Supporting Information

Methods

Dataset

Choosing the dataset was a crucial part of the analysis, since we required pairs with high sequence identity, which can be readily compared and at the same time high confidence in the oligomeric state, i.e. the correct biological unit. We analysed ten SCOP (1) protein families, which, according to the 3DComplex database (2), have at least one dimer and one homologous tetramer or hexamer with the same dimeric binding mode and sequence identity higher than 40%. Homomeric tetramers and hexamers with dihedral symmetry are, as explained in the main text, the simplest protein complexes with more than one type of interface. Once a pair of such structures was found, other family members going down to 30% sequence identity were also analysed (Table S1). Note that on average, about half of structures will have a conserved oligomeric state at the level of 40% sequence identity, and this proportion grows drastically with large sequence identities (3).

In the case of the phosphoribosyltransferase family (SCOP family identifier: 53272), all members exist as either dimers or tetramers, but the binding mode changes, i.e. interfaces between them do not structurally overlap. Therefore, the members of the family were split into two paralogous groups: the PyrR and the uracil phosphoribosyltransferase family - thus our dataset consists of eleven groups (subfamilies) of proteins belonging to ten SCOP families.

As a result of these dataset selection requirements, all but one selected subfamily contained only orthologues. The exception is the interleukin 8-like chemokine family. This high proportion of orthologues is not surprising, as paralogues are almost always more divergent in sequence than orthologues as a natural consequence of their functional divergence.

Biological units

Oligomeric state of the protein is decided on based on the annotation in the manually curated 3DComplex database. The biological unit files were obtained either from the biological unit file in the Protein Data Bank (PDB) as assigned by the authors, or from PISA (4), depending on which of them agreed with the 3DComplex database annotation.

Calculating geometric parameters

When comparing the crystal structures within a family, four protein regions were defined:

(i) the dimeric interface – the interface conserved in all of the homologues within the family, both dimers and tetramers/hexamers.
(ii) the tetrameric (or hexameric) interface – the interface which exists only in homologues with higher oligomeric state.

(iii) the region on the surface of the dimeric homologue which corresponds to the residues involved in the interface in homologues of higher oligomeric state (tetrameric or hexameric).
(iv) evolutionary core - structurally conserved core of a protein subunit.

The subunit evolutionary core was defined by using sieve fit as defined by Lesk (5). In this method, all atoms (or in this case all residue backbone atoms) that superimpose with an RMSD lower than some threshold (here 0.5 Å) are referred to as the *subunit core* and only those are used for the structural fit. The 0.5 Å threshold is empirical and represents an error window for high-resolution crystal structures.

A schematic illustration of all the structural fits is provided in Fig. S1 .

First A' to A superposition was done using only residues which make dimeric interface contacts in all analyzed homologues of the family (common dimeric residues, shown in green in Fig. S2). The dimeric interface rotation angle (dRot) translation vector (dTrans) and RMSD (dRMSD) were defined by superimposing the same residues of the B subunits - these parameters illustrate the contribution of local differences within the conserved (dimeric) interface to the overall structure geometry.

After superimposing the centres of mass of subunit evolutionary cores (A' to A) to an RMSD of approximately 0.5 Å, two types of translations were done. First, we translated and rotated subunit B to fit its evolutionary core. These geometric parameters (sRot and sTrans) show the difference in relative orientations of whole subunits around the dimeric interface. Secondly, after superimposing the whole subunit evolutionary core (A' to A) we translated and rotated only the tetrameric (or hexameric) interface residues of subunit A' to A. Tetrameric/hexameric interface rotation angle (tRot), translation vector (tTrans) and RMSD (tRMSD) yield the differences in position of the interface residues relative to the subunit evolutionary core.

Since all of the tetrameric and hexameric homologue structures are symmetric, the values we provide are for a chosen pair of A and B subunits. We have also calculated these values for other combinations of subunits, and values for RMSD vary by less than 0.05 Å, rotation angles by less than 1°, and translation vectors by less than 0.1 Å.

All the dimeric interface parameters were calculated using the common set of interface residues for each of the families analyzed. Geometric parameters for the tetrameric/hexameric interface of a dimer/tetramer (or dimer/hexamer) pair were calculated by superimposing the tetramer/hexamer interface residues of one to the corresponding surface residues of the other. In cases where two tetramers (or two hexamers) were compared, interfaces were fitted from the perspective of both of the structures yielding two values for each of the parameters for these pairs. The hexameric D-ribulose-5-phosphate 3-epimerase family analysed (SCOP identifier: 51372) has a cyclic, face-to-back type of hexameric interface. While dimeric and tetrameric interfaces in our dataset are all of the face-to-face type and consist of a single interface patch, hexameric interfaces consist of two distinct sets of residues, contacting different chains within the oligomeric structures. We have calculated all the parameters for the two halves of the hexameric interfaces separately – provided as A and B values in the Table S3. All the RMSD values are for backbone atoms only.

Hypothetical tetramers/hexamers and dDist

After superimposing evolutionary cores of the dimer and one half of the tetramer (or one third of the hexamer), the same was done for the other half (thirds). In this way, a hypothetical higher oligomer was constructed. This is by no means an attempt at docking, but rather a method to illustrate how geometry of the dimer influences the geometry of the tetramer/hexamer. It shows how symmetrical a structure of a dimer "forced" into a homologous tetramer (or hexamer) would be, without any kind of refinement.

We used another simple measure to illustrate this: the difference in distance, dDist (in Å) of interacting residues between subunits A' and C', and the corresponding residues of subunits B' and D'. Natural tetramers with dihedral symmetry have symmetric tetrameric interfaces, where corresponding residue pairs form contacts between subunits B and D, in the same way as between subunits A and C. Large differences in distance between A':C' and B':D' contact residues suggests that the homolog would require significant changes in the dimer geometry or form a tetramer via a different (although probably at least partially overlapping) surface.

Contact residues and protein interfaces

Interface residues are defined as all the residues making atomic contacts with residues from another subunit. Two atoms are considered to be in contact if the distance between them is equal to, or less than, the sum of their van der Waals radii plus 0.5 Å. The van der Waals radii used are defined in (6).

Accessible surface area

Accessible surface areas were calculated using the NACCESS algorithm (7). Protein regions – interior, surface and interface core, rim and support - were defined according to the thresholds described in (8).

Geometric parameter ratios and density plots

The ratios of geometric parameters per family in Fig. S5 for pairs of structures where oligomeric states are conserved versus where oligomeric has changed, are calculated as:

$$\mathrm{P_{ratio}} = \sum \frac{P_{conserved\_pair}}{N_{conserved\_pair}} \Big/ \sum \frac{P_{changed\_pair}}{N_{changed\_pair}},$$

where P is the value of the parameter for either a pair of structures with oligomeric state changed or conserved and N is the total number of pairs.

A_{rel} of homodimers

Relative accessible surface area (A_{rel}) is a ratio between a protein accessible surface area and the accessible surface area predicted for a protein of its molecular weight (9). Plotting molecular weight versus total accessible surface area (ASA) for a non-redundant set of

2748 homodimers, yields a simple power-law relationship of ASA and molar mass of the dimer (M). Calibration based on homodimers defined predicted ASA of a homodimer as:

$ASA = 4.30 M^{0.780}$.

Comparison of geometric parameters

Significant differences in intersubunit geometry are larger between homologues, than between different crystal structures of the same protein

Throughout this work, we calculate geometric variation between homologues and compare it to the variation between multiple crystal structures of the same protein wherever possible. This allows us to distinguish geometric variation that corresponds to functional allosteric changes or simply flexibility of a protein, from genuine variation in evolution across homologues.

For example, in the case of PyrR family homologues we compared six available crystal structures of two homologues - *Bacillus subtilis* PyrR (BsPyrR) and *Bacillus caldolyticus* (BcPyrR) (Table S5). Our data shows that although degree of geometrical change measured is influenced significantly by the crystallization conditions - both the crystal form and ligands bound - a clear difference can be seen when comparing different structures *versus* different homologues. A similar analysis for each of the eleven families is provided in Table S4.

Geometric and sequence conservation parameters and change in oligomeric state

In addition to the study of individual families, we have also evaluated each of the geometric and sequence parameters mentioned in the main text in the context of the whole dataset. Fig. S5 shows density plots for ratios of the sequence and geometric parameters between pairs which conserve and those which change their oligomeric state.

The density plots for ratios of the three sequence conservation parameters show how, for this set of high sequence identity homologues, sequence conservation presents a good predictor of oligomeric state change: in seven out of eleven families, the sequence identities of pairs of homologues with conserved oligomeric states are greater than those with different oligomeric states. However, the ratios are usually close around 1, so it is difficult to predict the sequence identity cut off based on which one could predict the oligomeric state. A similar conclusion holds for residue conservation of the dimeric interface. Tetrameric interface conservation has some larger ratios, and can be a good predictor of oligomeric state change, but only for some families (e.g. TIM or GABA aminotransferase-like family).

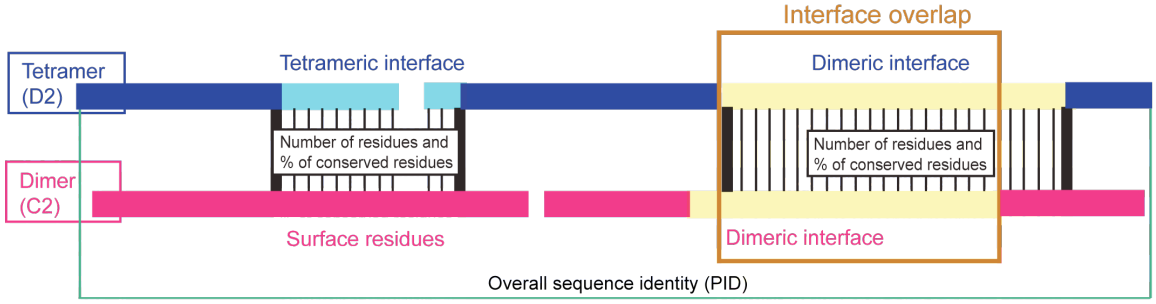
Figure S5B shows density plots for ratios of mean geometric parameter values. Parameters with ratios close to 0 could be used as predictors of oligomeric state change within a family. Tetrameric/hexameric interface parameters are in general better predictors of oligomeric state change than simple sequence conservation. In just over half of the families analysed (six out of eleven), subunit centre of mass rotation around the dimeric interface (sRot) correlates well with oligomeric state change.

