Figure S1

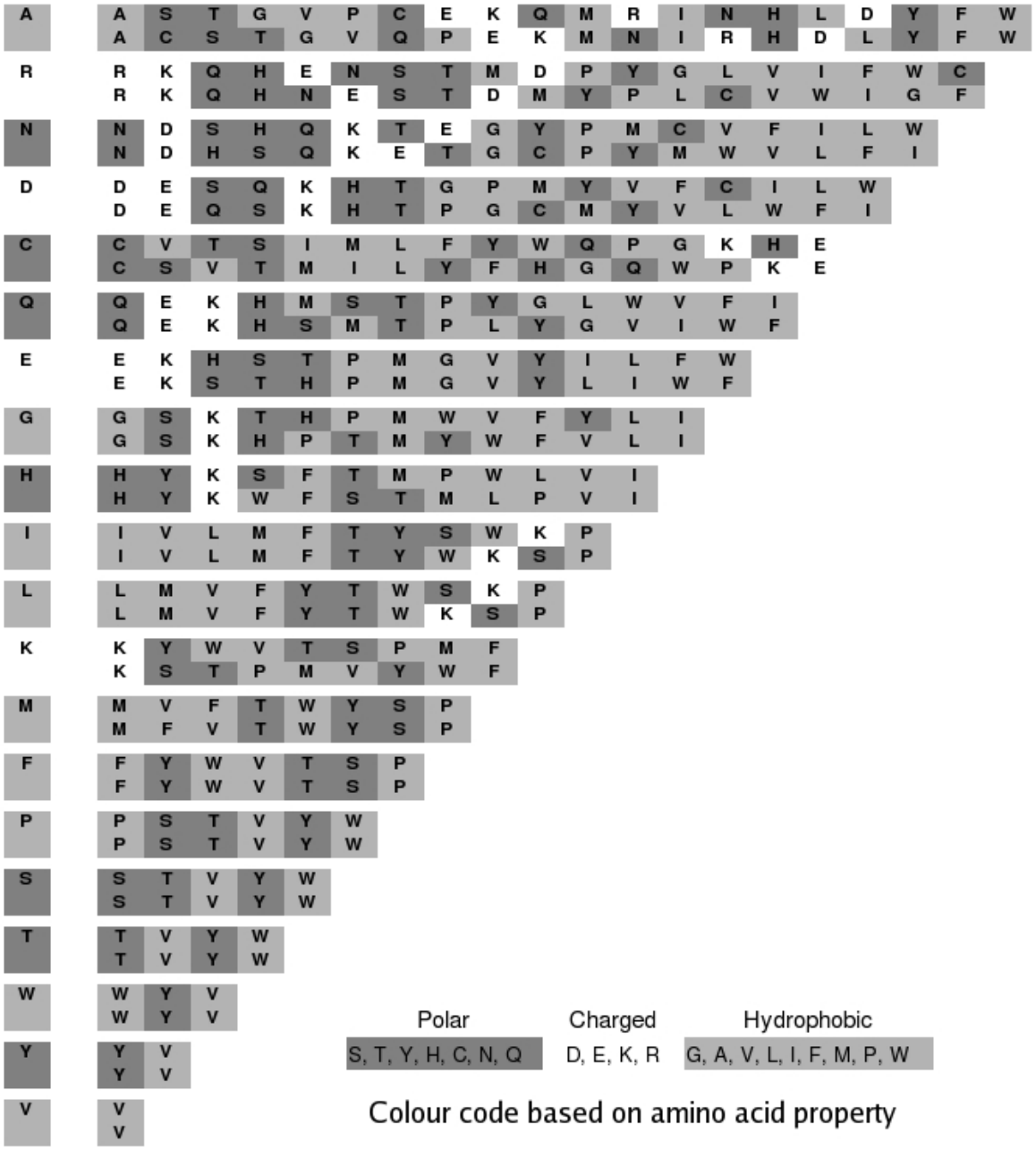


Figure S1. A half-diagonal representation of BLOSUM90 and Smat80 matrices showing differences in their substitution preferences The boxes in the left represent each of the 20 amino acids pairing with all other amino acids (adjacent rows) in the BLOSUM90 (odd rows of alphabets) and Smat80 (even rows of amino acid alphabets) matrices compared here. The amino acid pairs are arranged in a decreasing order of their lodscore values i.e. most preferred followed by the least preferred. Colour code is used to represent different classes of amino acids i.e. dark grey fill for polar amino acids, no fill for the charged amino acids and light grey fill for the hydrophobic amino acids.

