T	Q	Y	C	V	E	Q	V	I
	G	T	Q	Y	A	L	K	R	L	

		1	2	3	4	5	6	7	8	9	10
Similarity Conservation number for sub-group pair	A vs. B	10					8	9			
	A vs. C	10	10		9		8				
	B vs. C	10					8				
Difference Conservation number for sub-group pair	A vs. B										
	A vs. C								6		
	B vs. C			7		8					
Frequency of Similarity (Percent of similar sub-group pairs)											
Mean conservation number for similar sub-group pairs		10	10		9	8	8	9			
Mean conservation number for different sub-group pairs				7					6		
Frequency of Difference (Percent of different sub-group pairs)											

KEY	
D	Identical across all sequences
Y	Identical within one sub-group
Q	Conserved within one sub-group
E	Unconserved within one sub-group
S	Unconserved across all sequences

Figure 3

Principles of sub-family analysis

See what happens to conservation when you put two sub-families Together.

Does it stay high?
Implies- position is important to both and doing a similar job.

Does it go from high to low?
Implies- position is important to both but the position is important for novel features of the two sub-families.

References on Sub-family analysis

- Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. Comput Appl Biosci. 1993 Dec;9(6):745-56.
- Identification of functional residues and secondary structure from protein multiple sequence alignment. Methods Enzymol. 1996;266:497-512.

- Methods Enzymol paper includes summary of first paper.
- Copies are available on: http://www.compbio.dundee.ac.uk/ftp/pdf/

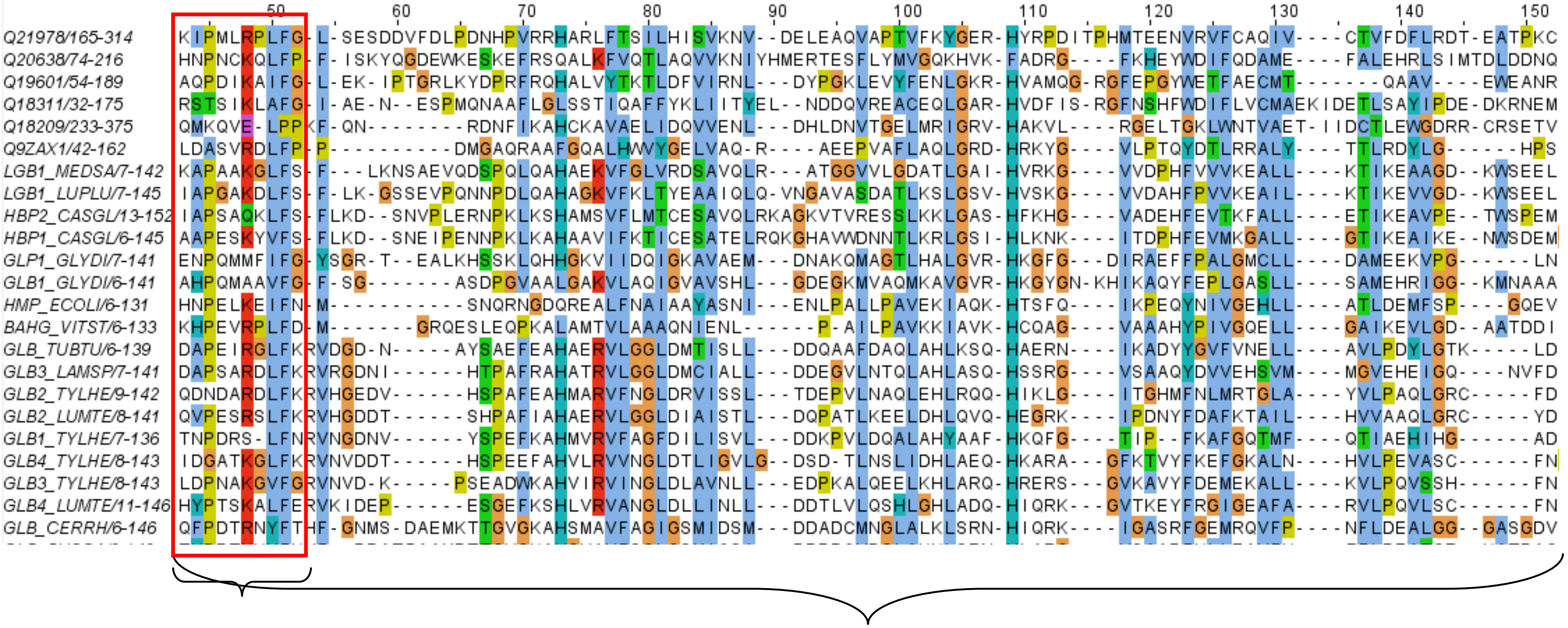
Identification of functional sites

- Whole alignment methods
  - Simple visualisation
  - Calculation of “conservation values”

- Sub-family analysis
  - AMAS analysis
  - “Tree determinant” positions
  - “Evolutionary trace”

Tree determinants

- Positions in the alignment that are most responsible for the topology of the phylogenetic tree derived from the complete alignment
- These positions may be functionally important.