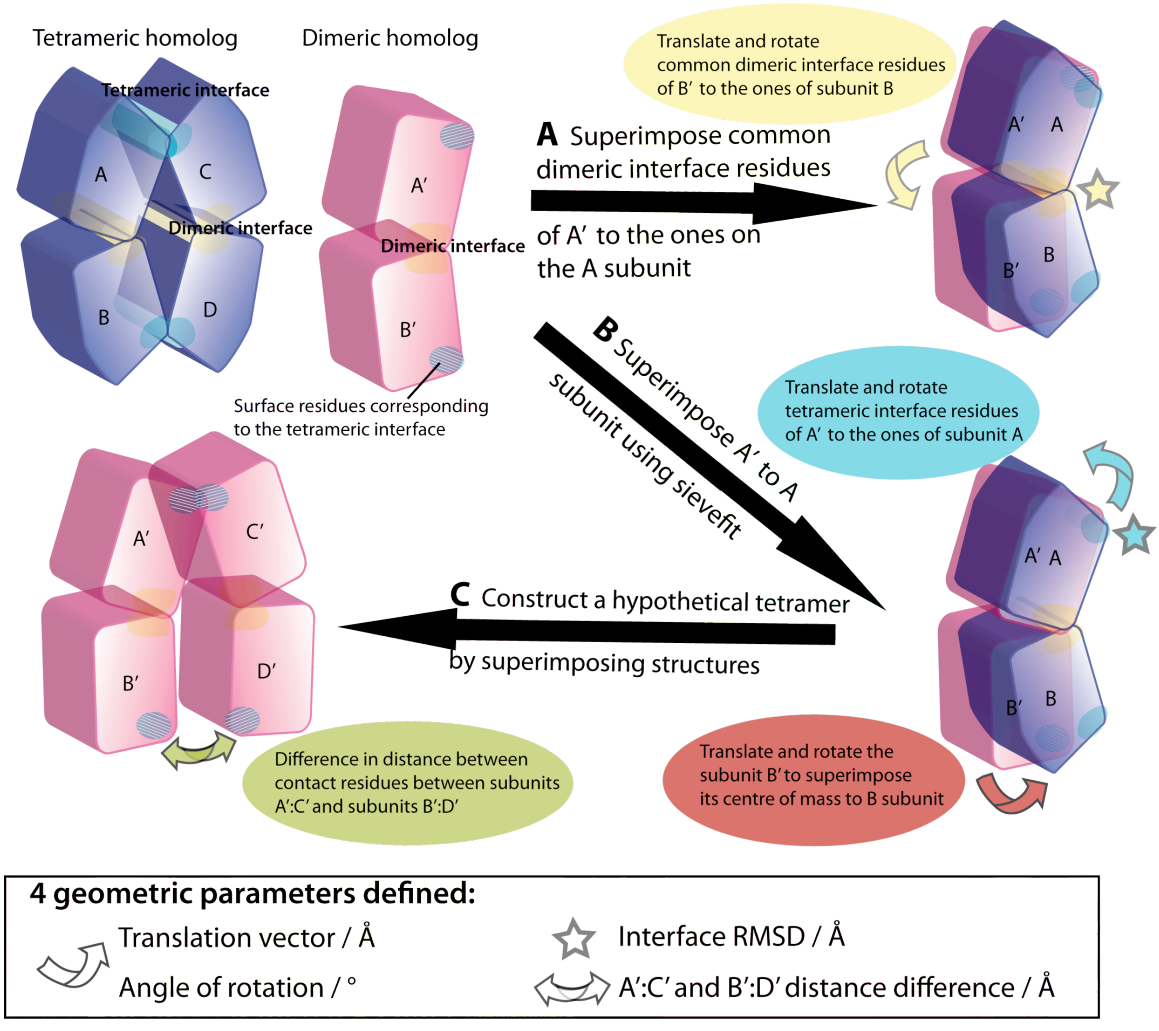
In conclusion, no single parameter is entirely predictive of oligomeric state change. Rather, a combination of tetrameric/hexameric interface parameters and subunit centre of mass rotation correlate best with changes in oligomeric state.

Structural plasticity of protein complexes

Analyzing geometric changes in eleven families revealed large differences in plasticity across families. In other words, close homologues in one family can have larger geometric differences than distant homologues in another family. Thus, geometric changes easily accommodated in one family can imply a change in oligomeric state in another. To explore these differences in plasticity between families further, we compared their relative accessible surface areas (A_{rel}) (Table S6). Proteins with high A_{rel} in the bound conformation are predicted to undergo large conformational changes upon binding. Proteins with high A_{rel} in the free state are predicted to be more flexible than average (9). Families assigned to the direct model (IV) and geometric model (III) have average homodimeric A_{rel} values of 0.9 and 1.0, respectively. This means that the dimers from families assigned to the direct model IV have are less flexible and more conformationally constrained than the ones from families assigned to the geometric model III. The direct model IV families in turn exhibit larger interface sequence changes across homologues with different oligomeric states.

Supplementary Figures



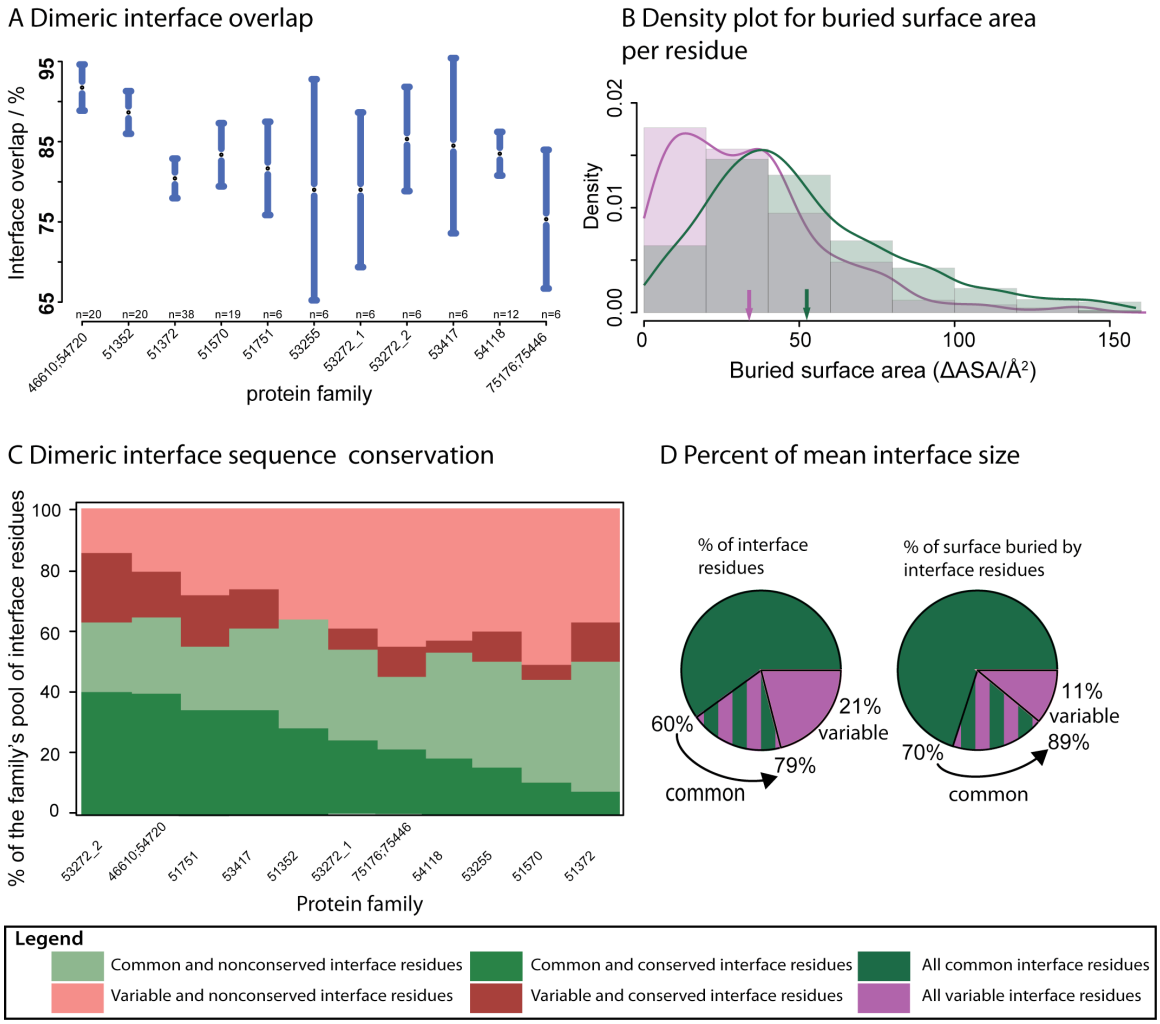


Supplementary Figure 1

(A) Scheme of pairwise sequence comparisons of interfaces. All members of a family (Table S1) were compared with each other. The dimeric interface (here and in the main text Fig. 1 in yellow) was defined as the one conserved between all the homologues, and the tetrameric/hexameric interface was the one forming in only some of the homologues. Dimeric homologues do not have tetrameric/hexameric interface residues, but corresponding surface residues can be defined from the sequence alignment. Interface conservation and overlap were calculated as percentages of conserved or overlapping residues from the perspective of each of the homologues.