Figure S2

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
	0	-2	-2	0	-2	-1	-2	-1	-1	0	-2	-1	0	-1	-1	0	-2	0	-2	0	C
		-1	-1	0	0	0	0	0	0	0	0	1	0	1	1	1	1	2	1	1	S
C	9		0	1	1	1	1	0	1	0	1	0	0	1	1	1	0	1	2	1	T
S	-2	5		0	1	1	0	0	0	1	1	1	0	2	1	1	1	1	1	1	P
T	-2	1	6		0	1	0	0	1	0	1	1	1	0	1	1	0	1	1	1	A
P	-4	-2	-2	8		-1	0	1	1	0	1	2	1	2	2	2	1	1	1	2	G
A	-1	1	0	-1	5		0	0	0	0	0	0	0	0	1	0	0	1	0	-1	N
G	-4	-1	-3	-3	0	6		0	-1	0	0	0	0	1	2	1	1	2	1	0	D
N	-4	0	0	-3	-2	-1	7		0	1	1	1	0	1	1	1	1	1	0	1	E
D	-5	-1	-2	-3	-3	-2	1	7		0	1	1	0	1	0	0	0	0	0	1	Q
E	-6	-1	-1	-2	-1	-3	-1	1	6		-1	1	0	0	1	0	1	0	0	-2	H
Q	-4	-1	-1	-2	-1	-3	0	-1	2	7		-1	0	1	1	1	1	1	0	0	R
H	-5	-2	-2	-3	-2	-3	0	-2	-1	1	8		0	1	0	1	1	1	1	0	K
R	-5	-1	-2	-3	-2	-3	-1	-3	-1	1	0	6		-1	0	0	0	-1	0	0	M
K	-4	-1	-1	-2	-1	-2	0	-1	0	1	-1	2	6		0	0	0	0	1	0	I
M	-2	-2	-1	-3	-2	-4	-3	-4	-3	0	-3	-2	-2	7		0	0	0	0	0	L
I	-2	-3	-1	-4	-2	-5	-4	-5	-4	-4	-4	-4	-4	1	5		0	0	0	1	V
L	-2	-3	-2	-4	-2	-5	-4	-5	-4	-3	-4	-3	-3	2	1	5		0	0	-1	F
V	-2	-2	-1	-3	-1	-5	-4	-5	-3	-3	-4	-3	-3	0	3	0	5		0	0	Y
F	-3	-3	-3	-4	-3	-5	-4	-5	-5	-4	-2	-4	-4	-1	-1	0	-2	7		-1	W
Y	-4	-3	-2	-4	-3	-5	-3	-4	-4	-3	1	-3	-3	-2	-2	-2	-3	3	8		
W	-4	-4	-4	-5	-4	-4	-5	-6	-5	-3	-3	-4	-5	-2	-4	-3	-3	0	2	11	
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	

Figure S2. The difference in the lodscore values of BLOSUM90 and Smat80 matrices. The lower half-diagonal represents the BLOSUM90 lodscore values and the upper half-diagonal shows the difference in the lodscore values of BLOSUM90 & Smat80 (i.e. BLOSUM90-Smat80). The half-bit values have been considered here. Comparison is made with BLOSUM90 as the entropy and scaling is similar to Smat80.

