ONLINE SUPPLEMENT

Plasmodium Interspersed Repeats: The major multigene superfamily of malaria parasites

Christoph S. Janssen, R. Stephen Phillips, C. Michael R. Turner, Michael P. Barrett

Institute of Biomedical and Life Sciences, Division of Infection & Immunity, IBLS, University of Glasgow, Glasgow G12 8QQ, UK.

MATERIALS AND METHODS

Protein detection and localisation

CIR protein was localized in *P. chabaudi* infected erythrocytes using indirect fluorescent antibody techniques (IFAT). Peptides were designed using expressed mRNA sequence (7) as a template and synthesised commercially (Genosphere Biotech.). 15mer peptides were designed to both variant and conserved regions, as determined from a CIR/ YIR alignment of 73 putative protein sequences. Sequences of the two peptides used successfully were as follows: cir_pep_2: CSQKASEFVKSFKEL (conserved) and cir_pep_1: CKINQHPNKKFGTND (variant). Rabbit antibodies to the peptides were commercially produced (Genosphere Biotech.) and used in binding assays on blood smears infected with the same stabilate of *P. chabaudi* that had provided the *cir* mRNA sequence. Thin blood smears were fixed in acetone for 15 min at room temperature, or for 4 h in 2% Formaldehyde/ PBS at room temperature. Smears were re-hydrated for 10 min in PBS before being blocked for 20 min with 0.01% Tween 20, 0.05% Sodium Azide and 1% BSA. Incubations with the rabbit anti-*cir*

RESULTS

Rodent malaria *cir/bir/yir* and *P. falciparum rif* introns

Table 1

Conserved sequence motifs in the <u>first</u> intron from rodent malaria *cir* homologues

MOTIF WIDTH BEST POSSIBLE MATCH

----- ----- -------------------

1	15	AATGTTTTCGTAGTG
2	15	TGGTATATGCACTTT
3	15	TATAAATATCATTAA
4	15	TAGTTTCCGTAACAT
5	15	CAATAATACGCGCTT
6	15	TTTCGGATGTTGCAT
7	15	ACTTTATTTTAATAA
8	11	TGTTTAATTCG

Conserved sequence motifs in the <u>second</u> intron from rodent malaria *cir* homologues

MOTIF WIDTH BEST POSSIBLE MATCH

----- ----- -------------------

1	9	AAGGTAAAT
2	9	TATTAGTAT
3	9	TAAGGAATT
4	9	TATGCAAAC
5	9	TACGCTTCT
6	9	TTTTCATTA
7	9	GTTAACAAA
8	9	TAACATTTT

1st *cir* introns:

 Based on conserved sequence motifs, the first intron sequences can be subdivided in into at least 10 conserved sequence families.

2nd *cir* introns:

 The presence and distribution of the consensus intron sequence motifs indicated that second intron sequences can be assigned to at least 7 conserved sub-families of intron sequences.

Figure 1

Maximum likelihood tree (dnaml) of an alignment of 503 *rif* and *cir/yir* 2nd introns

SEE SPEPARATE FILE "Supplementary Figure 1" FOR LARGER AND LEGIBLE IMAGE

Figure 1

Maximum likelihood tree (dnaml) of an alignment of 503 *rif* and *cir/yir* 2nd introns

key:
- yir
- cir
- rif
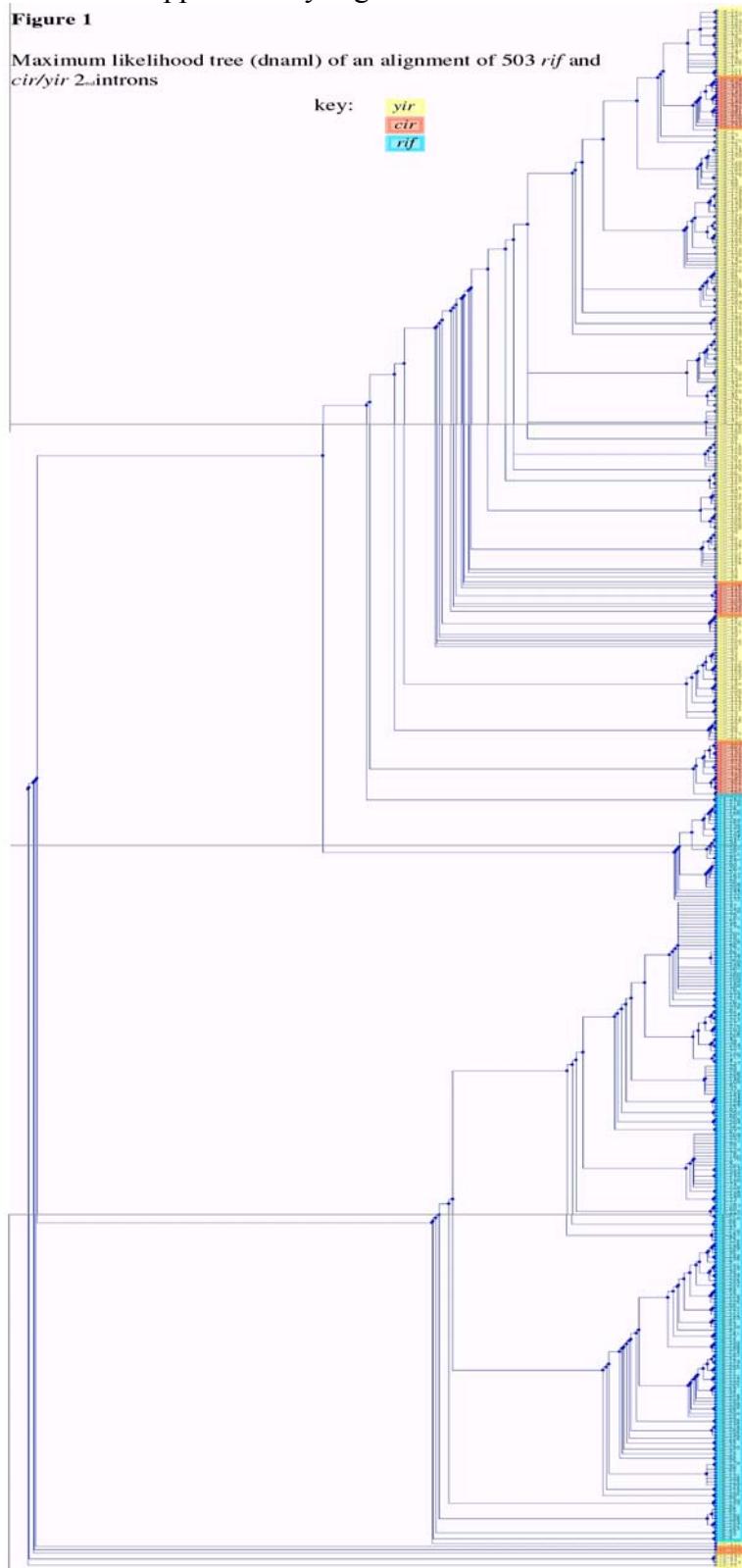
Tree -reconstruction analysis of an alignment of 503 *rif* and *cir/yir* 2nd introns shows 35 *rif* introns as more similar to the rodent malaria introns than those from other *rifs* (see main text).

Superfamily trees

A total of 157 amino acid sequences of members from *rif, vir, kir, yir* and *cir* were aligned using clustalW and dialign. The best global alignment (clustalW) was chosen for phylogenetic analysis. Poorly aligned positions and divergent regions of the alignment were eliminated and only conserved regions chosen (185 aa) (see main text). Thus, the variant domains of the proteins were eliminated from the analyses. Phylogenetic analysis was performed on the dataset using the maximum likelihood method applied to pairwise sequence distances calculated using quartet puzzling, which automatically assigns estimations of support to each internal branch (see main text) and using the maximum likelihood method implemented by the proml algorithm.