(B) Geometric comparisons of homologous oligomeric structures.

Each pair of structures was superimposed in two ways, first by superimposing common dimeric interface regions (corresponding to green residues in multiple sequence alignments) and then by superimposing the evolutionary cores of subunits A and A'. After superimposing the common dimeric residues (A), dimeric interface rotation angle and translation vector were defined by superimposing the same residues of B subunits (shown in yellow). After superimposing subunit evolutionary cores (B), two types of translations were carried out: on the evolutionary core and on tetrameric/hexameric interface residues.



Supplementary Figure 2

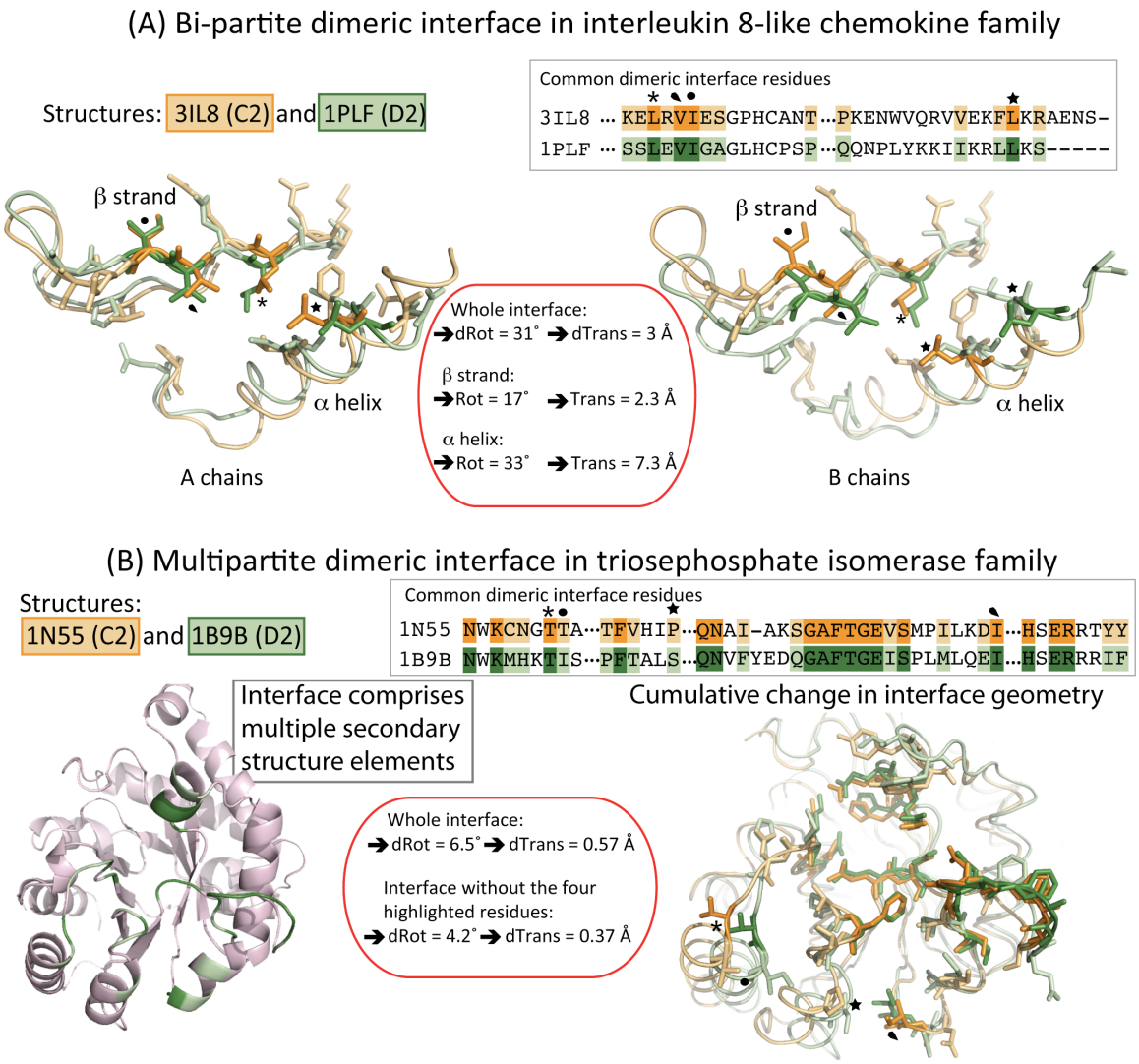
(A) Sequence overlap of dimeric interfaces. Means of dimeric interface overlaps (black dots), with blue bars showing 0.95 confidence level for each of the eleven families. Dimeric interface overlaps range from 59% to 100%. Mean values range from 75% to 92% for D-ribose-5-phosphate isomerase and Fe,Mn superoxide dismutase family, respectively.

(B) Common dimeric interface residues bury, on average more surface than the variable interface residues. The mean values are 52 and 34 Å^2, respectively (p-value < 2.2e-16, independent 2-group Mann-Whitney U test) as indicated by the two arrows.

(C) Sequence conservation (in the simplest conserved/non conserved form) of the dimeric interface residues. Each of the families analysed has a set of interface residues which are common for all of the dimeric interfaces (light green and green) and interface residues which make interface contacts in only a subset of the structures (light red and dark red). Interface residue pool of a family is a set of all the residues which make at least one interface contact in at least one of the structures, or a union of all the green and red residues. Green and dark red represent proportion of sequence conserved residues. The

proportion of common (both palegreen and green) residues represents 43% to 68% of the interface pool, depending on the protein family.

(D) The proportion of common residues represents from 60% to 81% of family's mean number of dimeric interface residues. Since common residues are on average more buried, they at the same time represent from 70 to 90% of the dimeric interface surface area. In addition, there is a small number of variable, often unique residues, which contribute to the remaining 10-30% of the buried surface.

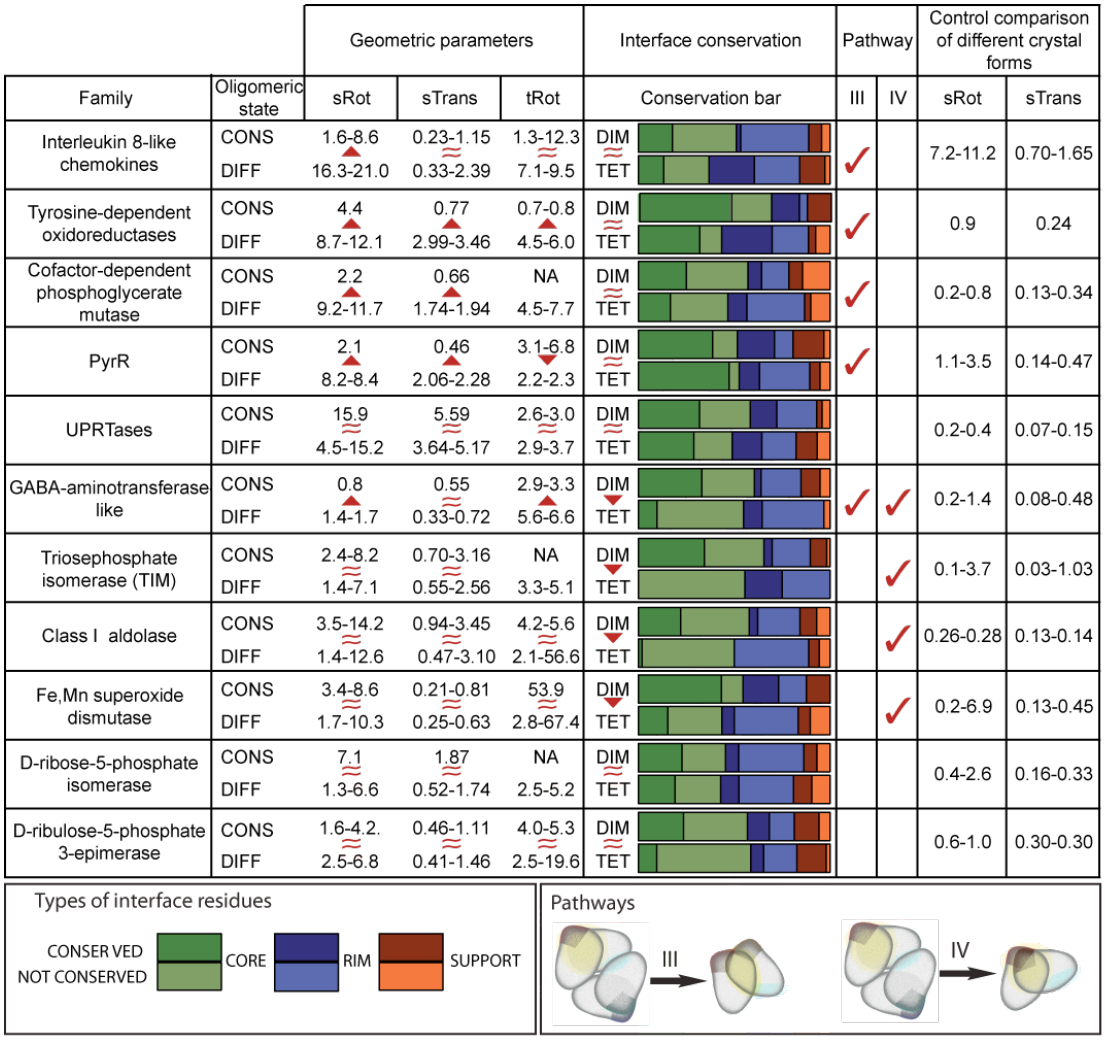


Supplementary Figure 3. Evolution of multipartite interfaces comprised of several secondary structure elements. Protein interfaces can have significant residue overlaps, as a consequence of their common ancestry, but at the same time exhibit large geometric differences. This can be explained by the fact that interfaces often comprise two or more secondary structure elements. For example, large geometric differences between homologous interfaces with conserved binding modes in the chemokine family are enabled by its bi-partite structure. In the triosephosphate isomerase family, the dimeric

interface comprises multiple secondary structure elements, but their relative positions are structurally conserved.