Figure S3

a)

```
>>gi|19703667|ref|NP_603229.1| fructose-bisphosphate ald (295 aa)
initn: 153 init1:  90 opt: 124  Z-score: 284.0  bits: 60.9 E(): 5.2e-14
Smith-Waterman score: 124; 26.708% identity (59.006% similar) in 161 aa overlap
(79-235:67-216)


       50        60        70        80        90       100
P14223 RFDNIKLENTIENRASYRDLLFGTKGLGKFISGAILFEETLFQKNEAGVPMVNLLHNENI
                                    : ::::::.:. .: ..      : .....
gi|197 NEYSNDKEMFDLIHKMRTRIIKSPAFNESKILGAILFEQTMDSKIDGKYTADFLWEEKKV
         40        50        60        70        80        90

      110       120       130       140       150       160
P14223 IPGIKVDKGLVNIPCTDEE--KSTQGLDGLAERCKEYYKAGARFAKWRTVLVIDTAKGKP
       .: .:.:::: ..     .  : .  :  : .: .: .  :.   : :.:.     :..:
gi|197 LPFLKIDKGLNDLDADGVQTMKPNPTLADLLKRANERHIFGT---KMRSVI----KKASP
        100       110       120       130          140

        170       180       190       200       210       220
P14223 TDLS-IHETAWGLARYASICQQNRLVPIVEPEILADGPHSIEVCAVVTQKVLSCVFKALQ
       . .. . :  . .:  :..   . :.::.:::.  ..  ... :  . .  .   ..::
gi|197 AGIARVVEQQFEVA--AQVVAAG-LIPIIEPEVDINNVDKVQ-CEEILRDEIRKHLNALP
     150       160       170       180       190       200

          230       240       250       260       270       280
P14223 E-NGVLLEGALLKPNMVTAGYECTAKTTTQDVGFLTVRTLRRTVPPALPGVVFLSGGQSE
       : ..:.:. .:
gi|197 ETSNVMLKLTLPTVENLYEEFTKHPRVVRVVALSGGYSREKANDILSKNKGVIASFSRAL
         210       220       230       240       250       260
```

b)

```
>>gi|19703667|ref|NP_603229.1| fructose-bisphosphate ald (295 aa)
initn: 526 init1: 163 opt: 489  Z-score: 1438.2  bits: 274.4 E(): 2.7e-78
Smith-Waterman score: 591; 27.077% identity (64.308% similar) in 325 aa overlap
(27-333:3-286)

               10        20          30         40        50
P14223 MAHCTEYMNAPKKLPADVAEELATTAQKL--VQAGKGILAA-DESTQTIKKRFDNIKL-E
                                .::  .. :::..:: :.:  ...: ..   . :
gi|197                        MNEKLEKMRNGKGFIAALDQSGGSTPKALKLYGVNE
                                      10        20        30

         60          70              80        90       100
P14223 NTIENRASYRDLL--FGTKGL--GKF----ISGAILFEETLFQKNEAGVPMVNLLHNEN-
       :. .:  .. ::.  . :. .   .:    : ::::::.:. .: . :  ....:..:.
gi|197 NEYSNDKEMFDLIHKMRTRIIKSPAFNESKILGAILFEQTMDSKID-GKYTADFLWEEKK
         40        50        60        70        80         90

       110       120       130       140       150       160
P14223 IIPGIKVDKGLVNIPC--TDEEKSTQGLDGLAERCKEYYKAGARFAKWRTVLVIDTAKGK
       ..: .:.:::: .. .  ....:... :. : .:..:..  :.   :.:.:.     :..
gi|197 VLPFLKIDKGLNDLDADGVQTMKPNPTLADLLKRANERHIFGT---KMRSVIK----KAS
         100       110       120       130          140
```

```
         170        180       190       200       210       220
P14223 PTDLS-IHETAWGLARYASICQQNRLVPIVEPEILADGPHSIEVCAVVTQKVLSCVFKAL
       :. .. . :. . .:  :..... :.::.:::.  .  .....  ..... ... .::
gi|197 PAGIARVVEQQFEVA--AQVVAAG-LIPIIEPEVDINNVDKVQCEEILRDEIRKHL-NAL
      150       160         170        180       190       200

           230       240       250       260       270       280
P14223 QENG-VLLEGALLKPNMVTAGYECTAKTTTQDVGFLTVRTLRRTVPPALPGVVFLSGGQS
       .:.. :.:.  :  :. :.  ::  .: .         :..:         ::.:::: :
gi|197 PETSNVMLK--LTLPT-VENLYEEFTKHP---------RVVR---------VVALSGGYS
          210          220       230                          240

           290       300       310       320        330       340
P14223 EEEASVNLNSINALGPHPWALTFSYGRALQASVLNTWQGKKE-NVAKAREVLLQRAEANS
       .:.:. .. : :       .. .:..:::. . :.. : ..: : . : ..
gi|197 REKAN-DILSKNK------GVIASFSRALTEG-LSAQQTDEEFNKTLAASIDGIYEASVK
            250             260        270       280       290
```