Figure 2

tree reconstruction using maximum likelihood method applied to pairwise sequence distances calculated using quartet puzzling
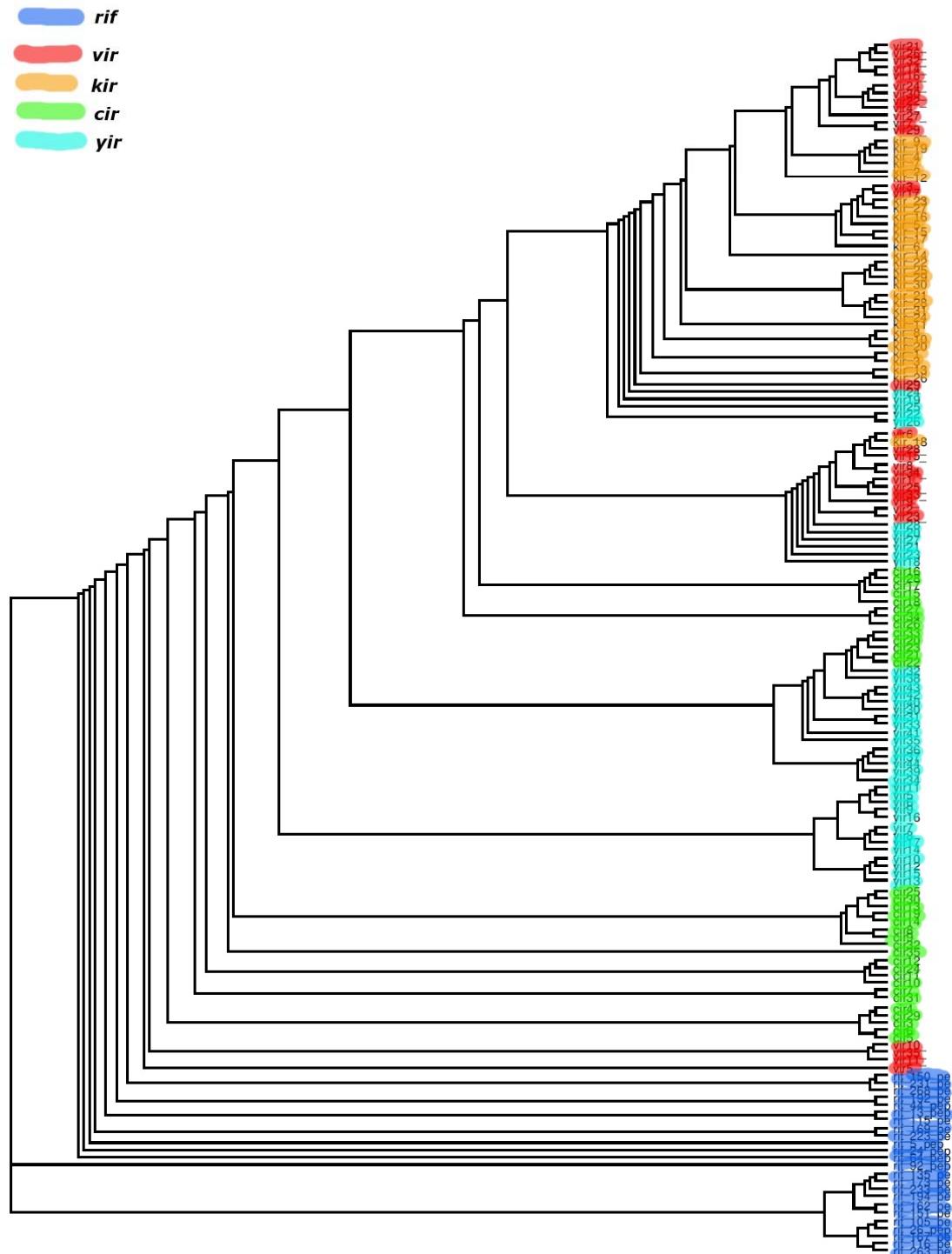
Figure 3

tree reconstruction using the maximum likelihood method implemented by the proml algorithm
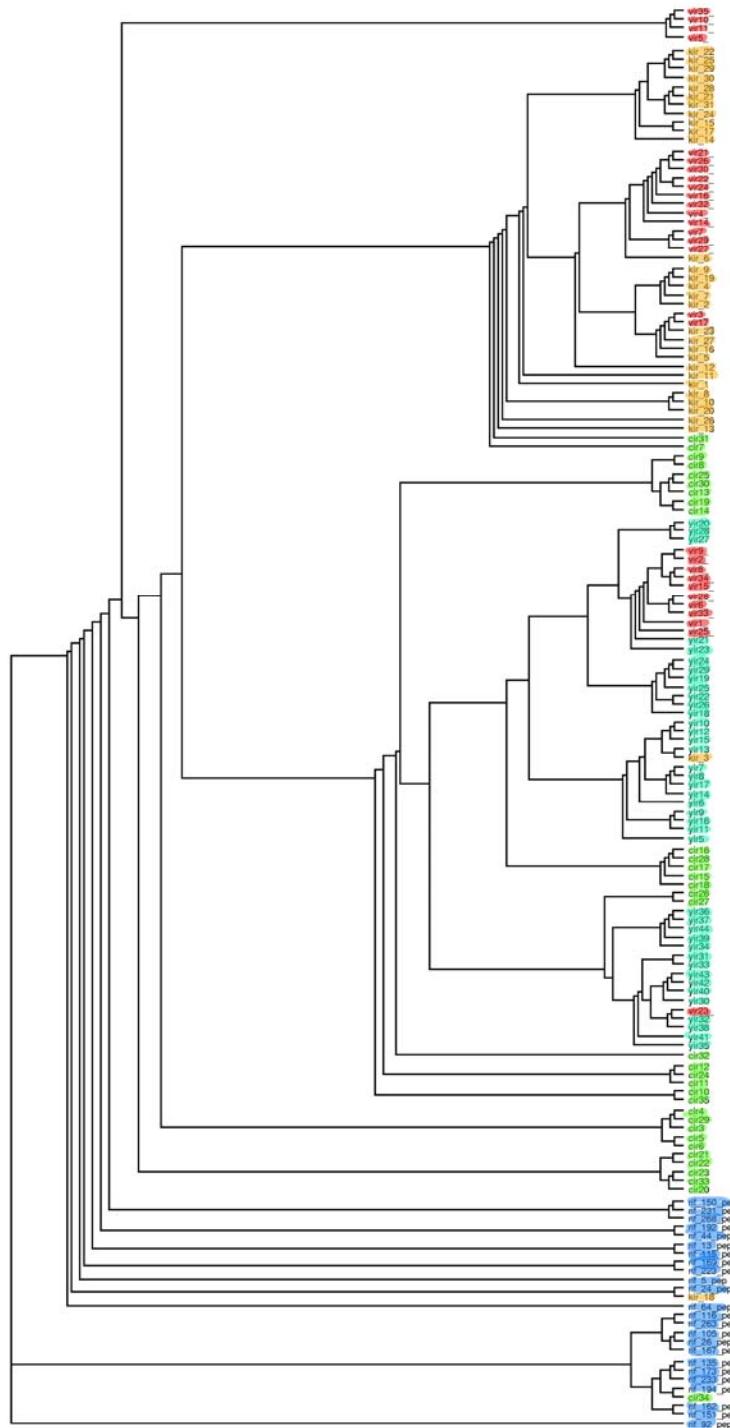
Protein detection and localisation

Localisation of CIR protein in the parasite and infected erythrocyte using anti-CIR peptide antibodies in immuno-fluorescence assays revealed that CIR proteins are exported from the parasite into the erythrocyte and appear to localize with the erythrocyte outer cell membrane. Late trophozoite stage parasite populations, which were previously shown to express specific *cir* mRNA sequences (7), could be labelled with antibodies to gene-specific peptide sequences (Figure 4). Not all parasites in a population sample showed binding of the anti-CIR antibodies; about 20% failed to label. Those parasites that did label showed an association of CIR proteins with the infected erythrocyte membrane (Figure 4) in the majority of cases. A minority of infected erythrocytes were labeled in the cytosol only. Uninfected erythrocytes failed to bind any anti-CIR antibodies and did not label. Both acetone and formaldehyde fixed smears showed binding of antibodies to infected erythrocytes. Formaldehyde fixed smears gave superior labelling of CIR protein associated with the erythrocyte membrane.