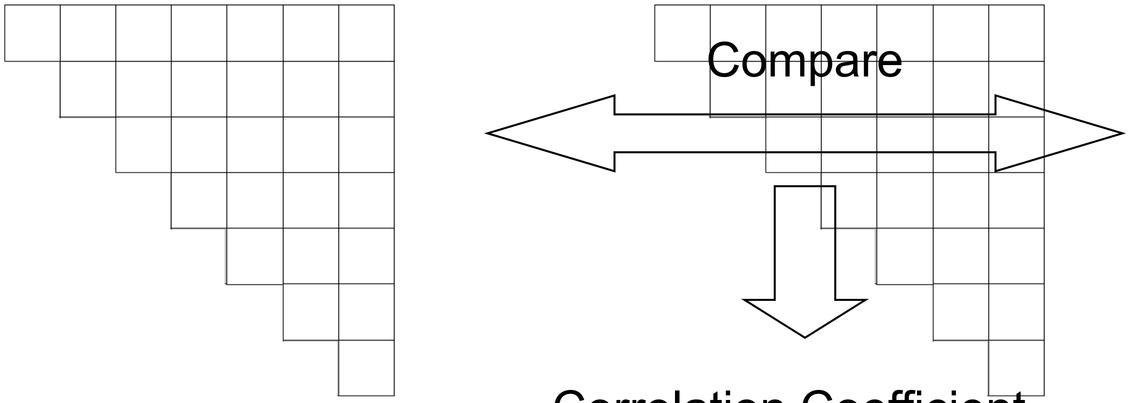


```
                        50        60        70        80        90       100       110       120       130       140       150
Q21978/165-314    KIPMLRPLFG-L-SESDDVFDLPDNHPVRRHARLFTSILHISVKNV--DELEAQVAPTVFKYGER-HYRPDITPHMTEENVRVFCAQIV----CTVFDFLRDT-EATPKC
Q20638/74-216     HNPNCKQLFP-F-ISKYQGDEWKESKEFRSQALKFVQTLAQVVKNIYHMERTESFLYMVGQKHVK-FADRG----FKHEYWDIFQDAME----FALEHRLSIMTDLDDNQ
Q19601/54-189     AQPDIKAIFG-L-EK-IPTGRLKYDPRFRQHALVYTKTLDFVIRNL---DYPGKLEVYFENLGKR-HVAMQG-RGFEPGYWETFAECMT-------QAAV----EWEANR
Q18311/32-175     RSTSIKLAFG-I-AE-N--ESPMQNAAFLGLSSTIQAFFYKLIITYEL--NDDQVREACEQLGAR-HVDFIS-RGFNSHFWDIFLVCMAEKIDETLSAYIPDE-DKRNEM
Q18209/233-375    QMKQVE-LPPKF-QN--------RDNFIKAHCKAVAELIDQVVENL---DHLDNVTGELMRIGRV-HAKVL----RGELTGKLWNTVAET-IIDCTLEWGDRR-CRSETV
Q9ZAX1/42-162     LDASVRDLFP-P----------DMGAQRAAFGQALHWVYGELVAQ-R----AEEPVAFLAQLGRD-HRKYG----VLPTQYDTLRRALY----TTLRDYLG------HPS
LGB1_MEDSA/7-142  KAPAAKGLFS-F---LKNSAEVQDSPQLQAHAEKVFGLVRDSAVQLR---ATGGVVLGDATLGAI-HVRKG----VVDPHFVVVKEALL----KTIKEAAGD--KWSEEL
LGB1_LUPLU/7-145  IAPGAKDLFS-F-LK-GSSEVPQNNPDLQAHAGKVFKLTYEAAIQLQ-VNGAVASDATLKSLGSV-HVSKG----VVDAHFPVVKEAIL----KTIKEVVGD--KWSEEL
HBP2_CASGL/13-152 IAPSAQKLFS-FLKD--SNVPLERNPKLKSHAMSVFLMTCESAVQLRKAGKVTVRESSLKKLGAS-HFKHG----VADEHFEVTKFALL----ETIKEAVPE--TWSPEM
HBP1_CASGL/6-145  AAPESKYVFS-FLKD--SNEIPENNPKLKAHAAVIFKTICESATELRQKGHAWDNNTLKRLGSI-HLKNK----ITDPHFEVMKGALL----GTIKEAIKE--NWSDEM
GLP1_GLYDI/7-141  ENPQMMFIFG-YSGR-T--EALKHSSKLQHHGKVIIDQIGKAVAEM---DNAKQMAGTLHALGVR-HKGFG---DIRAEFFPALGMCLL----DAMEEKVPG------LN
