Table 5. Sequence pairs associated with structural data

NCBI gi numbers	Bit scores		Alignment extension	
	BLOSUM-62 matrix	Adjusted matrix	Left	Right
30749342 15609900	54.6	57.8	0	0
2624735 21219111	82.5	86.9	53	2
28948857 6137489	57.7	55.5	97	0
22219360 18418224	31.5	34.3	0	0
23613598 15674896	43.6	45.0	-1	0
123499 4502171	31.0	32.2	0	-5
3319034 19703667	36.6	39.9	0	90
2392498 33358140	37.5	35.2	18	0
28209987 15610383	77.3	80.0	0	0
15894223 28948854	97.0	97.4	0	0
28212114 27574180	46.2	44.2	0	0
28210575 22219354	57.6	58.5	4	12
16079353 15610243	28.2	28.9	0	3
15895916 15609357	41.6	42.9	0	0
18309683 15609676	66.1	70.1	1	67
28210205 13399468	30.6	28.6	0	0
18310891 6014910	91.5	89.6	0	0
15895827 15610982	61.3	63.3	0	0
15004757 14278695	93.1	96.8	0	0
15894130 28373649	32.5	32.0	0	0
15895511 18158792	44.4	49.8	1	8
15893474 1127200	39.9	44.0	48	91
15893312 3915100	29.9	30.9	0	0
15894597 17943056	62.1	59.8	-1	1
15894033 3024624	25.4	26.7	0	0
28210529 3892001	44.7	43.3	0	13
15893754 1633298	34.5	35.3	0	28
15594366 1346693	67.4	64.5	0	2
11497009 15827115	68.0	69.5	1	38
29726767 11139534	37.9	41.3	0	-1
16805184 15607948	29.7	31.8	182	1
102245 2498360	31.3	35.7	19	0

A test set of sequence pairs for which three-dimensional structural evidence provides support for alignment accuracy and at least one of the sequences is from an organism with strong compositional bias. The sequence pairs meet the following criteria: (*i*) the normalized BLOSUM-62 alignment is <100 bits, with as many as possible <40 bits; (*ii*) a crystal or NMR structure exists for each sequence or (because relatively few proteins from strongly biased organisms have known structures) for a homologous sequence that is closely related enough to be aligned unambiguously; and (*iii*) a structural superposition can be made that covers the aligned region well enough to define the corresponding secondary structure elements. The adjusted BLOSUM-62 matrices were constructed by using background frequencies from each sequence pair and pseudocounts as defined by the ‡ footnote in Table 1 of the main article; details of the bit score calculation and statistics are given in the * footnote of Table 1. The lengths, in residues, of alignment extensions that were yielded by matrix adjustment (columns 5 and 6 above) include alignment positions in which amino acids are aligned with gaps. Of the organisms represented in this test set, the following are considered to have strong compositional bias in the sense discussed in ref. 1: *Plasmodium falciparum*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium ulcerans*, *Streptomyces coelicolor*, *Streptomyces lividans*, *Streptomyces fradiae*, *Fusobacterium nucleatum*, *Clostridium tetani*, *Clostridium acetobutylicum*, *Clostridium perfringens*, *Borrelia burgdorferi*, and *Dictyostelium discoideum*.

1. Wan, H. & Wootton, J. C. (2000) *Comput. Chem.* **24**, 71–94.