Figure 4. CIR protein labeling

Figure 4.1. *P. chabaudi* late trophozoites labelled with anti-*cir* antibodies.
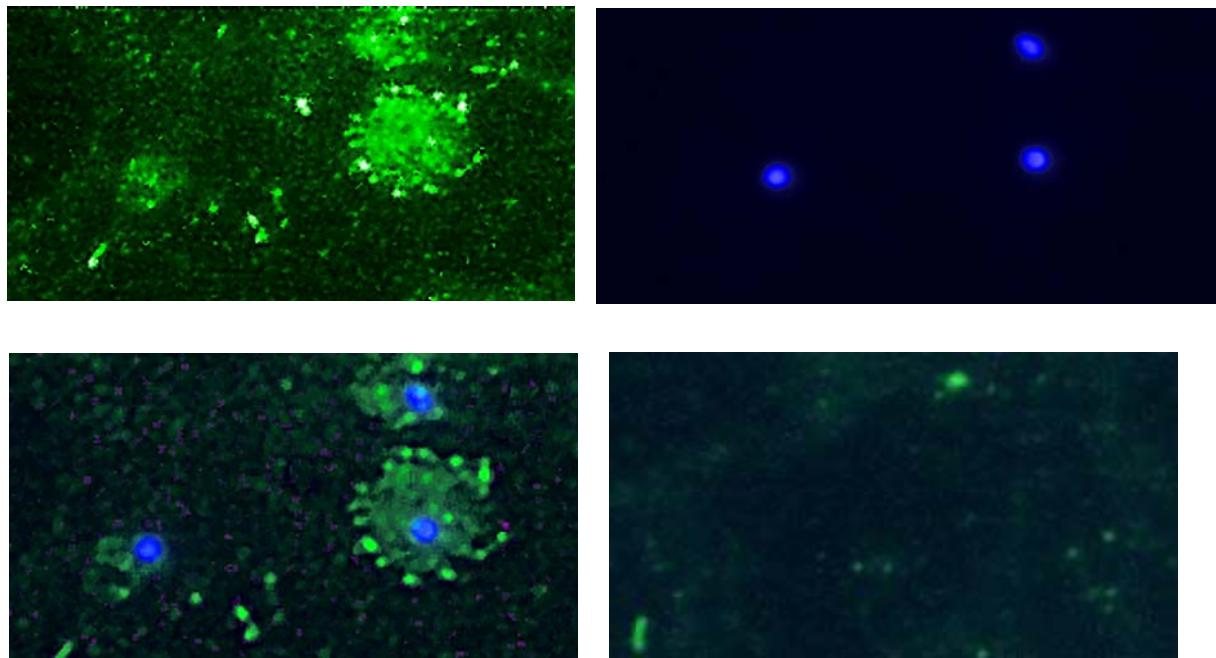
Figure 4.2. Western blot of *P. chabaudi* late trophozoite SDS-soluble protein fraction probed with anti-*cir* peptide antibodies.

—37.5 Kd

—28 Kd
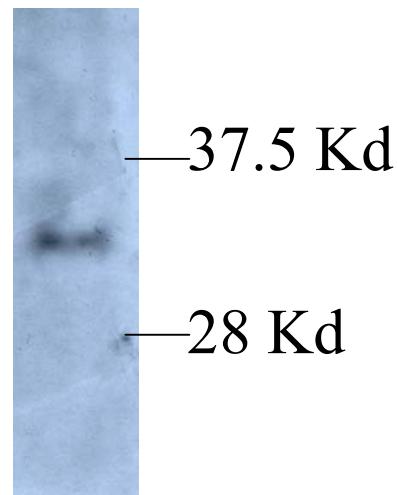
Figure 4. Sub-cellular localisation of CIR

4.1

Thin blood smears were fixed in 2% formaldehyde for 4 h prior to processing with standard IFAT techniques and DAPI.   Top left-hand panel - FITC labelled anti-*cir* peptide antibodies bound to cir proteins in/on *P. chabaudi* infected erythrocytes. Top right panel - The same parasites as depicted in top left stained with DAPI. Bottom left-hand panel DAPI and FITC stains merged. Bottom right panel FITC labelled pre-immune control serum.

4.2

Parasite extracts and erythrocyte ghosts were obtained by saponin lysis of *P. chabaudi* late trophozoite infected blood and subsequent centrifugation. The triton X-100 insoluble/ SDS soluble fraction obtained from the parasite/ ghost pellet was electrophoresed on a 5% - 20% gradient SDS-PAGE gel and blotted onto nitrocellulose. The blot was probed with anti-*cir* peptide antibodies using standard methods. Only one distinct labelled band was detected by using the Phototope_HRP detection system (Cell Signalling Tech.).

Superfamily sequence motifs

Neural network analysis reveals that at least four amino acid sequence motifs are conserved in all members of the PIR superfamily. The VIR and KIR members of the family share a further three amino acid motifs not found in other members. Interestingly the position of some motifs showed shifts in relative position in different members of the superfamily. For example motif 7 in the RIFIN sequences resides upstream of its relative position in the CIR family.