(A) Bi-partite dimeric interface in the interleukin 8-like chemokine family. The dimeric structure (PDB: 3IL8, in yellow) and the tetramer (PDB: 1PLF, in green) represent the homologous pair with the largest dRot value (31°) in the family. Common dimeric interface residues of chain A can be superimposed with an RMSD of 1.3 Å. The other half of the dimeric interface, from chain B, shows a large 31° rotation, which comes from a large shift of the helix in the bi-partite interface. The conserved Leu residue, marked with a star, best illustrates the large shift of the helix. When the geometric comparisons are done for each of the two parts of the interface (α helix and β strand) separately, one sees that the large shift comes from the α helix part of the interface, while the majority of common interface residues come from the β strand.

(B) The triosephosphate isomerase interface comprises multiple secondary structure elements. The dimeric structure (PDB: 1N55, in yellow) and the tetrameric (PDB: 1B9B, in green) represent the homologous pair with the largest dRot value (6.5°) in the family. The cartoon representation of the 1B9B structure in pink, with interface residues in green, shows the interface comprises residues from at least eight different helices and loops. Their relative arrangements are conserved, and the geometric difference between interfaces is an effect of cumulative small changes. For example, the four interface residues which superimpose worst between the 1B9B and 1N55 structures, belong to three different secondary structure elements (marked in the alignment). When each of those residues is excluded from the superpositions, the dRot value decreses slightly. When all four of them together are removed, the values add up and dRot goes down to 4.2°.



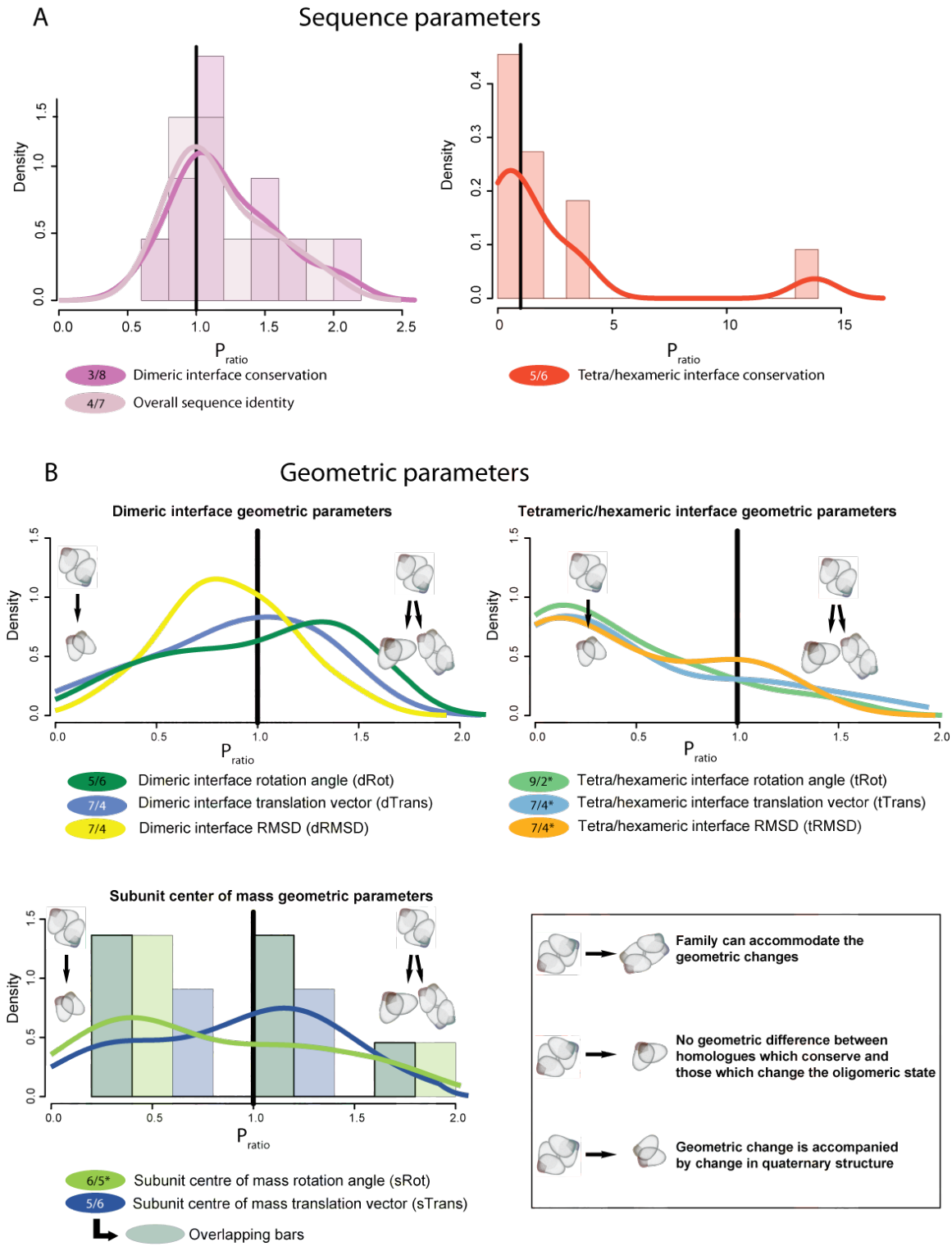
Supplementary Figure 4 Summary of main geometric parameters and interface conservation for each of the eleven families analysed.

We have assigned the *a priori* defined model (see main text, Fig. 1) to each of the eleven families based on the geometric parameters and interface conservation. We have compared ranges of geometric values (sRot and sTrans and tRot) between all pairs with different and ones with conserved oligomeric states within a family. Higher geometric variation between homologues with different (DIFF) oligomeric state, than the ones with conserved (CONS) oligomeric state indicates towards the geometric evolutionary model III.

We have also compared sequence conservation of both dimeric and tetrameric/hexameric interface between all pairs with different oligomeric states. Lower sequence conservation of the tetrameric, than the dimeric interface between pairs of homologues with different oligomeric state indicates the direct evolutionary model IV.

The patterns in geometric and sequence parameters lead us to conclude that three families follow the direct model IV, and four follow the geometric model III. For the remaining

four families it is not possible to make an unambiguous decision on the model since data either points towards both of the pathways (e.g. for GABA-aminotransferase like family) or not clearly towards either of them (e.g. UPRTase subfamily).



Supplementary Figure 5 Ratios of sequence and geometric conservation parameters between pairs which conserve and those which change their oligomeric state.

Panel A shows density plots for ratios of the three sequence conservation parameters between pairs which conserve and those which change their oligomeric state. Values are larger than 1 for families where sequence conservation is higher in pairs which conserve their oligomeric state. Spread of values is larger than for geometric parameters, which means that for some families sequence conservation, especially of the tetrameric/hexameric interface, can be a very good predictor of oligomeric state change. For some, though, oligomeric state change does not correlate at all with the sequence change.

Panel B shows density plots for ratios of mean geometric parameter values between pairs which conserve and those which change their oligomeric state. Ratio smaller than 1 for some families means geometric differences are on average larger in cases where oligomeric state has changed. Ratios of families which have geometric parameter ratios smaller and larger than 1, repectively are given in the legends below each plot.

How resolved the two modes of the plot are is more informative than the simple ratio of families. For parameters which have values significantly lower than 1, the value of this parameter could be used as a predictor of oligomeric state change within that family.

This holds true for four geometric parameters measured, emphised by an asterisk: Angle of rotation between centers of mass of the two subnits in the dimer (sRot) as well as for all three measures of the tetrameric/hexameric interface structural comparison parameters (angle of rotation (tRot), translation vector (tTrans) and tRMSD).

Supplementary Table 1 All pairs of protein structures from ten SCOP families (11 subfamilies). Each familiy was selected from the 3DComplex database as having at least one pair of homologous dimers and tetramers (or hexamer) with the same dimeric binding mode and sequence identity higher than 40%. Pairs in the table are ordered by descending pairwise percent sequence identity. *Code* refers to the PDB identifier of each protein and *sym* defines the oligomeric state of the biological unit according to the 3DComplex database and literature. C2 and C3 are cyclic dimer and trimer and D2 and D3 are dihedral tetramer and hexamer, respectively. *Crystal* defines the space group in which the X-ray structure is crystallised. *Family summary* is the number of dimers and tetramers/hexamers analysed per family.