GLB1_GLYDI/6-141  AHPQMAAVFG-F-SG-------ASDPGVAALGAKVLAQIGVAVSHL---GDEGKMVAQMKAVGVR-HKGYGN-KHIKAQYFEPLGASLL----SAMEHRIGG--KMNAAA
HMP_ECOLI/6-131   HNPELKEIFN-M-----------SNQRNGDQREALFNAIAAYASNI---ENLPALLPAVEKIAQK-HTSFQ----IKPEQYNIVGEHLL----ATLDEMFSP----GQEV
BAHG_VITST/6-133  KHPEVRPLFD-M-------GRQESLEQPKALAMTVLAAAQNIENL------P-AILPAVKKIAVK-HCQAG----VAAAHYPIVGQELL----GAIKEVLGD--AATDDI
GLB_TUBTU/6-139   DAPEIRGLFKRVDGD-N-----AYSAEFEAHAERVLGGLDMTISLL---DDQAAFDAQLAHLKSQ-HAERN----IKADYYGVFVNELL----AVLPDYLGTK-----LD
GLB3_LAMSP/7-141  DAPSARDLFKRVRGDNI------HTPAFRAHATRVLGGLDMCIALL---DDEGVLNTQLAHLASQ-HSSRG----VSAAQYDVVEHSVM----MGVEHEIGQ----NVFD
GLB2_TYLHE/9-142  QDNDARDLFKRVHGEDV------HSPAFEAHMARVFNGLDRVISSL---TDEPVLNAQLEHLRQQ-HIKLG----ITGHMFNLMRTGLA----YVLPAQLGRC-----FD
GLB2_LUMTE/8-141  QVPESRSLFKRVHGDDT------SHPAFIAHAERVLGGLDIAISTL---DQPATLKEELDHLQVQ-HEGRK----IPDNYFDAFKTAIL----HVVAAQLGRC-----YD
GLB1_TYLHE/7-136  TNPDRS-LFNRVNGDNV------YSPEFKAHMVRVFAGFDILISVL---DDKPVLDQALAHYAAF-HKQFG----TIP--FKAFGQTMF----QTIAEHIHG------AD
GLB4_TYLHE/8-143  IDGATKGLFKRVNVDDT------HSPEEFAHVLRVVNGLDTLIGVLG--DSD-TLNSLIDHLAEQ-HKARA---GFKTVYFKEFGKALN----HVLPEVASC------FN
GLB3_TYLHE/8-143  LDPNAKGVFGRVNVD-K-----PSEADWKAHVIRVINGLDLAVNLL---EDPKALQEELKHLARQ-HRERS---GVKAVYFDEMEKALL----KVLPQVSSH------FN
GLB4_LUMTE/11-146 HYPTSKALFERVKIDEP------ESGEFKSHLVRVANGLDLLINLL---DDTLVLQSHLGHLADQ-HIQRK---GVTKEYFRGIGEAFA----RVLPQVLSC------FN
GLB_CERRH/6-146   QFPDTRNYFTHF-GNMS-DAEMKTTGVGKAHSMAVFAGIGSMIDSM---DDADCMNGLALKLSRN-HIQRK----IGASRFGEMRQVFP----NFLDEALGG--GASGDV
```