PROBILITY MATRICES AND HMMs

CIR and RIFIN motifs:

```
motif 2
letter-probability matrix: alength= 20 w= 12 n= 61247 E= 5.6e-561
 0.041253  0.000918  0.000890  0.005653  0.001162  0.024662  0.000593  0.002422
0.038952  0.054441  0.001231  0.002057  0.033664  0.000958  0.015391  0.646481
0.041035  0.021537  0.000280  0.066420
 0.003690  0.000924  0.001394  0.001609  0.013133  0.008112  0.000630  0.620597
0.002038  0.050449  0.020008  0.005026  0.000936  0.001286  0.001443  0.002668
0.014007  0.246446  0.000484  0.005120
 0.154238  0.000631  0.000316  0.000597  0.041920  0.005784  0.000442  0.216308
0.000705  0.140900  0.001643  0.010716  0.040955  0.000629  0.000695  0.052017
0.072842  0.247369  0.000291  0.011001
 0.554914  0.002384  0.001025  0.001569  0.057840  0.022159  0.000588  0.006524
0.005823  0.072827  0.001413  0.027246  0.000610  0.001035  0.018769  0.100994
0.029541  0.093685  0.000348  0.000705
 0.012605  0.000474  0.000216  0.000378  0.000821  0.051849  0.000268  0.610146
0.102628  0.068839  0.000936  0.000407  0.061671  0.000405  0.000447  0.001398
0.052504  0.033333  0.000170  0.000504
 0.017576  0.005681  0.000318  0.000601  0.127736  0.010823  0.000443  0.054324
0.000709  0.460520  0.001686  0.000595  0.035836  0.000635  0.000701  0.026666
0.012094  0.231782  0.000292  0.010983
 0.036192  0.010265  0.000490  0.000760  0.016053  0.010402  0.000472  0.491811
0.000920  0.055813  0.002403  0.000759  0.010135  0.000734  0.000815  0.011205
0.011870  0.337567  0.000322  0.001013
 0.047976  0.000630  0.000316  0.000597  0.036860  0.005784  0.000442  0.509809
0.000704  0.080174  0.001640  0.000595  0.167464  0.000628  0.000694  0.026717
0.012118  0.055078  0.040773  0.011003
 0.078910  0.001204  0.000848  0.001248  0.010508  0.010156  0.000703  0.199286
0.001181  0.095087  0.001690  0.000888  0.001004  0.000884  0.001301  0.011029
0.012887  0.570204  0.000281  0.000701
 0.006102  0.000524  0.000422  0.000849  0.011639  0.027349  0.000482  0.068264
0.000926  0.728868  0.004506  0.000560  0.076656  0.001073  0.001099  0.018936
0.014849  0.031256  0.004812  0.000829
 0.021557  0.000693  0.000542  0.005524  0.030033  0.005557  0.000482  0.429427
0.000985  0.178681  0.012060  0.010238  0.005286  0.000766  0.000852  0.039397
0.002371  0.244735  0.000331  0.010484
 0.053036  0.000631  0.000316  0.000597  0.036860  0.036145  0.000442  0.226429
0.005765  0.049818  0.441873  0.005656  0.000474  0.000629  0.005755  0.011536
0.072842  0.050024  0.000291  0.000881

motif 3

letter-probability matrix: alength= 20 w= 11 n= 61452 E= 7.1e-922
 0.000145  0.016872  0.000072  0.000080  0.007934  0.000106  0.000465  0.000218
```

```
 0.000110  0.000524  0.005739  0.000175  0.000080  0.000154  0.000151  0.000187
0.000142  0.000235  0.000457  0.966154
 0.391095  0.000680  0.000342  0.000646  0.045336  0.011722  0.000477  0.042400
0.017176  0.432193  0.007267  0.000642  0.000513  0.000681  0.000753  0.001524
0.002152  0.037665  0.000314  0.006422
 0.003859  0.000966  0.001458  0.001683  0.006647  0.001345  0.000660  0.769536
0.002132  0.056692  0.006651  0.001687  0.000979  0.001345  0.001508  0.002790
0.011166  0.122984  0.000508  0.005407
 0.001699  0.000546  0.000437  0.000880  0.002826  0.000566  0.000500  0.044514
0.000959  0.906235  0.028772  0.000580  0.000739  0.001113  0.001139  0.001106
0.001504  0.004662  0.000360  0.000866
 0.000167  0.000079  0.000087  0.000125  0.007815  0.000121  0.017257  0.000221
0.006016  0.006117  0.000146  0.000208  0.000088  0.000236  0.436768  0.005808
0.000170  0.000244  0.515161  0.003167
 0.000145  0.000079  0.000066  0.000076  0.080537  0.000096  0.000419  0.011598
0.000104  0.431977  0.000268  0.000159  0.000077  0.000149  0.000142  0.000168
0.000146  0.005927  0.000418  0.467449
 0.002075  0.077578  0.000667  0.000690  0.006842  0.039328  0.000486  0.030465
0.000798  0.014379  0.001290  0.105586  0.000435  0.006149  0.424549  0.271169
0.001715  0.003499  0.000272  0.012026
 0.001218  0.000275  0.035804  0.001583  0.006029  0.001381  0.034005  0.000773
0.001863  0.000928  0.000285  0.014235  0.000489  0.000988  0.462043  0.068758
0.001228  0.000633  0.000107  0.367375
 0.002780  0.000669  0.001104  0.008573  0.000467  0.001288  0.000884  0.018863
0.877792  0.001524  0.035300  0.015819  0.001007  0.001427  0.023319  0.001693
0.005348  0.001201  0.000353  0.000590
 0.003317  0.000996  0.000925  0.006840  0.001940  0.001261  0.000656  0.180262
0.390709  0.301266  0.018431  0.023336  0.000807  0.001056  0.017493  0.013265
0.008281  0.021758  0.000388  0.007013
 0.001838  0.000172  0.018058  0.002302  0.022329  0.001018  0.000859  0.011574
0.510101  0.012017  0.000457  0.265819  0.000584  0.013037  0.090830  0.035051
0.007019  0.000851  0.000121  0.005962
```