SCOP Family	Protein family	Family summary	code1	sym1	crystal1	code2	sym2	crystal2
54118	Interleukin 8-like chemokines	N(C2) = 1 N(D2) = 3	1F9S	D2	$P2_12_12_1$	1PLF	D2	$P2_12_12_1$
			1TVX	D2	$P2_1$	1PLF	D2	$P2_12_12_1$
			1TVX	D2	$P2_1$	1F9S	D2	$P2_12_12_1$
			1TVX	D2	$P2_1$	3IL8	C2	$P3_121$
			1F9S	D2	$P2_12_12_1$	3IL8	C2	$P3_121$
			1PLF	D2	$P2_12_12_1$	3IL8	C2	$P3_121$
51751	Tyrosine-dependent oxidoreductases	N(C2) = 1 N(D2) = 2	1T2A	D2	$P2_12_12_1$	1DB3	C2/D3	$P6_322$
			1N7G	D2	$P2_12_12_1$	1DB3	C2/D3	$P6_322$
			1N7G	D2	$P2_12_12_1$	1T2A	D2	$P2_12_12_1$
53255	Cofactor-dependent phosphoglycerate mutase	N(C2) = 2 N(D2) = 1	1QHF	D2	C2	1E58	C2	$P2_12_12$
			1QHF	D2	C2	2HHJ	C2	$P2_12_12_1$
			2HHJ	C2	$P2_12_12_1$	1E58	C2	$P2_12_12$
53272	Phosphorybosyltransferases (PRTases)	N(C2) = 1 N(D2) = 2	1NON	D2	C2	1A3C	C2	C2
			1W30	D2	$P3_121$	1NON	D2	C2
			1W30	D2	$P3_121$	1A3C	C2	C2
		N(C2) = 1 N(D2) = 2	1O5O	D2	C2	1I5E	C2	$P3_121$
			1BD3	D2	$P2_1$	1I5E	C2	$P3_121$
			1BD3	D2	$P2_1$	1O5O	D2	C2
53417	GABA-aminotransferase-like	N(C2) = 1 N(D2) = 2	1EJI	D2	$P4_32_12$	1RV3	D2	$P4_12_12$
			1RV3	D2	$P4_12_12$	1KKP	C2	$P2_12_12_1$
			1EJI	D2	$P4_32_12$	1KKP	C2	$P2_12_12_1$
51352	Triosephosphate isomerase (TIM)	N(C2) = 4 N(D2) = 1	1KV5	C2	$P2_12_12_1$	1N55	C2	C2
			1KV5	C2	$P2_12_12_1$	1R2R	C2	$P2_12_12_1$
			1R2R	C2	$P2_12_12_1$	1N55	C2	C2
			1B9B	D2	$P3_221$	2BTM	C2	$P2_12_12$
			1N55	C2	C2	2BTM	C2	$P2_12_12$
			1B9B	D2	$P3_221$	1N55	C2	C2
			1B9B	D2	$P3_221$	1R2R	C2	$P2_12_12_1$
			1B9B	D2	$P3_221$	1KV5	C2	$P2_12_12_1$
			1KV5	C2	$P2_12_12_1$	2BTM	C2	$P2_12_12$
			2BTM	C2	$P2_12_12$	1R2R	C2	$P2_1$
51570	Class I aldolase	N(C2) = 3 N(D2) = 2	1J2W	D2	$P2_12_12_1$	3R12	C2	$P2_1$
			1MZH	C2	$C222_1$	3R12	C2	$P2_1$
			1J2W	D2	$P2_12_12_1$	1MZH	C2	$C222_1$
			1VCV	C2	$P6_3$	1MZH	C2	$C222_1$
			1VCV	C2	$P6_3$	3R12	C2	$P2_1$
			1N7K	D2	$P2_12_12$	3R12	C2	$P2_1$
			1N7K	D2	$P2_12_12$	1J2W	D2	$P2_12_12_1$
			1N7K	D2	$P2_12_12$	1MZH	C2	$C222_1$
			1J2W	D2	$P2_12_12_1$	1VCV	C2	$P6_3$
			1N7K	D2	$P2_12_12$	1VCV	C2	$P6_3$
46610;54720	Fe,Mn superoxide dismutase	N(C2) = 3 N(D2) = 2	1GV3	C2	H3	1JR9	C2	$P4_32_12$
			1IXB	C2	$P2_1$	1JR9	C2	$P4_32_12$
			3MDS	D2	$P4_12_12$	1JR9	C2	$P4_32_12$
			3MDS	D2	$P4_12_12$	1IXB	C2	$P2_1$
			3MDS	D2	$P4_12_12$	1GV3	C2	H3
			1IXB	C2	$P2_1$	1GV3	C2	H3
			1KKC	D2	$P2_12_12_1$	1IXB	C2	$P2_1$
			1KKC	D2	$P2_12_12_1$	3MDS	D2	$P4_12_12$
			1KKC	D2	$P2_12_12_1$	1JR9	C2	$P4_32_12$
			1KKC	D2	$P2_12_12_1$	1GV3	C2	H3
75176;75446	D-ribose-5-phosphate isomerase	N(C2) = 2 N(D2) = 1	1O8B	C2	P1	1M0S	C2	$P2_12_12_1$
			1LK5	D2	$P2_12_12_1$	1M0S	C2	$P2_12_12_1$
			1LK5	D2	$P2_12_12_1$	1O8B	C2	P1
51372	D-ribulose-5-phosphate 3-epimerase	N(C2) = 2 N(D3) = 3	1RPX	D3	$P3_221$	1TQJ	D3	$P2_1$
			1TQX	C2	$P2_12_12_1$	1H1Y	C2	$P2_1$
			2FLI	D3	$P2_1$	1H1Y	C2	$P2_1$
			1RPX	D3	$P3_221$	2FLI	D3	$P2_1$
			1TQJ	D3	$P2_1$	2FLI	D3	$P2_1$
			1TQJ	D3	$P2_1$	1H1Y	C2	$P2_1$
			1RPX	D3	$P3_221$	1H1Y	C2	$P2_1$
			2FLI	D3	$P2_1$	1TQX	C2	$P2_12_12_1$
			1TQJ	D3	$P2_1$	1TQX	C2	$P2_12_12_1$
			1RPX	D3	$P3_221$	1TQX	C2	$P2_12_12_1$

Supplementary Table 2 Sequence comparison parameters for all homologues pairs. Pairs within families are sorted by descending overall protein sequence identity (PID). Cells of the table are shaded according to the methods illustrated in Figure S1. Sequence comparison parameters are all provided twice, from the perspective of each of the two structures (represented by *code1* and *code2* PDB identifiers, respectively). Residue conservation for each of the interfaces is shown as percentage of identical residues. Interface overlap is defined as the intersection of the two sets of interface residues, and percentage overlap is defined as ratio of a number of residues in the intersection set and all of the interface residues. *Code* refers to the PDB identifier of each protein and *sym* defines the oligomeric state of the biological unit according to the 3DComplex database and literature. C2 and C3 are cyclic dimer and trimer and D2 and D3 are dihedral tetramer and hexamer, respectively.

Pairwise sequence comparisons - Part 1

						Dimeric interface						Tetrameric interface					
family	code1	sym1	code2	sym2	PID	N residues		Residue conservation		Interface Overlap		N residues		Residue conservation		Interface Overlap	
54118	1F9S	D2	1PLF	D2	0.73	17	17	0.82	0.76	0.88	0.88	12	14	0.83	0.71	0.67	0.57
	1TVX	D2	1PLF	D2	0.56	16	17	0.44	0.41	0.88	0.82	13	14	0.85	0.79	0.62	0.57
	1TVX	D2	1F9S	D2	0.51	16	17	0.44	0.47	0.88	0.82	13	12	0.77	0.75	0.38	0.42
	1TVX	D2	3IL8	C2	0.45	16	18	0.38	0.39	0.88	0.78	13	/	0.62	/	/	/
	1F9S	D2	3IL8	C2	0.34	17	18	0.29	0.22	0.82	0.78	12	/	0.42	/	/	/
	1PLF	D2	3IL8	C2	0.33	17	18	0.24	0.22	0.82	0.78	14	/	0.43	/	/	/
51751	1T2A	D2	1DB3	C2	0.60	23	23	0.78	0.74	0.78	0.78	22	/	0.64	/	/	/
	1N7G	D2	1DB3	C2	0.57	25	23	0.72	0.61	0.76	0.83	29	/	0.66	/	/	/
	1N7G	D2	1T2A	D2	0.55	25	23	0.64	0.65	0.84	0.91	29	22	0.52	0.64	0.69	0.91
53255	1QHF	D2	1E58	C2	0.53	16	13	0.31	0.38	0.81	1.00	19	/	0.26	/	/	/
	1QHF	D2	2HHJ	C2	0.48	16	17	0.38	0.35	0.81	0.76	19	/	0.53	/	/	/
	2HHJ	C2	1E58	C2	0.47	17	13	0.53	0.54	0.59	0.77	/	/	/	/	/	/
53272_2	1NON	D2	1A3C	C2	0.73	19	18	0.84	0.83	0.84	0.89	11	/	1	/	/	/
	1W30	D2	1NON	D2	0.55	16	19	0.63	0.68	0.94	0.79	13	11	0.38	0.55	0.69	0.91
	1W30	D2	1A3C	C2	0.52	16	18	0.63	0.61	0.88	0.78	13	/	0.38	/	/	/
53272_1	1O5O	D2	1I5E	C2	0.63	45	38	0.62	0.68	0.78	0.92	23	/	0.78	/	/	/
	1BD3	D2	1I5E	C2	0.37	45	38	0.40	0.45	0.64	0.76	21	/	0.43	/	/	/
	1BD3	D2	1O5O	D2	0.36	45	45	0.40	0.36	0.82	0.82	21	23	0.38	0.35	0.71	0.65
53417	1EJI	D2	1RV3	D2	0.90	95	85	0.95	0.95	0.86	0.96	15	15	0.87	0.93	0.93	0.93
	1RV3	D2	1KKP	C2	0.47	85	72	0.47	0.58	0.75	0.89	15	/	0.13	/	/	/
	1EJI	D2	1KKP	C2	0.47	95	72	0.45	0.58	0.69	0.92	15	/	0.13	/	/	/
51352	1KV5	C2	1N55	C2	0.70	34	32	0.85	0.88	0.94	1.00	/	/	/	/	/	/
	1KV5	C2	1R2R	C2	0.51	34	33	0.50	0.48	0.88	0.91	/	/	/	/	/	/
	1R2R	C2	1N55	C2	0.51	33	32	0.48	0.53	0.91	0.94	/	/	/	/	/	/
	1B9B	D2	2BTM	C2	0.49	34	37	0.56	0.57	0.91	0.84	9	/	0.11	/	/	/
	1N55	C2	2BTM	C2	0.45	32	37	0.50	0.51	0.91	0.78	/	/	/	/	/	/
	1B9B	D2	1N55	C2	0.43	34	32	0.50	0.53	0.88	0.94	9	/	/	/	/	/
	1B9B	D2	1R2R	C2	0.42	34	33	0.56	0.55	0.79	0.82	9	/	0.11	/	/	/
	1B9B	D2	1KV5	C2	0.42	34	34	0.44	0.50	0.88	0.88	9	/	0.11	/	/	/
	1KV5	C2	2BTM	C2	0.41	34	37	0.56	0.51	0.94	0.86	/	/	/	/	/	/
	2BTM	C2	1R2R	C2	0.39	37	33	0.54	0.61	0.81	0.91	/	/	/	/	/	/