Generate Local Matrix

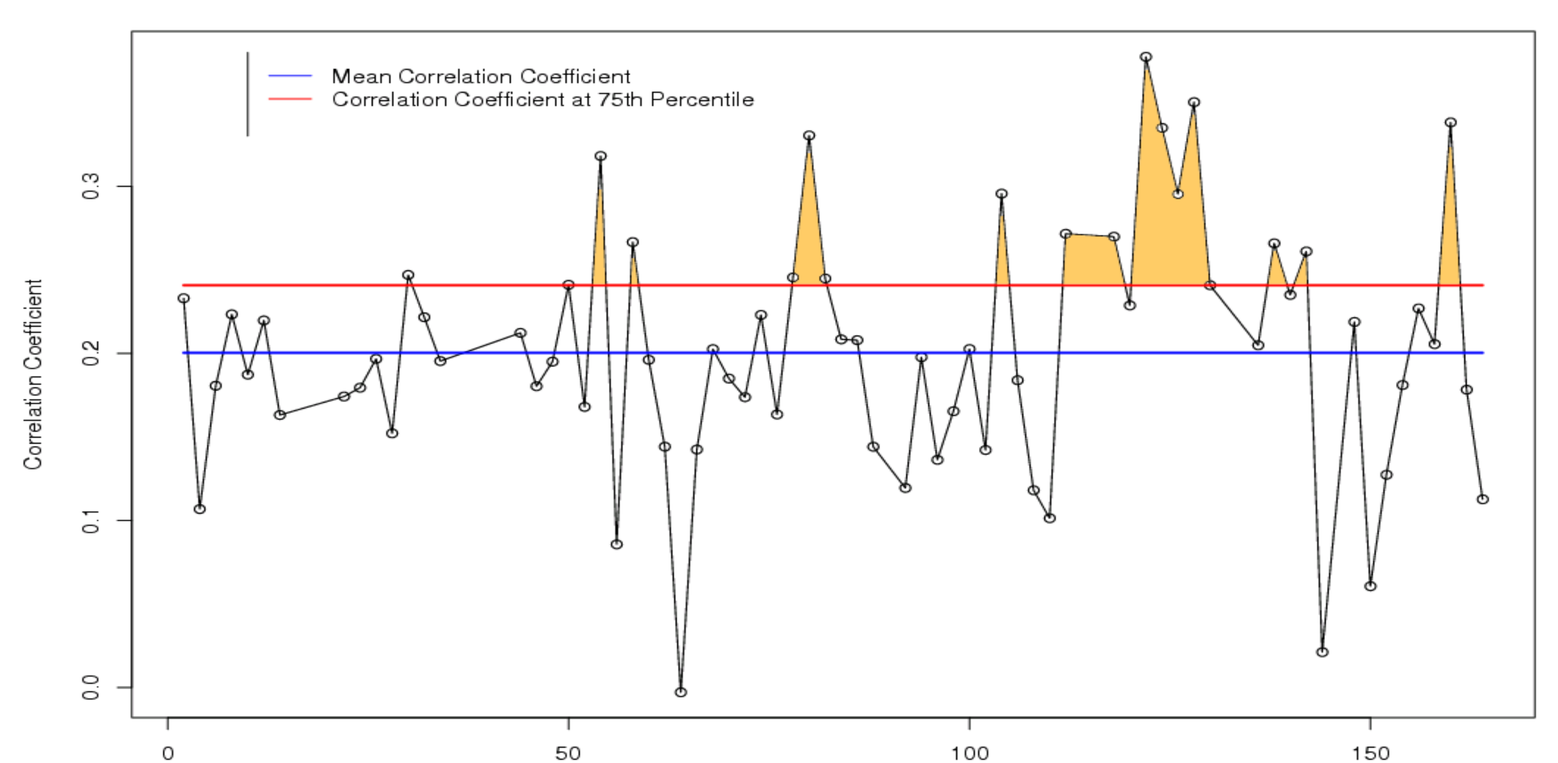
Generate Global Distance Matrix

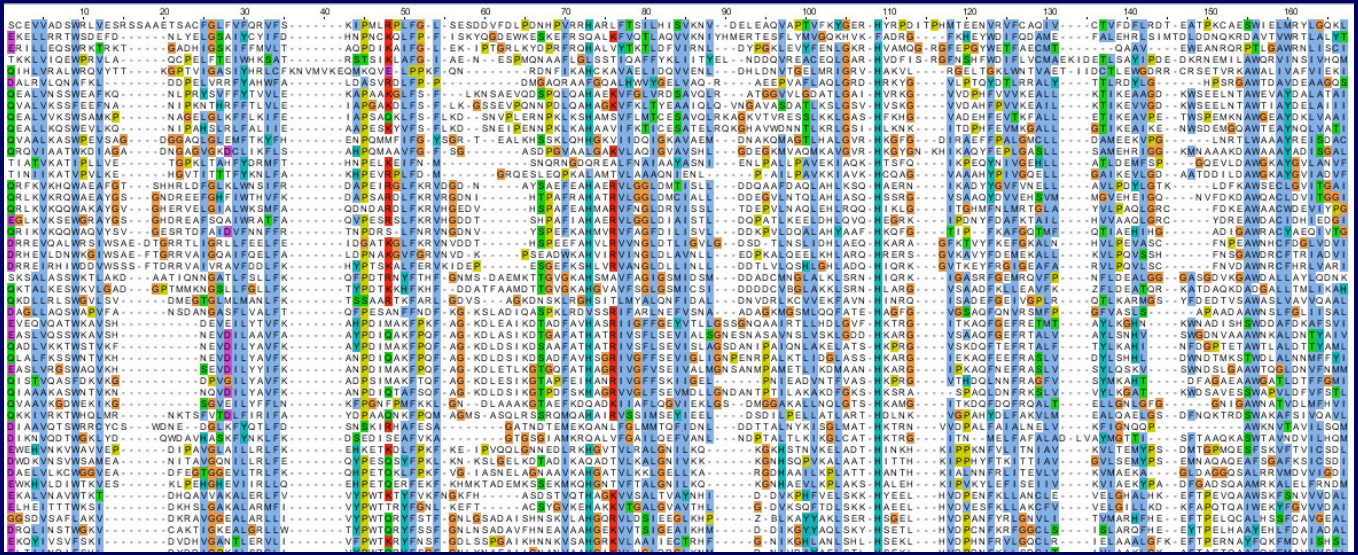
Saved Global Matrix



Compare

Correlation Coefficient



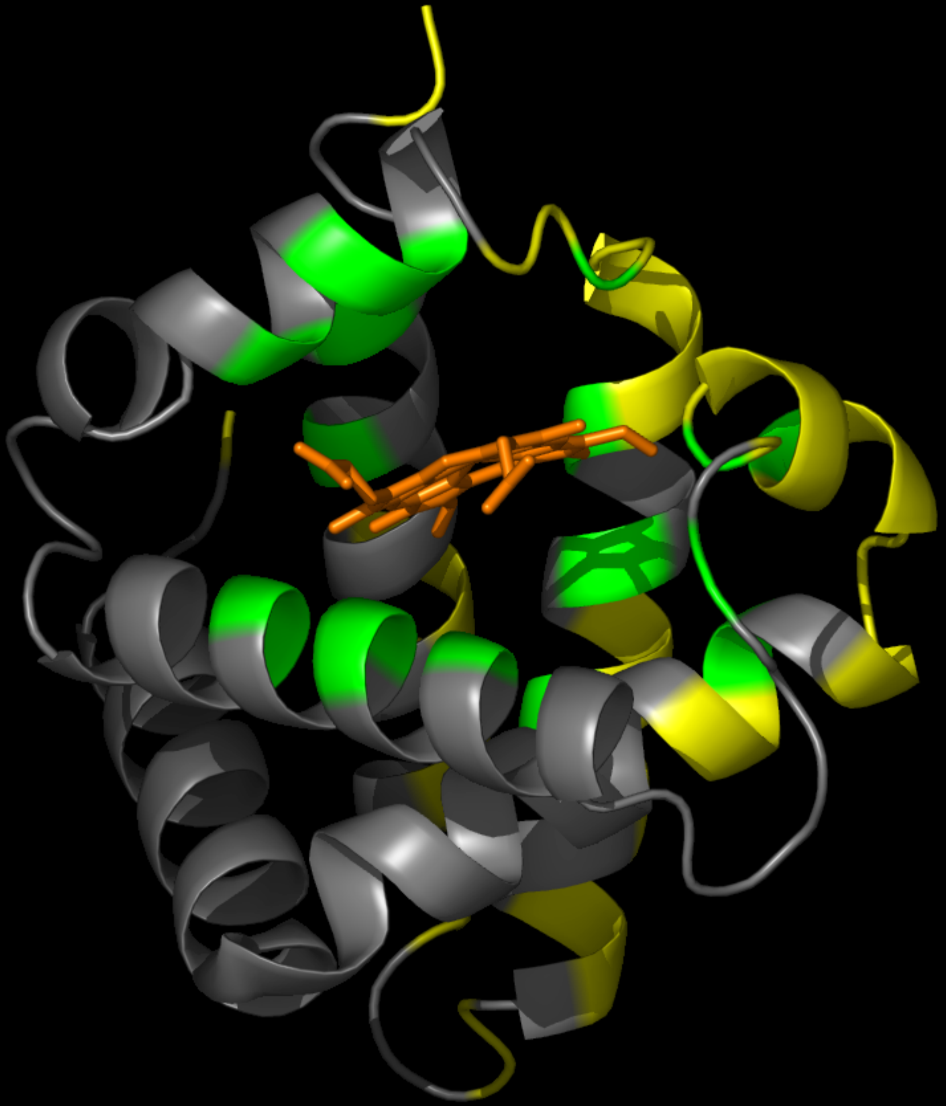


Partial Multiple Sequence Alignment