motif 4

letter-probability matrix: alength= 20 w= 12 n= 61247 E= 1.3e-616
```
 0.017590  0.498985  0.000441  0.007950  0.243945  0.001004  0.000620  0.047872
0.015276  0.005860  0.002250  0.000830  0.000655  0.000886  0.121980  0.009077
0.002750  0.013252  0.000405  0.008372
 0.001073  0.000120  0.767109  0.110904  0.000168  0.065737  0.022108  0.000199
0.008299  0.000323  0.000124  0.018087  0.000430  0.001035  0.000583  0.002057
0.001003  0.000311  0.000054  0.000276
 0.003611  0.000889  0.001185  0.008884  0.000608  0.001595  0.001100  0.002100
0.907254  0.001978  0.000975  0.009433  0.001309  0.005233  0.032463  0.012651
0.002424  0.001554  0.000468  0.004285
 0.010317  0.000214  0.223640  0.334869  0.014679  0.015651  0.001127  0.000796
0.018226  0.036892  0.000659  0.101800  0.000998  0.024490  0.186829  0.017330
0.009413  0.001275  0.000154  0.000643
 0.010425  0.000901  0.000520  0.000907  0.029879  0.001050  0.000631  0.653938
0.001081  0.132536  0.002637  0.007846  0.000695  0.000924  0.028786  0.085352
0.002917  0.030319  0.000420  0.008235
 0.005584  0.001123  0.020872  0.030863  0.001229  0.001782  0.070242  0.002509
0.046406  0.011054  0.003639  0.013537  0.007151  0.760884  0.010315  0.004474
0.003566  0.002737  0.000797  0.001237
 0.001205  0.000063  0.009614  0.108473  0.000097  0.000544  0.000477  0.029184
0.639195  0.000582  0.000247  0.029724  0.000404  0.023654  0.154095  0.000917
0.000766  0.000509  0.000057  0.000194
 0.003280  0.000848  0.007588  0.000892  0.002018  0.000987  0.007740  0.660084
0.008135  0.019986  0.002205  0.000835  0.007764  0.221589  0.000979  0.001925
0.002695  0.027369  0.000408  0.022673
 0.010532  0.000887  0.007558  0.022177  0.002021  0.001018  0.014844  0.595540
0.000994  0.020229  0.009418  0.015061  0.000667  0.000885  0.000978  0.009102
```

```
0.017033  0.027629  0.000409  0.243017
 0.020170  0.000703  0.012551  0.073050  0.015615  0.006724  0.018624  0.031650
0.007235  0.726168  0.012040  0.000751  0.000956  0.001440  0.013460  0.013416
0.001942  0.041930  0.000463  0.001112
 0.008364  0.000177  0.007869  0.001376  0.014516  0.000706  0.000753  0.014899
0.652475  0.001039  0.000433  0.015518  0.000404  0.001960  0.254435  0.008389
0.008233  0.007956  0.000124  0.000373
 0.001103  0.000088  0.627646  0.250606  0.000128  0.008478  0.007797  0.000179
0.008435  0.000308  0.007338  0.053530  0.000430  0.001217  0.000552  0.016234
0.000925  0.000309  0.000039  0.014657


motif 5

letter-probability matrix: alength= 20 w= 10 n= 61657 E= 5.2e-407
 0.008134  0.000699  0.000563  0.001132  0.039346  0.000727  0.000643  0.019548
0.007191  0.817166  0.011970  0.000746  0.000951  0.001432  0.001466  0.072896
0.001929  0.011895  0.000461  0.001106
 0.020048  0.000699  0.006519  0.001132  0.015522  0.000727  0.000643  0.091025
0.036973  0.745683  0.006013  0.000746  0.006907  0.001431  0.001466  0.007379
0.013842  0.041677  0.000460  0.001106
 0.003101  0.010866  0.001100  0.001321  0.851897  0.001216  0.000734  0.009495
0.001183  0.040932  0.007358  0.001062  0.001323  0.000811  0.000979  0.003075
0.001550  0.009118  0.001661  0.051218
 0.343007  0.029099  0.000457  0.085468  0.065458  0.001054  0.000622  0.019350
0.001009  0.041196  0.002285  0.050200  0.050020  0.000892  0.000986  0.171283
0.080365  0.020362  0.007457  0.029430
 0.052098  0.000866  0.000451  0.000858  0.183397  0.000987  0.007591  0.186729
0.001004  0.449875  0.009565  0.000824  0.000685  0.042751  0.014953  0.008902
0.002720  0.034115  0.000410  0.001219
 0.017474  0.000635  0.007624  0.001164  0.008176  0.001120  0.000613  0.085419
0.021855  0.134095  0.008291  0.014574  0.604752  0.001087  0.001109  0.064040
0.009221  0.017660  0.000275  0.000816
 0.002181  0.000699  0.006521  0.001132  0.081070  0.000727  0.000643  0.108929
0.001234  0.703900  0.006005  0.000746  0.012868  0.001430  0.007424  0.007382
0.019808  0.035733  0.000460  0.001106
 0.008289  0.000868  0.014635  0.001376  0.100985  0.002049  0.003794  0.019981
0.002244  0.008983  0.001453  0.737336  0.001188  0.002160  0.001705  0.061501
0.003717  0.014506  0.000583  0.012649
 0.013227  0.001122  0.001649  0.001919  0.016608  0.001550  0.000765  0.703732
0.011182  0.088225  0.007536  0.001923  0.001124  0.001543  0.001730  0.003209
0.048308  0.092014  0.000585  0.002052
 0.002132  0.000684  0.000550  0.001106  0.033511  0.102524  0.000633  0.019459
0.001206  0.811155  0.011875  0.000730  0.000929  0.001399  0.001432  0.001392
0.001887  0.005815  0.000456  0.001123


motif 7

letter-probability matrix: alength= 20 w= 10 n= 61657 E= 1.7e-497
 0.134744  0.000395  0.000122  0.000139  0.000114  0.014952  0.000078  0.000160
0.000115  0.021001  0.000120  0.000238  0.007235  0.000172  0.000161  0.811676
0.007904  0.000562  0.000036  0.000075
 0.000722  0.000244  0.000070  0.000102  0.000332  0.013894  0.000036  0.527850
0.000103  0.001799  0.028009  0.000097  0.000087  0.000070  0.000075  0.000198
0.021232  0.404933  0.000027  0.000119
 0.198366  0.000839  0.000420  0.007523  0.001911  0.007691  0.000587  0.166513
0.027852  0.120066  0.002181  0.000791  0.135203  0.000835  0.000924  0.001882
0.049757  0.275101  0.000386  0.001170
 0.534527  0.002267  0.000998  0.001580  0.055244  0.004261  0.000692  0.009496
0.007498  0.064099  0.001930  0.066674  0.000725  0.001151  0.025249  0.151404
0.015654  0.055136  0.000423  0.000992
 0.008200  0.000392  0.000158  0.000276  0.000731  0.000334  0.000175  0.613970
```

```
0.178341   0.043875   0.000899   0.000272   0.000224   0.000265   0.000294   0.000623
0.117491   0.032982   0.000118   0.000382
 0.000538   0.000193   0.000076   0.000139   0.007847   0.000117   0.000086   0.037436
0.000148   0.668103   0.001409   0.000105   0.000135   0.000189   0.000169   0.000191
0.000454   0.282317   0.000103   0.000243
 0.000622   0.000226   0.000060   0.000083   0.000297   0.000087   0.000021   0.679740
0.000080   0.008625   0.000428   0.000077   0.000070   0.000049   0.000052   0.000132
0.000513   0.308725   0.000018   0.000092
 0.010109   0.000390   0.000123   0.000140   0.000119   0.001091   0.000078   0.512343
0.000116   0.000251   0.000122   0.000237   0.464064   0.000171   0.000160   0.008804
0.000968   0.000598   0.000037   0.000080
 0.069150   0.001867   0.001421   0.001995   0.001880   0.012068   0.001035   0.197245
0.001782   0.017371   0.002133   0.001266   0.001573   0.001237   0.002012   0.002567
0.016589   0.665776   0.000327   0.000707
 0.002081   0.000665   0.000534   0.001074   0.009166   0.000694   0.000611   0.127662
0.001171   0.795798   0.005702   0.000709   0.000903   0.001358   0.001391   0.007091
0.036237   0.005652   0.000439   0.001063
```

motif 8

```
letter-probability matrix: alength= 20 w= 10 n= 61657 E= 3.7e-856
 0.012805   0.005800   0.005758   0.207482   0.001733   0.006048   0.000596   0.466060
0.011614   0.088472   0.006701   0.001304   0.000729   0.148273   0.000976   0.002094
0.012343   0.014577   0.000347   0.006289
 0.000121   0.000070   0.000065   0.000071   0.017138   0.000092   0.000420   0.000194
0.000099   0.005473   0.000125   0.000158   0.045150   0.000139   0.000136   0.015190
0.000125   0.000209   0.005422   0.909602
 0.316610   0.000612   0.000307   0.010401   0.025949   0.010522   0.000429   0.141159
0.015415   0.402038   0.001599   0.005488   0.000461   0.000611   0.000675   0.001374
0.001940   0.043632   0.010103   0.010676
 0.010827   0.000269   0.000152   0.000242   0.060244   0.000286   0.000223   0.728470
0.000288   0.028211   0.001013   0.000255   0.000195   0.000262   0.000275   0.000543
0.010654   0.077709   0.079020   0.000861
 0.005940   0.000510   0.000411   0.000827   0.002636   0.000531   0.000469   0.009925
0.000902   0.931725   0.021791   0.000545   0.000694   0.001045   0.001071   0.001038
0.001409   0.008686   0.004686   0.005158
 0.002240   0.000580   0.000301   0.005486   0.011244   0.049833   0.015181   0.008375
0.015467   0.028507   0.006422   0.000565   0.000441   0.000605   0.413643   0.006232
0.001842   0.009016   0.359162   0.064857
 0.000452   0.000147   0.000167   0.000219   0.117454   0.000198   0.005697   0.030273
0.025086   0.308926   0.000268   0.000264   0.000159   0.000236   0.054929   0.000416
0.000315   0.015437   0.010425   0.428933
 0.051557   0.054455   0.000672   0.001230   0.006159   0.030245   0.005481   0.008090
0.011162   0.023313   0.006352   0.089118   0.000563   0.005954   0.476376   0.212462
0.002065   0.008739   0.000267   0.005742
 0.002628   0.010539   0.005624   0.001035   0.016331   0.005868   0.025115   0.003846
0.006262   0.009238   0.001601   0.005993   0.000604   0.000855   0.491833   0.055807
0.002190   0.009185   0.000310   0.345136
 0.002958   0.000729   0.000967   0.004379   0.000498   0.001305   0.000900   0.007488
0.909697   0.001621   0.015217   0.013478   0.001072   0.001396   0.020807   0.007459
0.007752   0.001274   0.000383   0.000620
```