Supplementary Table 2 – continued

Pairwise sequence comparisons - Part 2						Dimeric interface						Tetrameric interface					
family	code1	sym1	code2	sym2	PID	N residues		Residue conservation		Interface Overlap		N residues		Residue Conservation		Overlap	
51570	1J2W	D2	3R12	C2	0.45	28	32	0.57	0.53	0.96	0.84	10	/	0.2	/	/	/
	1MZH	C2	3R12	C2	0.44	33	32	0.42	0.41	0.94	0.97	/	/	/	/	/	/
	1J2W	D2	1MZH	C2	0.40	28	33	0.46	0.39	0.96	0.82	10	/	0.2	/	/	/
	1VCV	C2	1MZH	C2	0.39	31	33	0.39	0.36	0.84	0.79	/	/	/	/	/	/
	1VCV	C2	3R12	C2	0.38	31	32	0.45	0.41	0.81	0.78	/	/	/	/	/	/
	1N7K	D2	3R12	C2	0.36	26	32	0.35	0.34	0.85	0.69	21	/	0.05	/	/	/
	1N7K	D2	1J2W	D2	0.36	26	28	0.46	0.5	0.85	0.79	21	10	0	0.1	0	0
	1N7K	D2	1MZH	C2	0.36	26	33	0.35	0.36	0.88	0.7	21	/	0.05	/	/	/
	1J2W	D2	1VCV	C2	0.35	28	31	0.5	0.45	0.82	0.74	10	/	0.2	/	/	/
	1N7K	D2	1VCV	C2	0.34	26	31	0.31	0.29	0.81	0.68	21	/	0.1	/	/	/
46610;54720	1GV3	C2	1JR9	C2	0.57	17	16	0.88	0.88	0.94	1	/	/	/	/	/	/
	1IXB	C2	1JR9	C2	0.56	16	16	0.88	0.88	0.94	0.94	/	/	/	/	/	/
	3MDS	D2	1JR9	C2	0.56	18	16	0.89	0.94	0.89	1	15	/	0.27	/	/	/
	3MDS	D2	1IXB	C2	0.55	18	16	0.89	0.94	0.89	1	15	/	0.2	/	/	/
	3MDS	D2	1GV3	C2	0.55	18	17	0.89	0.94	0.94	1	15	/	0.27	/	/	/
	1IXB	C2	1GV3	C2	0.51	16	17	0.94	0.88	1	0.94	/	/	/	/	/	/
	1KKC	D2	1IXB	C2	0.49	17	16	0.53	0.69	0.82	0.88	16	/	0.38	/	/	/
	1KKC	D2	3MDS	D2	0.47	17	18	0.59	0.67	0.88	0.83	16	15	0.5	0.2	0.13	0.07
	1KKC	D2	1JR9	C2	0.45	17	16	0.65	0.63	0.82	0.88	16	/	0.31	/	/	/
	1KKC	D2	1GV3	C2	0.40	17	17	0.65	0.76	0.88	0.88	16	/	0.38	/	/	/
75176;75446	1O8B	C2	1M0S	C2	0.65	18	21	0.72	0.62	0.83	0.71	/	/	/	/	/	/
	1LK5	D2	1M0S	C2	0.44	24	21	0.29	0.33	0.71	0.81	22	/	0.32	/	/	/
	1LK5	D2	1O8B	C2	0.43	24	18	0.46	0.61	0.63	0.83	22	/	0.41	/	/	/
51372	1RPX_A	D3	1TQJ_A	D3	0.68	21	23	0.86	0.78	0.9	0.83	18	13	0.61	0.54	0.67	0.92
	1RPX_B	D3	1TQJ_B	D3	0.68	21	23	0.86	0.78	0.9	0.83	8	12	0.63	0.58	0.88	0.58
	1TQX	C2	1H1Y	C2	0.50	24	21	0.54	0.52	0.88	1	/	/	/	/	/	/
	2FLI_A	D3	1H1Y	C2	0.49	21	21	0.57	0.57	0.81	0.81	12	/	0.58	/	/	/
	2FLI_B	D3	1H1Y	C2	0.49	21	21	0.57	0.57	0.81	0.81	10	/	0.3	/	/	/
	1RPX_A	D3	2FLI_A	D3	0.47	21	21	0.38	0.38	0.81	0.81	18	12	0.72	0.83	0.56	0.83
	1RPX_B	D3	2FLI_B	D3	0.47	21	21	0.38	0.38	0.81	0.81	8	10	0.38	0.4	0.75	0.6
	1TQJ_A	D3	2FLI_A	D3	0.45	23	21	0.35	0.33	0.87	0.95	13	12	0.62	0.75	0.69	0.75
	1TQJ_B	D3	2FLI_B	D3	0.45	23	21	0.35	0.33	0.87	0.95	12	10	0.25	0.2	0.75	0.9
	1TQJ_A	D3	1H1Y	C2	0.41	23	21	0.43	0.43	0.7	0.76	13	/	0.38	/	/	/
	1TQJ_B	D3	1H1Y	C2	0.41	23	21	0.43	0.43	0.7	0.76	12	/	0.25	/	/	/
	1RPX_A	D3	1H1Y	C2	0.41	21	21	0.48	0.48	0.76	0.76	18	/	0.56	/	/	/
	1RPX_B	D3	1H1Y	C2	0.41	21	21	0.48	0.48	0.76	0.76	8	/	0	/	/	/
	2FLI_A	D3	1TQX	C2	0.38	21	24	0.43	0.38	0.86	0.75	12	/	0.25	/	/	/
	2FLI_B	D3	1TQX	C2	0.38	21	24	0.43	0.38	0.86	0.75	10	/	0.3	/	/	/
	1TQJ_A	D3	1TQX	C2	0.36	23	24	0.39	0.33	0.78	0.75	13	/	0.15	/	/	/
	1TQJ_B	D3	1TQX	C2	0.36	23	24	0.39	0.33	0.78	0.75	12	/	0	/	/	/
	1RPX_A	D3	1TQX	C2	0.362	21	24	0.43	0.46	0.76	0.67	18	/	0.33	/	/	/
	1RPX_B	D3	1TQX	C2	0.362	21	24	0.43	0.46	0.76	0.67	8	/	0	/	/	/

Supplementary Table 3 Geometric comparison parameters for all homologous pairs. Pairs within families are sorted by descending overall protein sequence identity (PID). Cells of the table are shaded according to the methods illustrated in Fig. S1. *Code* refers to the PDB identifier of each protein and *sym* defines the oligomeric state of the biological unit according to the 3DComplex database and literature. C2 and C3 are cyclic dimer and trimer and D2 and D3 are dihedral tetramer and hexamer, respectively. dRot (angle of rotation), dTrans (translation vector) and dRMSD are obtained by superimposing a set of dimeric interface residues common for all analysed members of the family (coloured green in the family multiple sequence alignments). sRot and sTrans are obtained by superimposing the evolutionary core defined for each pair of proteins by the set of residues whose backbone atoms superimpose with an RMSD of 0.5 Å. Since, by definition, there is no common set of tetrameric/hexameric interface residues, tRot, tTrans and tRMSD are provided from the perspective of each of the tetrameric interfaces in the pair.