Figure S3: Alignment of fructose-bisphosphate aldolase from AT-rich genomes of *P. falciparum* (P14223) and *Fusobacterium nucleatum* at a gap opening and extension penalties of 12 & 2 respectively. a) Alignment obtained with the best performing standard matrix, Blosum50 b) Alignment extension (highlighted in grey) obtained with PfFSmat60 that extends left and right of the alignment obtained with BLOSUM50, with a good conservation of residues. The alignment length improved 2 fold with an increase in similarity and identity. The query sequence overlap with Pam2 and BLOSUM100 for this example was 81-88 and 79-118 residues respectively.

Figure S4

a)

```
>>P38013|AHP1_YEAST Peroxiredoxin type-2 - Saccharomyces (176 aa)
initn: 213 init1: 183 opt: 222  Z-score: 593.9  bits: 116.8 E(): 2.8e-31
Smith-Waterman score: 222; 30.081% identity (67.480% similar) in 123 aa overlap
(101-218:46-164)

               80        90       100       110       120       130
MAL7P1 MIDVRNMNNISDTDGSPNDFTSIDTHELFNNKKILLISLPGAFTPTCSTKMIPGYEEEYD
                                       ::::.. . :.::.:::... :::: .  :
P38013 FQYIAISQSDADSESCKMPQTVEWSKLISENKKVIITGAPAAFSPTCTVSHIPGYINYLD
          20        30        40        50        60        70

              140       150       160          170         180
MAL7P1 YFIKENNFDDIYCITNNDIYVLKSWFKSMDIK---KIKYISDGNSSFTESMN--MLVDKS
        ..::.. :..  .: .. .. ..: ::. .:   .::. :: . .::.:..  . :  .
P38013 ELVKEKEVDQVIVVTVDNPFANQAWAKSLGVKDTTHIKFASDPGCAFTKSIGFELAVGDG
          80        90       100       110       120       130

         190       200       210       220       230       240
MAL7P1 NFFMGMRPWRFVAIVENNILVKMFQEKDKQHNIQTDPYDISTVNNVKEFLKNNQL
        .. :    :.. .:::.:..   .: .   ..
P38013 VYWSG----RWAMVVENGIVTYAAKETNPGTDVTVSSVESVLAHL
         140           150       160       170
```

b)

```
>>P38013|AHP1_YEAST Peroxiredoxin type-2 - Saccharomyces (176 aa)
initn: 518 init1: 315 opt: 536  Z-score: 1586.9  bits: 300.6 E(): 1.4e-86
Smith-Waterman score: 536; 28.261% identity (73.913% similar) in 184 aa overlap
(62-235:2-176)

              40        50        60        70         80        90
MAL7P1 IVSKRGNSKNRFSQKVYESKNIDLENDIKENDLIPNVKVMIDVR-NMNNISDTDGSPNDF
                                     .::. .   . : . .  .::..:.....
P38013                             MSDLVNKKFPAGDYKFQYIAISQSDADSESC
                                           10        20        30

                  100       110       120       130       140
MAL7P1 TSIDTHE----LFNNKKILLISLPGAFTPTCSTKMIPGYEEEYDYFIKENNFDDIYCITN
       .. .: :    . .:::.....:.::.:::... :::: . .: ..::...:.. ..:
P38013 KMPQTVEWSKLISENKKVIITGAPAAFSPTCTVSHIPGYINYLDELVKEKEVDQVIVVTV
              40        50        60        70        80        90

        150       160          170         180       190       200
MAL7P1 NDIYVLKSWFKSMDIKK---IKYISDGNSSFTESMN--MLVDKSNFFMGMRPWRFVAIVE
       .. .. ..:.::. .:.   ::..:: ...::.:..  ..: .. .. :    :....::
P38013 DNPFANQAWAKSLGVKDTTHIKFASDPGCAFTKSIGFELAVGDGVYWSG----RWAMVVE
             100       110       120       130       140

             210       220       230       240
MAL7P1 NNILVKMFQEKDKQHNIQTDPYDISTVNNVKEFLKNNQL
       :.:..     ..:..:  :: ...:.:..: ..:
P38013 NGIVTY----AAKETNPGTD-VTVSSVESVLAHL
       150           160        170
```

Figure S4: **Alignment of a putative peroxiredoxin of *P. falciparum* (MAL7P1.159) with the yeast peroxiredoxin at a gap opening and extension penalty of 12 & 2 respectively.** a) Alignment obtained with the BLOSUM50 matrix b) The alignment extension (highlighted regions) with PfFSmat60 that spans left and right of the alignment obtained with the standard. The overlap was 1.5 times above the standard. The query sequence overlap with the PAM2 and BLOSUM100 matrices was 112-117 and 101-178 residues respectively.