PIR superfamily motifs:

motif 1

```
letter-probability matrix: alength= 20 w= 15 n= 38516 E= 1.1e-252
 0.054060   0.000634   0.223503   0.124795   0.001447   0.095285   0.025996   0.036601
0.056582   0.015390   0.013321   0.120545   0.002475   0.041828   0.006666   0.100077
0.006284   0.014874   0.000457   0.059179
```

```
 0.054060   0.000634   0.177641   0.147726   0.012912   0.015029   0.025997   0.025135
0.297351   0.015388   0.001856   0.074685   0.013940   0.041829   0.041063   0.031287
0.017750   0.003407   0.000456   0.001853
 0.042513   0.035129   0.028363   0.009668   0.013057   0.049393   0.048893   0.048497
0.044677   0.210946   0.013447   0.017132   0.094264   0.018591   0.155670   0.019566
0.017679   0.003791   0.000470   0.128253
 0.194693   0.544700   0.000843   0.013391   0.038749   0.001730   0.024695   0.020502
0.001841   0.009856   0.003793   0.001487   0.012946   0.001595   0.013541   0.027034
0.016513   0.033935   0.012484   0.025673
 0.008200   0.000634   0.062990   0.193586   0.012912   0.015029   0.014532   0.036600
0.102444   0.026854   0.013321   0.109080   0.013940   0.190875   0.086923   0.042752
0.029215   0.003407   0.011922   0.024784
 0.036924   0.011386   0.003541   0.004877   0.244292   0.003302   0.012042   0.005721
0.004591   0.074027   0.003626   0.003685   0.002877   0.003246   0.013565   0.026284
0.004920   0.016105   0.012806   0.512183
 0.266873   0.001228   0.000605   0.001144   0.241472   0.013303   0.000854   0.055571
0.001349   0.296238   0.027430   0.001138   0.000920   0.001228   0.001339   0.002695
0.003851   0.032758   0.000587   0.049417
 0.005634   0.001376   0.000991   0.001869   0.003201   0.013562   0.012973   0.267943
0.061593   0.021105   0.015399   0.155157   0.001154   0.001857   0.013986   0.015305
0.004609   0.033299   0.000651   0.368333
 0.012879   0.000235   0.000258   0.000313   0.166272   0.000318   0.025961   0.000652
0.000372   0.321382   0.000416   0.000485   0.000254   0.000427   0.000456   0.025237
0.000449   0.000722   0.001096   0.441815
 0.003289   0.001358   0.002228   0.002480   0.011866   0.002664   0.006162   0.002579
0.002676   0.010904   0.011668   0.002318   0.001564   0.002467   0.003193   0.002965
0.002655   0.008581   0.891655   0.026727
 0.003705   0.001306   0.001378   0.002301   0.023902   0.001628   0.001163   0.102471
0.011739   0.678480   0.045362   0.001831   0.001719   0.002409   0.002545   0.002920
0.012494   0.099646   0.000774   0.002227
 0.030502   0.092986   0.017069   0.009117   0.001337   0.350820   0.002817   0.013625
0.021169   0.003610   0.001690   0.109818   0.002246   0.018203   0.017530   0.157871
0.005731   0.003162   0.011966   0.128730
 0.008132   0.000636   0.166139   0.078822   0.024546   0.060869   0.048938   0.013663
0.056453   0.003913   0.013309   0.109062   0.002457   0.030276   0.006596   0.088548
0.006229   0.003391   0.000467   0.277557
 0.031077   0.000652   0.005544   0.147626   0.001486   0.003516   0.003017   0.105619
0.354760   0.061424   0.047795   0.028733   0.002440   0.099088   0.029509   0.019704
0.029187   0.015032   0.000461   0.013329
 0.003868   0.001157   0.000825   0.011327   0.034026   0.001241   0.000987   0.243432
0.021194   0.543331   0.036908   0.010881   0.001360   0.002000   0.011766   0.002429
0.003339   0.048096   0.000693   0.021143
```

motif 2

```
letter-probability matrix: alength= 20 w= 10 n= 39116 E= 1.1e-215
 0.046582   0.000547   0.113554   0.087774   0.011127   0.012949   0.012521   0.090818
0.236465   0.003382   0.001600   0.084113   0.012012   0.026162   0.035382   0.017078
0.015294   0.190649   0.000393   0.001597
 0.096134   0.021542   0.000640   0.001208   0.033299   0.011592   0.031306   0.169799
0.011564   0.231624   0.013428   0.021478   0.071926   0.011406   0.092652   0.002844
0.004009   0.079948   0.000583   0.093018
 0.156971   0.305451   0.000633   0.011337   0.084000   0.001451   0.021165   0.017721
0.001414   0.180942   0.013431   0.001194   0.000950   0.001260   0.011533   0.033256
0.014147   0.110378   0.010722   0.022045
 0.007062   0.000547   0.054274   0.147047   0.021008   0.012948   0.042159   0.239021
0.048748   0.013267   0.001602   0.074232   0.012011   0.174354   0.055138   0.026955
0.005414   0.022703   0.010273   0.031236
 0.032909   0.001890   0.002854   0.003954   0.201291   0.002728   0.017941   0.146653
0.003746   0.058750   0.011471   0.003015   0.002353   0.002672   0.003627   0.006582
0.004205   0.031204   0.011303   0.450851
 0.241882   0.000283   0.000104   0.000199   0.169576   0.000234   0.000171   0.098103
```

```
0.000228  0.418465  0.023328  0.000185  0.000206  0.000306  0.000264  0.000399
0.000740  0.013186  0.000213  0.031926
 0.004852  0.001256  0.000643  0.001213  0.003019  0.011582  0.011064  0.230437
0.011629  0.038891  0.053929  0.072139  0.000951  0.001279  0.011585  0.002838
0.003976  0.039280  0.000591  0.498848
 0.035147  0.001215  0.000610  0.001150  0.114648  0.001393  0.021190  0.251094
0.001358  0.272453  0.003185  0.001151  0.000914  0.001214  0.001341  0.023037
0.003841  0.059425  0.000584  0.205051
 0.000655  0.000283  0.000458  0.000483  0.002806  0.000595  0.000438  0.000531
0.000531  0.002107  0.000472  0.000488  0.000354  0.000504  0.000645  0.000614
0.000541  0.000783  0.965197  0.021513
 0.003279  0.001158  0.001223  0.002044  0.029215  0.001442  0.001032  0.050844
0.002404  0.794909  0.024238  0.001624  0.001528  0.002142  0.002262  0.002587
0.003067  0.072340  0.000687  0.001975
```