Pairwise geometric comparisons - Part 1

family	code1	sym1	code2	sym2	PID	dRot	dTrans	dRMSD	sRot	sTrans	tRot1	tTrans1	tRMSD1	tRot2	tTrans2	tRMSD2
54118	1F9S	D2	1PLF	D2	0.73	1.2	0.44	0.63	1.6	0.23	9.1	1.10	1.62	3.1	0.40	0.60
	1TVX	D2	1PLF	D2	0.56	10.7	1.76	1.18	8.6	1.15	12.3	0.51	1.58	2.3	0.19	0.79
	1TVX	D2	1F9S	D2	0.51	9.2	1.38	1.05	8.3	0.93	1.3	0.15	0.77	7.0	0.58	1.01
	1TVX	D2	3IL8	C2	0.45	21.9	1.58	0.80	21.0	2.39	7.1	0.57	0.91	/	/	/
	1F9S	D2	3IL8	C2	0.34	30.3	2.65	1.10	16.3	0.33	9.5	0.49	0.99	/	/	/
	1PLF	D2	3IL8	C2	0.33	31.6	2.99	1.33	18.3	0.38	9.0	0.55	1.41	/	/	/
51751	1T2A	D2	1DB3	C2	0.60	2.8	0.91	0.54	12.1	3.46	6.0	0.87	3.14	/	/	/
	1N7G	D2	1DB3	C2	0.57	1.3	0.47	0.56	8.7	2.99	4.5	1.05	3.44	/	/	/
	1N7G	D2	1T2A	D2	0.55	3.0	0.85	0.31	4.4	0.77	0.7	0.13	0.42	0.8	0.13	0.32
53255	1QHF	D2	1E58	C2	0.53	14.1	0.89	0.44	9.2	1.94	4.5	0.60	1.43	/	/	/
	1QHF	D2	2HHJ	C2	0.48	10.4	0.65	0.45	11.7	1.74	7.7	0.70	1.94	/	/	/
	2HHJ	C2	1E58	C2	0.47	4.9	0.30	0.37	2.2	0.66	/	/	/	/	/	/
53272_2	1NON	D2	1A3C	C2	0.73	1.9	0.21	0.31	8.2	2.28	2.3	0.15	0.57	/	/	/
	1W30	D2	1NON	D2	0.55	5.0	0.61	0.39	2.1	0.46	3.1	0.51	0.50	6.8	1.47	4.05
	1W30	D2	1A3C	C2	0.52	6.5	0.63	0.41	8.4	2.06	2.2	0.48	0.42	/	/	/
53272_1	1O5O	D2	1I5E	C2	0.63	2.6	0.65	1.01	4.5	3.64	3.7	0.43	1.89	/	/	/
	1BD3	D2	1I5E	C2	0.37	16.5	1.47	1.82	15.2	5.17	2.9	0.79	1.09	/	/	/
	1BD3	D2	1O5O	D2	0.36	14.2	1.02	2.01	15.9	5.59	2.6	0.64	1.54	3.0	1.02	2.17
53417	1EJI	D2	1RV3	D2	0.90	0.7	0.07	0.44	0.8	0.55	2.9	0.39	0.50	3.3	0.45	0.52
	1RV3	D2	1KKP	C2	0.47	0.6	0.81	1.42	1.4	0.33	6.6	0.89	0.74	/	/	/
	1EJI	D2	1KKP	C2	0.47	1.0	0.86	1.41	1.7	0.72	5.6	0.76	0.68	/	/	/
51352	1KV5	C2	1N55	C2	0.70	1.4	0.09	0.18	3.0	0.95	/	/	/	/	/	/
	1KV5	C2	1R2R	C2	0.51	1.4	0.18	0.82	2.4	0.70	/	/	/	/	/	/
	1R2R	C2	1N55	C2	0.51	2.7	0.21	0.82	5.5	1.78	/	/	/	/	/	/
	1B9B	D2	2BTM	C2	0.494	1.9	0.13	0.67	3.2	0.88	3.5	0.37	0.38	/	/	/
	1N55	C2	2BTM	C2	0.453	5.9	0.64	0.78	8.2	3.16	/	/	/	/	/	/
	1B9B	D2	1N55	C2	0.426	6.5	0.57	1.03	7.1	2.56	4.3	0.37	0.38	/	/	/
	1B9B	D2	1R2R	C2	0.421	4.6	0.41	0.79	1.4	0.55	5.1	1.04	0.38	/	/	/
	1B9B	D2	1KV5	C2	0.418	5.1	0.51	1.01	4.2	1.30	3.3	0.26	0.43	/	/	/
	1KV5	C2	2BTM	C2	0.409	4.5	0.58	0.73	4.9	2.11	/	/	/	/	/	/
	2BTM	C2	1R2R	C2	0.392	3.6	0.52	0.72	3.1	0.76	/	/	/	/	/	/

Supplementary Table 3 - continued

Pairwise geometric comparisons - Part 2

family	code1	sym1	code2	sym2	PID	dRot	dTrans	dRMSD	sRot	sTrans	tRot1	tTrans1	tRMSD1	tRot2	tTrans2	tRMSD2
51570	1J2W	D2	3R12	C2	0.45	1.5	0.57	0.47	2.7	0.90	2.1	0.23	0.81	/	/	/
	1MZH	C2	3R12	C2	0.44	5.2	0.28	0.70	8.2	2.56	/	/	/	/	/	/
	1J2W	D2	1MZH	C2	0.40	6.3	0.44	0.83	10.8	2.32	3.1	0.27	0.50	/	/	/
	1VCV	C2	1MZH	C2	0.39	9.8	0.73	1.39	10.8	2.66	/	/	/	/	/	/
	1VCV	C2	3R12	C2	0.38	5.7	0.55	1.20	3.5	0.94	/	/	/	/	/	/
	1N7K	D2	3R12	C2	0.36	3.3	0.51	0.84	12.6	3.10	43.8	1.85	2.93	/	/	/
	1N7K	D2	1J2W	D2	0.36	4.6	0.21	0.92	14.2	3.45	5.6	0.91	0.72	4.2	0.65	1.16
	1N7K	D2	1MZH	C2	0.36	1.9	0.33	0.74	6.7	1.40	56.6	3.28	3.37	/	/	/
	1J2W	D2	1VCV	C2	0.35	4.2	0.81	1.22	1.4	0.47	8.4	2.09	2.48	/	/	/
	1N7K	D2	1VCV	C2	0.34	8.2	0.89	1.54	11.9	2.66	7.7	2.71	2.19	/	/	/
46610;54720	1GV3	C2	1JR9	C2	0.57	3.3	0.06	0.34	3.4	0.21	/	/	/	/	/	/
	1IXB	C2	1JR9	C2	0.56	7.7	0.51	0.42	8.6	0.71	/	/	/	/	/	/
	3MDS	D2	1JR9	C2	0.56	2.1	0.10	0.33	3.2	0.25	3.3	0.45	1.32	/	/	/
	3MDS	D2	1IXB	C2	0.55	5.8	0.50	0.24	6.3	0.56	12.3	0.95	4.01	/	/	/
	3MDS	D2	1GV3	C2	0.55	1.6	0.11	0.15	1.7	0.33	2.8	0.45	1.26	/	/	/
	1IXB	C2	1GV3	C2	0.51	4.5	0.55	0.21	5.5	0.81	/	/	/	/	/	/
	1KKC	D2	1IXB	C2	0.49	8.3	0.42	0.33	10.3	0.44	67.4	6.76	7.62	/	/	/
	1KKC	D2	3MDS	D2	0.47	3.3	0.17	0.30	3.9	0.43	53.9	5.97	7.09	33.2	3.58	7.44
	1KKC	D2	1JR9	C2	0.45	2.8	0.18	0.44	2.7	0.59	58.1	6.25	7.19	/	/	/
	1KKC	D2	1GV3	C2	0.40	4.0	0.23	0.28	5.6	0.63	57.1	6.05	6.89	/	/	/
75176;75446	1O8B	C2	1M0S	C2	0.65	3.5	0.57	0.45	7.1	1.87	/	/	/	/	/	/
	1LK5	D2	1M0S	C2	0.44	1.8	0.38	0.75	1.3	0.52	5.2	0.68	1.84	/	/	/
	1LK5	D2	1O8B	C2	0.43	3.5	0.92	0.65	6.6	1.74	2.5	0.36	0.73	/	/	/
51372	1RPX_A	D3	1TQJ_A	D3	0.68	0.4	0.27	0.34	1.6	0.46	1.4	0.13	0.76	1.0	0.22	0.50
	1RPX_B	D3	1TQJ_B	D3	0.68	0.4	0.27	0.34	1.6	0.46	1.8	0.28	0.70	/1.87	0.36	0.70
	1TQX	C2	1H1Y	C2	0.50	0.9	0.19	0.80	4.2	1.11	/	/	/	/	/	/
	2FLI_A	D3	1H1Y	C2	0.49	1.1	0.32	0.57	4.1	1.10	3.6	0.42	0.59	/	/	/
	2FLI_B	D3	1H1Y	C2	0.49	1.1	0.32	0.57	4.1	1.10	5.7	1.39	1.69	/	/	/
	1RPX_A	D3	2FLI_A	D3	0.47	2.5	0.58	0.73	1.7	0.80	5.3	0.36	1.09	1.8	0.26	0.67
	1RPX_B	D3	2FLI_B	D3	0.47	2.5	0.58	0.73	1.7	0.80	4.0	0.44	0.50	4.0	0.48	0.44
	1TQJ_A	D3	2FLI_A	D3	0.45	2.4	0.48	0.75	2.5	0.64	1.9	0.20	0.97	2.3	0.22	0.88
	1TQJ_B	D3	2FLI_B	D3	0.45	2.4	0.48	0.75	2.5	0.64	3.1	0.76	0.63	3.3	0.70	0.55
	1TQJ_A	D3	1H1Y	C2	0.41	3.4	0.25	0.89	6.8	0.51	4.9	0.46	1.99	/	/	/
	1TQJ_B	D3	1H1Y	C2	0.41	3.4	0.25	0.89	6.8	0.51	7.7	0.47	1.43	/	/	/
	1RPX_A	D3	1H1Y	C2	0.41	3.5	0.29	0.83	4.2	0.41	2.5	0.60	1.47	/	/	/
	1RPX_B	D3	1H1Y	C2	0.41	3.5	0.29	0.83	4.2	0.41	17.2	0.98	1.68	/	/	/
	2FLI_A	D3	1TQX	C2	0.38	1.0	0.34	0.71	4.2	1.46	2.6	0.23	0.59	/	/	/
	2FLI_B	D3	1TQX	C2	0.38	1.0	0.34	0.71	4.2	1.46	17.1	2.90	2.04	/	/	/
	1TQJ_A	D3	1TQX	C2	0.36	3.3	0.30	1.29	4.6	1.23	4.8	0.49	1.48	/	/	/
	1TQJ_B	D3	1TQX	C2	0.36	3.3	0.30	1.29	4.6	1.23	19.6	2.78	1.92	/	/	/
	1RPX_A	D3	1TQX	C2	0.36	3.3	0.45	1.19	2.5	0.67	2.6	0.43	1.39	/	/	/
	1RPX_B	D3	1TQX	C2	0.36	3.3	0.45	1.19	2.5	0.67	14.7	3.32	2.58	/	/	/

Supplementary Table 4 Geometric parameters calculated by comparisons of different crystal structures (with different crystallisation space group and/or different ligands) of the same protein. For all eleven sub(families) at least one control pair was obtained. The aim of this type of control is to help distinguish geometric variation from real evolutionary differences between homologues of the family (values given in Table S3) and allosteric changes (in cases of different biological ligands) or structural flexibility of the protein (potentially sampled by different crystal forms). The quality of this control depends on the range of different structures in the PDB database, but from the available data we can conclude that, for each of the eleven (sub)families, geometric variations connected with evolutionary change in the oligomeric state are always larger than the geometric variations in its control set.