Figure S5

a)

```
>>B0RS18|B0RS18_XANCB Superoxidase dismutase - Xanthomon (203 aa)
 initn: 530 init1: 255 opt: 535  Z-score: 1583.7  bits: 299.9 E(): 2.1e-86
Smith-Waterman score: 535; 36.946% identity (73.892% similar) in 203 aa overlap
(1-196:1-202)

               10        20        30        40        50        60
PF08_0 MVITLPKLKYALNALSPHISEETLNFHYNKHHAGYVNKLNTLIKDTPFAEKSLLDIVKES
       :. :::.: :: .:: :.:. .:...:.:::  :.:..:. .. : .:.  . ..:..
B0RS18 MAYTLPQLPYAYDALEPNIDAQTMEIHHTKHHQTYINNVNAALEGTEYADLPVEELVSKL
               10        20        30        40        50        60
```

```
                    70        80        90       100       110
PF08_0 S-------GAIFNNAAQIWNHTFYWDSMGPDCGGEPHGEIKEKIQEDFGSFNNFKEQFSN
       .       : . ::..   ::...:  :.:. ::::.::. . :..:.:.:..::: :..
B0RS18 KSLPENLQGPVRNNGGGHANHSLFWTVMSPNGGGEPKGEVAKAIDKDIGGFEKFKEAFTK
               70        80        90       100       110       120

           120       130       140       150       160       170
PF08_0 ILCGHFGSGWGWLALNNNNKLVILQTHDAGNPIKDNTGIPILTCDIWEHAYYIDYRNDRA
         ..:::::.::... ..:.:. .: .  .:. ...  :::  :.:::::. :.: :
B0RS18 AALSRFGSGWAWLSVTPDKKVVVESTANQDSPLFEGN-TPILGLDVWEHAYYLKYQNRRP
              130       140       150       160       170

           180       190
PF08_0 SYVKAWWNLVNWNFANENLKKAMQK
       .:. :..:.:::. ...  . :.
B0RS18 DYIGAFYNVVNWDEVERRYHAAIA
     180       190       200
```

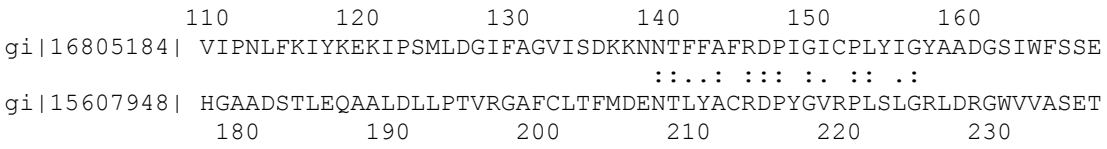
b)

```
>>B0RS18|B0RS18_XANCB Superoxidase dismutase - Xanthomon (203 aa)
 initn: 956 init1: 409 opt: 924  Z-score: 2813.8  bits: 527.5 E(): 6.3e-155
Smith-Waterman score: 924; 38.537% identity (83.902% similar) in 205 aa overlap
(1-197:1-203)

               10        20        30        40        50
PF08_0 MVITLPKLKYALNALSPHISEETLNFHYNKHHAGYVNKLNTLIKDTPFAEKSLLDIV-KE
       :. :::.:.:: .::.:.:...:...:.:::  :.:..:.... : .:. .. ..: :
B0RS18 MAYTLPQLPYAYDALEPNIDAQTMEIHHTKHHQTYINNVNAALEGTEYADLPVEELVSKL
               10        20        30        40        50        60

      60              70        80        90       100       110
PF08_0 SS------GAIFNNAAQIWNHTFYWDSMGPDCGGEPHGEIKEKIQEDFGSFNNFKEQFSN
       .:      :.. ::..   ::...:. :.:. ::::.::....:.:.:.:.::::.:..
B0RS18 KSLPENLQGPVRNNGGGHANHSLFWTVMSPNGGGEPKGEVAKAIDKDIGGFEKFKEAFTK
               70        80        90       100       110       120

           120       130       140        150       160       170
PF08_0 ILCGHFGSGWGWLALNNNNKLVILQTHDAGNPIKD-NTGIPILTCDIWEHAYYIDYRNDR
       .....:::::.::.....:.:. .:..  .:. . ::  ::: .:.::::::..:.:.:
B0RS18 AALSRFGSGWAWLSVTPDKKVVVESTANQDSPLFEGNT--PILGLDVWEHAYYLKYQNRR
              130       140       150         160       170

            180       190
PF08_0 ASYVKAWWNLVNWNFANENLKKAMQK
       ..:. :..:.:::. .... ..:..
B0RS18 PDYIGAFYNVVNWDEVERRYHAAIA
      180       190       200
```