MOTIFS CONVERTED TO HMM

```
hmm 1

HMMER2.0  [2.3.1]
NAME  motif6_cfh_m12
LENG  12
ALPH  Amino
RF    no
CS    no
MAP   yes
COM   hmmbuild -s m6_cfh_m12.hmm motif6_cfh_m12.aln
COM   hmmcalibrate m6_cfh_m12.hmm
NSEQ  16
DATE  Thu May 13 11:51:30 2004
CKSUM 7398
XT       -8455     -4      0       *  -8455     -4      0       *
NULT      -4  -8455
NULE     595  -1558     85    338   -294    453  -1158    197    249    902  -
1085   -142    -21   -313     45    531    201    384  -1998   -644
EVD   -5.874833   0.618609
HMM        A      C      D      E      F      G      H      I      K      L
M      N      P      Q      R      S      T      V      W      Y
        m->m    m->i    m->d    i->m    i->i    d->m    d->d    b->m    m->e
        -322       *   -2322
     1  -1041  -1794   -826  -1082  -3355   3103  -1745  -3348  -1830  -3441  -
2683   1798  -2250  -1530  -2196  -1133  -1366  -2544  -3384  -2934      1
     -    -149   -500    233     43   -381    399    106   -626    210   -466   -
720    275    394     45     96    359    117   -369   -294   -249
     -     -83  -7175  -8217   -894  -1115   -701  -1378  -1322  -4462
     2   -462  -1222   -713    981   1607  -1568   -240    757    722  -1131  -
373   -370  -1648     78   1120   -525   -396   -723  -1533  -1028      2
     -    -149   -500    233     43   -381    399    106   -626    210   -466   -
720    275    394     45     96    359    117   -369   -294   -249
     -     -86  -7178  -8220   -894  -1115   -701  -1378  -4781  -4395
     3  -1829  -3086   3651   -300  -3868    567  -1408  -3890  -1720  -3814  -
3189   -655  -2396  -1128  -2496  -1616  -1989  -3297  -3742  -3054      3
     -    -149   -500    233     43   -381    399    106   -626    210   -466   -
720    275    394     45     96    359    117   -369   -294   -249
     -     -89  -7181  -8224   -894  -1115   -701  -1378  -4781  -4325
     4  -2057  -2766  -2566  -1374  -3580  -2594   -560  -2946   3257  -2621  -
1947  -1295  -2559   -143   1969  -1941  -1765  -2707  -2470  -2349      4
     -    -149   -500    233     43   -381    399    106   -626    210   -466   -
720    275    394     45     96    359    117   -369   -294   -249
     -     -93  -7185  -8227   -894  -1115   -701  -1378  -4781  -4251
     5   -573  -1010  -1167   -589   -824  -1755   -403   -648   -204    422  -
193   -707  -1833   -255   1612    712   -509   -497  -1195   2104      5
     -    -149   -500    233     43   -381    399    106   -626    210   -466   -
```

```
720     275     394      45      96     359     117    -369    -294    -249
     -     -98   -7190   -8232    -894   -1115    -701   -1378   -4781   -4174
     6     -38    -399   -2114   -1521     908   -1899    -704      91   -1237    1462
408   -1303   -1957    -983    -385     179    -385     179    -788    1613       6
     -    -149    -500     233      43    -381     399     106    -626     210    -466     -
720     275     394      45      96     359     117    -369    -294    -249
     -    -103   -7196   -8238    -894   -1115    -701   -1378   -4781   -4092
     7   -1209   -2368   -1030    -440   -2813   -1971    2291   -2448    1988   -2287     -
1493    -623   -2040    2822     460   -1086   -1071   -2103   -2297   -1868       7
     -    -149    -500     233      43    -381     399     106    -626     210    -466     -
720     275     394      45      96     359     117    -369    -294    -249
     -    -108   -7201   -8243    -894   -1115    -701   -1378   -4781   -4005
     8    -498   -1889    1030     196   -2208   -1452    2259    -863    1211   -1883     -
987     -96   -1543     371    1692    -387    -429   -1513   -2051   -1415       8
     -    -149    -500     233      43    -381     399     106    -626     210    -466     -
720     275     394      45      96     359     117    -369    -294    -249
     -    -115   -7207   -8249    -894   -1115    -701   -1378   -4781   -3913
     9   -1001   -1124   -2000   -1465    -760   -2337   -1130      15   -1077    1307
186   -1492   -2374    2645   -1252   -1409    -948    1218   -1529   -1165       9
     -    -149    -500     233      43    -381     399     106    -626     210    -466     -
720     275     394      45      96     359     117    -369    -294    -249
     -    -122   -7215   -8257    -894   -1115    -701   -1378   -4781   -3814
    10   -1846   -2272   -1848   -1420    -410   -2534    -625   -2175    1966   -2054     -
1576   -1395   -2652    -798    -345   -1827   -1747   -2040    -930    3834      10
     -    -149    -500     233      43    -381     399     106    -626     210    -466     -
720     275     394      45      96     359     117    -369    -294    -249
     -    -131   -7223   -8265    -894   -1115    -701   -1378   -4781   -3708
    11    -478    -750   -1329    -748    -746    -253    1677     229    -421    1566
 74    -814   -1823    -394     492    -763    -414    -198   -1134    -716      11
     -    -149    -500     233      43    -381     399     106    -626     210    -466     -
720     275     394      45      96     359     117    -369    -294    -249
     -    -140   -7233   -8275    -894   -1115    -701   -1378   -4781   -3594
    12   -2126   -2788   -2709   -1452   -3623   -2637    -579   -2974    1876   -2641     -
1979   -1347   -2598    -164    3439   -2013   -1821   -2746   -2481   -2380      12
     -       *       *       *       *       *       *       *       *       *       *
  *       *       *       *       *       *       *       *       *       *
     -       *       *       *       *       *       *       *   -4781       0
//


hmm 2

HMMER2.0  [2.3.1]
NAME  motifs_crh_m12b
LENG  25
ALPH  Amino
RF    no
CS    no
MAP   yes
COM   hmmbuild -f m_cfh_m12_fragb.hmm motifs_crh_m12b.aln
COM   hmmcalibrate m_cfh_m12_fragb.hmm
NSEQ  16
DATE  Thu May 13 15:55:16 2004
CKSUM 4635
XT      -8455      -4   -1000   -1000   -8455      -4   -8455      -4
NULT      -4   -8455
NULE     595   -1558      85     338    -294     453   -1158     197     249     902     -
1085    -142     -21    -313      45     531     201     384   -1998    -644
EVD   -7.945719   0.645052
HMM         A       C       D       E       F       G       H       I       K       L
M           N       P       Q       R       S       T       V       W       Y
        m->m    m->i    m->d    i->m    i->i    d->m    d->d    b->m    m->e
        -415       *   -2000
     1    -590   -1320    -549    -756   -2777    3060   -1315   -2721   -1334   -2854     -
```

```
2097    906  -1837  -1103  -1662   -698   -904  -1991  -2850  -2398     1
   -   -149   -500    233     43   -381    399    106   -626    210   -466    -
720    275    394     45     96    359    117   -369   -294   -249
   -    -52  -6703  -7745   -894  -1115   -701  -1378  -1415  -5586
   2   -290   -952   -632    454   1314  -1423   -100    661    920   -850    -
109   -262  -1500    187   1053   -389   -224   -440  -1279   -794     2
   -   -149   -500    233     43   -381    399    106   -626    210   -466    -
720    275    394     45     96    359    117   -369   -294   -249
   -    -52  -6703  -7745   -894  -1115   -701  -1378  -6000  -5555
   3  -1362  -2623   3511    107  -3304     99   -943  -3268  -1172  -3232    -
2575   -245  -1973   -650  -1899  -1160  -1498  -2726  -3219  -2522     3
   -   -149   -500    233     43   -381    399    106   -626    210   -466    -
720    275    394     45     96    359    117   -369   -294   -249
   -    -53  -6704  -7746   -894  -1115   -701  -1378  -6000  -5525
   4  -1525  -2322  -1764   -851  -3005  -2129   -254  -2471   3179  -2226    -
1532   -853  -2146    158   1509  -1424  -1300  -2209  -2135  -1932     4
   -   -149   -500    233     43   -381    399    106   -626    210   -466    -
720    275    394     45     96    359    117   -369   -294   -249
   -    -54  -6705  -7747   -894  -1115   -701  -1378  -6000  -5493
   5   -394   -797  -1024   -446   -558  -1584   -217   -421    -38    360
22   -549  -1660    -96   1384    100   -329   -278   -948   2166     5