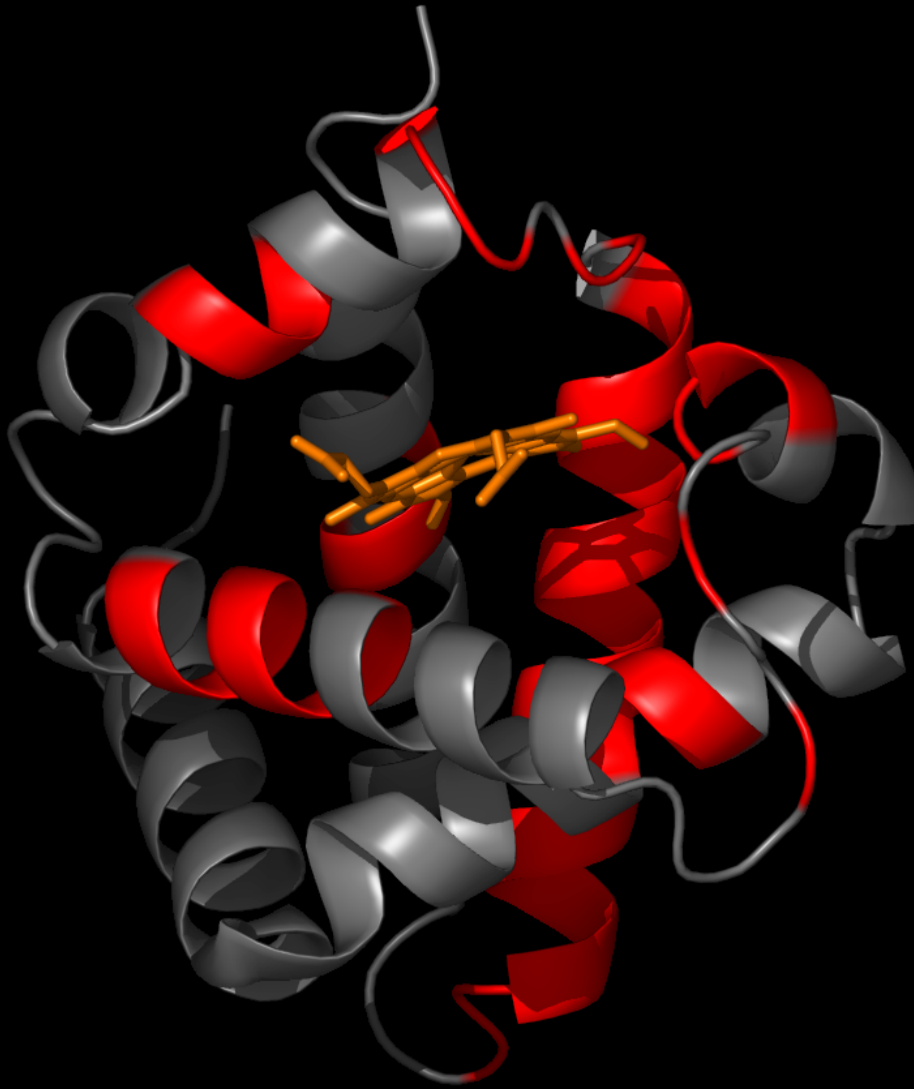
(Globins)

Observed Haeme Contacts (green)

Observed Protein Contacts (yellow)



Prediction (window size 3, interval 2)



Papers on tree determinants/predicting functional sites

- del Sol Mesa A., Pazos F., Valencia A. Automatic methods for predicting functionally important residues. J. Mol. Biol. 2003;326:1289–1302. [PubMed]

- Mihalek I., Res I., Lichtarge O. A family of evolution-entropy hybrid methods for ranking protein residues by importance. J. Mol. Biol. 2004;336:1265–1282. [PubMed]

Practicals on sub-family analysis