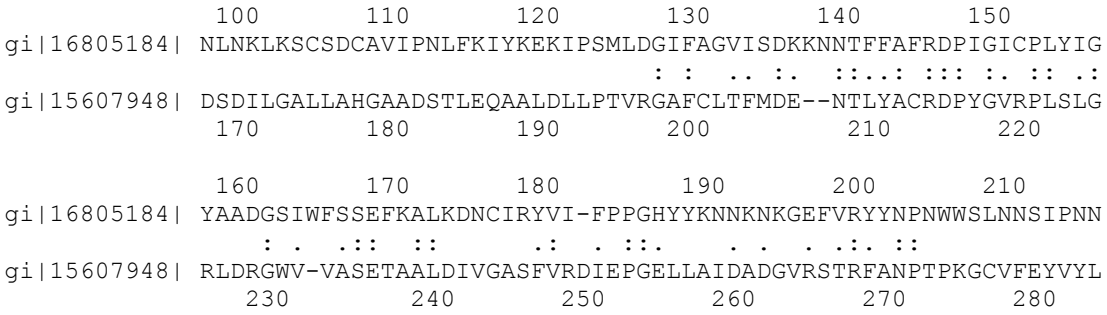
Figure S5: **Alignment of a typical Fe-SOD of *P. falciparum* with the *Xanthomonas campestris* ortholog at a gap opening and extension penalties of 12 & 2 respectively.** a) Alignment obtained with BLOSUM50 b) Alignment with PfFSmat60 with an improvement in identity, similarity and extension (1 residue only, highlighted in grey). The extra identities are shown in red text.

Figure S6

a)



b)



c)

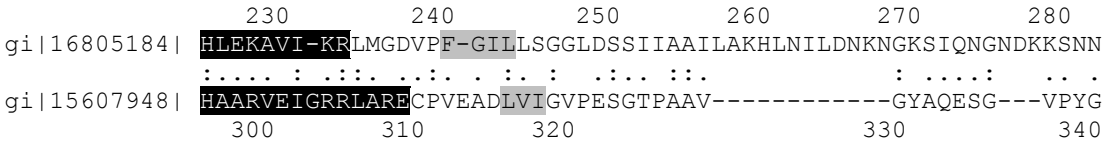


Figure S6. Alignment extension with PfFSmat60 for *P. falciparum* putative asparagine synthetase and *M. tuberculosis* PurF protein. The sequences compared are *P. falciparum* putative asparagine synthetase (NCBI gi 16805184) (top line) and *M. tuberculosis* PurF protein (NCBI gi 15607948). a) Alignment with BLOSUM100 with a bit score of 23.2. b) The alignment with BLOSUM50 with a bit score of 42.5. c) The alignment region with PfFSmat60 where the structural elements highlighted corresponds closely to the three-dimensional structural superposition of the common domain shared between PurF and asparagine synthase families. The bit score was equal to 369.6 at an E-value of 1.7e-106. The known crystal structures of *E. coli* asparagine synthetase B (1CT9, chainA) and *B. subtilis* PurF protein (1GPH, chain 3) were used to assign the secondary structure elements. The beta sheets are shaded grey and the alpha elements are shown in white text and a black background.