   -   -149   -500    233     43   -381    399    106   -626    210   -466    -
720    275    394     45     96    359    117   -369   -294   -249
   -    -54  -6705  -7747   -894  -1115   -701  -1378  -6000  -5459
   6     41   -236  -1836  -1247    963  -1698   -496    256   -974   1435
578  -1066  -1756   -738   -377   -117   -198    343   -622    799     6
   -   -149   -500    233     43   -381    399    106   -626    210   -466    -
720    275    394     45     96    359    117   -369   -294   -249
   -    -55  -6706  -7749   -894  -1115   -701  -1378  -6000  -5427
   7   -830  -2022   -652    -83  -2422  -1635   2187  -2077   2052  -1954    -
1144   -289  -1712   2231    622   -714   -709  -1725  -2005  -1534     7
   -   -149   -500    233     43   -381    399    106   -626    210   -466    -
720    275    394     45     96    359    117   -369   -294   -249
   -    -56  -6707  -7749   -894  -1115   -701  -1378  -6000  -5393
   8   -338  -1661    423    330  -1957  -1299   2341   -744   1064  -1644    -
766     45  -1389    519   1575   -240   -267  -1273  -1833  -1217     8
   -   -149   -500    233     43   -381    399    106   -626    210   -466    -
720    275    394     45     96    359    117   -369   -294   -249
   -    -57  -6708  -7750   -894  -1115   -701  -1378  -6000  -5359
   9   -706   -988  -1420   -892   -615  -1947   -700     11   -518   1358
326   -989  -2002   2370   -747  -1000   -653    561  -1322   -923     9
   -   -149   -500    233     43   -381    399    106   -626    210   -466    -
720    275    394     45     96    359    117   -369   -294   -249
   -    -58  -6709  -7751   -894  -1115   -701  -1378  -6000  -5323
  10  -1167  -1745  -1202   -737   -398  -1975   -251  -1578   2053  -1562    -
989   -805  -2086   -233     74  -1170  -1076  -1412   -847   3263    10
   -   -149   -500    233     43   -381    399    106   -626    210   -466    -
720    275    394     45     96    359    117   -369   -294   -249
   -    -59  -6710  -7752   -894  -1115   -701  -1378  -6000  -5286
  11   -276   -695   -897   -327   -720     -7   1540    214    -63   1007
132   -458  -1570    -34    595   -484   -214   -159  -1063   -618    11
   -   -149   -500    233     43   -381    399    106   -626    210   -466    -
720    275    394     45     96    359    117   -369   -294   -249
   -    -60  -6711  -7753   -894  -1115   -701  -1378  -6000  -5249
  12  -1631  -2342  -1987   -990  -3049  -2201   -285  -2498   1498  -2247    -
1573   -946  -2211    121   3372  -1541  -1389  -2256  -2141  -1974    12
   -   -149   -500    233     43   -381    399    106   -626    210   -466    -
720    275    394     45     96    359    117   -369   -294   -249
   -    -61  -6712  -7754   -894  -1115   -701  -1378  -6000  -5210
  13     29    276    -40    359   -411   -792    613   -320    465   -632
485    255   -679    621    145     -1    148   -194    455    -83    13
   -   -149   -491    232     43   -379    397    109   -625    210   -467    -
714    275    393     46     96    358    117   -369   -284   -247
   -  -1877   -524  -7755  -3166   -170   -701  -1378  -6000  -5171
```

```
    14  -1243    -941   -3389   -2895    -784   -3084   -2211    2973   -2594     935
295  -2655   -2995   -2325   -2573   -2270    -324    1072   -1940   -1632      46
     -    -149    -500     233      43    -381     399     106    -626     210    -466    -
720    275     394      45      96     359     117    -369    -294    -249
     -     -63   -6714   -7756    -894   -1115    -701   -1378   -6000   -5130
    15   1139    -701   -1084    -615   -1368   -1145    -601    -208    -484   -1215    -
466   -604    2207    -375    -816     154     612    -619   -1659   -1220      47
     -    -149    -500     233      43    -381     399     106    -626     210    -466    -
720    275     394      45      96     359     117    -369    -294    -249
     -     -64   -6715   -7757    -894   -1115    -701   -1378   -6000   -5088
    16  -1403    -986   -3866   -3469   -1339   -3572   -3096    2660   -3293    -247    -
160  -3223   -3437   -3115   -3342   -2850   -1393    2622   -2715   -2300      48
     -    -149    -500     233      43    -381     399     106    -626     210    -466    -
720    275     394      45      96     359     117    -369    -294    -249
     -     -66   -6717   -7759    -894   -1115    -701   -1378   -6000   -5046
    17   -413    -254   -2506   -1904     872   -1982    -870    2167   -1571     454
580  -1536   -2018   -1251   -1480    -471     246    1142    -803    -452      49
     -    -149    -500     233      43    -381     399     106    -626     210    -466    -
720    275     394      45      96     359     117    -369    -294    -249
     -     -67   -6718   -7760    -894   -1115    -701   -1378   -6000   -5001
    18   -885    -777   -2601   -2091    1336   -2321   -1003     230   -1782    2211
594  -1792   -2345   -1444   -1742    -369    -845      63    -760    -157      50
     -    -149    -500     233      43    -381     399     106    -626     210    -466    -
720    275     394      45      96     359     117    -369    -294    -249
     -     -68   -6719   -7761    -894   -1115    -701   -1378   -6000   -4955
    19    269    -312   -1775   -1200    -340     451    -599    1558    -963     417
464  -1069   -1795    -748   -1073    -767     687     696    -829    -455      51
     -    -149    -500     233      43    -381     399     106    -626     210    -466    -
720    275     394      45      96     359     117    -369    -294    -249
     -     -70   -6722   -7764    -894   -1115    -701   -1378   -6000   -4909
    20  -1465   -1060   -3836   -3392   -1025   -3565   -2872    3173   -3169     419
127  -3172   -3378   -2891   -3165   -2814   -1439    1539   -2411   -2084      52
     -    -149    -500     233      43    -381     399     106    -626     210    -466    -
720    275     394      45      96     359     117    -369    -294    -249
     -     -72   -6723   -7765    -894   -1115    -701   -1378   -6000   -4860
    21    623    -336   -1906   -1333    -333   -1767    -682     815   -1081    1420
487  -1169   -1854    -854   -1164     316    -297     717    -861    -497      53
     -    -149    -500     233      43    -381     399     106    -626     210    -466    -
720    275     394      45      96     359     117    -369    -294    -249
     -     -74   -6725   -7767    -894   -1115    -701   -1378   -6000   -4809
    22  -2742   -2290   -3014   -3139     701   -3008    -615   -2258   -2836   -1879    -
1835  -2399   -3220   -2421   -2640   -2733   -2781   -2314      18    4561      54
     -    -149    -500     233      43    -381     399     106    -626     210    -466    -
720    275     394      45      96     359     117    -369    -294    -249
     -     -76   -6727   -7769    -894   -1115    -701   -1378   -6000   -4757
    23   -712   -1004   -1447    -832    -958   -1942    -583    1750    2053      83
-8   -917   -1986    -366    -290   -1001    -639     143   -1452   -1042      55
     -    -149    -500     233      43    -381     399     106    -626     210    -466    -
720    275     394      45      96     359     117    -369    -294    -249
     -     -78   -6729   -7771    -894   -1115    -701   -1378   -6000   -4702
    24   -897    -717   -2633   -2136     679   -2286    -387     582   -1789    -313
155  -1626   -2304   -1390   -1687   -1377    -832     623    3719    2930      56
     -    -149    -500     233      43    -381     399     106    -626     210    -466    -
720    275     394      45      96     359     117    -369    -294    -249
     -     -80   -6731   -7773    -894   -1115    -701   -1378   -6000   -4646
    25   -295    -473   -1325    -786    -335   -1598    -380     -44    -545    1511
322    293   -1693    -422    -748     478    -248      69    1222    -302      57
     -       *       *       *       *       *       *       *       *       *       *
*       *       *       *       *       *       *       *       *       *
     -       *       *       *       *       *       *       *  -6000       0
//
```