Part 1

SCOP family	code1	sym1	crystal1	ligands1	code2	sym2	crystal2	ligands2	dRot	dTrans	dRMSD	sRot	sTrans
54118	1O7Y	D2	C2	SO4	1O7Z	C2	P41212	free	7.1	0.42	0.84	11.2	1.03
	1O7Y	D2	C2	SO4	1O80	C2	P6522	free	14.0	1.19	0.75	7.2	0.70
	1O7Z	C2	P41212	free	1O80	C2	P6522	free	9.1	0.91	0.42	9.1	1.65
51751	1N7G	D2	P212121	GDR, NDP	1N7H	D2	C2221	GDP, NDP	1.5	0.21	0.28	0.9	0.24
53255	1QHF	D2	C2	3PG, SO4	5PGM	D2	P21	ALA, SO4	1.6	0.12	0.19	0.6	0.31
	1QHF	D2	C2	3PG, SO4	4PGM	D2	P21	free	1.4	0.22	0.21	0.2	0.34
	1QHF	D2	C2	3PG, SO4	1BQ4	D2	P21	BHC, SO4	1.4	0.16	0.26	0.3	0.24
	5PGM	D2	P21	ALA, SO4	4PGM	D2	P21	free	0.3	0.11	0.16	0.5	0.26
	5PGM	D2	P21	ALA, SO4	1BQ4	D2	P21	BHC, SO4	0.7	0.05	0.24	0.8	0.14
	4PGM	D2	P21	free	1BQ4	D2	P21	BHC, SO4	0.4	0.13	0.23	0.4	0.19
53272	2GIB	D2	P21212	EDO, U5P	1NON	D2	C2	free	1.5	0.17	0.32	0.1	1.10
	2GIB	D2	P21212	EDO, U5P	1XZ8	D2	H32	MG, 5GP, 3GP, U5P	2.4	0.21	0.30	0.7	2.32
	1XZ8	D2	H32	MG, 5GP, 3GP, U5P	1NON	D2	C2	free	1.7	0.23	0.16	0.9	2.40
	1XZ8	D2	H32	MG, 5GP, 3GP, U5P	1XZN	D2	P41212	MG, SO4	2.9	0.33	0.48	0.3	2.89
	1NON	D2	C2	free	1XZN	D2	P41212	MG, SO4	1.6	0.16	0.22	0.6	2.92
	2GIB	D2	P21212	EDO, U5P	1XZN	D2	P41212	MG, SO4	3.0	0.13	0.28	0.5	3.45
	1A3C	C2	C2	SO4, SM	1A4X	C2	H32	SO4, SM	2.8	0.15	0.27	0.1	2.79
	1BD3	D2	P21	PO4	1BD4	D2	P21	PO4, URA	0.2	0.04	0.23	0.2	0.15
	1BD3	D2	P21	PO4	1UPF	D2	P21	SO4, URF	0.7	0.03	0.16	0.4	0.13
	1BD3	D2	P21	PO4	1UPU	D2	P21	PO4, U5P	0.2	0.13	0.20	0.2	0.07
	1BD4	D2	P21	PO4, URA	1UPF	D2	P21	SO4, URF	0.6	0.10	0.18	0.3	0.14
	1BD4	D2	P21	PO4, URA	1UPU	D2	P21	PO4, U5P	0.2	0.09	0.18	0.2	0.14
	1UPF	D2	P21	SO4, URF	1UPU	D2	P21	PO4, U5P	0.6	0.06	0.18	0.4	0.15
53417	1RV3	D2	P41212	GLY, PLP, PO4	1RVU	D2	P41212	PLP, PO4	0.5	0.06	0.26	0.2	0.15
	1RV3	D2	P41212	GLY, PLP, PO4	1LS3	D2	P41	GLY, GOL, PLP, TGF	0.1	0.10	0.42	0.8	0.44
	1RV3	D2	P41212	GLY, PLP, PO4	1CJ0	D2	P41212	PLP	0.8	0.21	0.45	1.4	0.31
	1RV3	D2	P41212	GLY, PLP, PO4	1RV4	D2	P41212	PLP, PO4	0.3	0.06	0.17	0.3	0.11
	1RV3	D2	P41212	GLY, PLP, PO4	1RVY	D2	P41212	PLG, PLP, PO4	0.5	0.95	0.31	0.2	0.11
	1RVU	D2	P41212	PLP, PO4	1LS3	D2	P41	GLY, GOL, PLP, TGF	0.4	0.04	0.29	0.9	0.30
	1RVU	D2	P41212	PLP, PO4	1CJ0	D2	P41212	PLP	1.1	0.19	0.44	1.3	0.26
	1RVU	D2	P41212	PLP, PO4	1RV4	D2	P41212	PLP, PO4	0.6	0.05	0.28	0.4	0.06
	1RVU	D2	P41212	PLP, PO4	1RVY	D2	P41212	PLG, PLP, PO4	0.4	0.09	0.15	0.2	0.08
	1LS3	D2	P41	GLY, GOL, PLP, TGF	1CJ0	D2	P41212	PLP	0.8	0.20	0.46	1.0	0.48
	1LS3	D2	P41	GLY, GOL, PLP, TGF	1RV4	D2	P41212	PLP, PO4	0.3	0.07	0.41	0.5	0.35
	1LS3	D2	P41	GLY, GOL, PLP, TGF	1RVY	D2	P41212	PLG, PLP, PO4	0.4	0.12	0.32	0.9	0.35
	1CJ0	D2	P41212	PLP	1RV4	D2	P41212	PLP, PO4	0.8	0.24	0.48	1.1	0.28
	1CJ0	D2	P41212	PLP	1RVY	D2	P41212	PLG, PLP, PO4	1.1	0.21	0.45	1.4	0.31
	1RV4	D2	P41212	PLP, PO4	1RVY	D2	P41212	PLG, PLP, PO4	0.7	0.15	0.16	0.3	0.11

Supplementary Table 4 - continued

Part 2

SCOP family	code1	sym1	crystal1	ligands1	code2	sym2	crystal2	ligands2	dRot	dTrans	dRMSD	sRot	sTrans
51352	1IF2	C2	C2	129	1QDS	C2	C2	PGA	0.2	0.01	0.08	0.1	0.03
	1IF2	C2	C2	129	1AMK	C2	C2	PGA	0.7	0.05	0.18	0.2	0.05
	1QDS	C2	C2	PGA	1AMK	C2	C2	PGA	0.2	0.04	0.21	0.2	0.04
	1N55	C2	C2	ACY, GOL, PGA	1AMK	C2	C2	PGA	1.7	0.12	0.28	0.6	0.12
	1N55	C2	C2	ACY, GOL, PGA	1QDS	C2	C2	PGA	1.6	0.09	0.19	2.7	0.61
	1N55	C2	C2	ACY, GOL, PGA	1IF2	C2	C2	129	1.6	0.09	0.20	3.0	0.76
	1R2R_1	C2	P212121	DMS, MG, TRS	1R2S_1	C2	P212121	free	1.0	0.10	0.21	0.1	0.61
	1R2S_2	C2	P212121	free	1HTI	C2	P212121	PGA	0.4	0.14	0.35	0.5	0.41
	1R2S_1	C2	P212121	free	1HTI	C2	P212121	PGA	1.7	0.16	0.34	1.1	0.49
	1R2R_1	C2	P212121	DMS, MG, TRS	1HTI	C2	P212121	PGA	0.7	0.13	0.28	1.1	0.15
	1HTI	C2	P212121	PGA	1WYI	C2	P212121	free	0.4	0.08	0.33	1.2	0.31
	1R2R_1	C2	P212121	DMS, MG, TRS	1R2S_2	C2	P212121	free	0.4	0.07	0.21	1.5	0.47
	1R2S_1	C2	P212121	free	1R2S_2	C2	P212121	free	1.3	0.03	0.07	1.6	0.30
	1R2R_2	C2	P212121	DMS, MG, TRS	1HTI	C2	P212121	PGA	0.9	0.06	0.28	1.7	0.65
	1R2R_2	C2	P212121	DMS, MG, TRS	1R2S_2	C2	P212121	free	0.7	0.09	0.24	1.9	0.75
	1R2R_2	C2	P212121	DMS, MG, TRS	1WYI	C2	P212121	free	0.5	0.09	0.20	2.0	0.79
	1R2S_2	C2	P212121	free	1WYI	C2	P212121	free	0.2	0.14	0.34	2.1	0.85
	1R2R_2	C2	P212121	DMS, MG, TRS	1R2S_1	C2	P212121	free	1.3	0.11	0.23	2.3	0.75
	1R2S_1	C2	P212121	free	1WYI	C2	P212121	free	0.8	0.30	0.35	2.4	1.03
	1R2R_1	C2	P212121	DMS, MG, TRS	1WYI	C2	P212121	free	0.5	0.10	0.22	2.5	0.83
	1R2R_1	C2	P212121	DMS, MG, TRS	1R2R_2	C2	P212121	DMS, MG, TRS	0.7	0.10	0.10	3.7	1.01
51570	1J2W	D2	P212121	free	1UB3	D2	P212121	HPD	0.1	0.07	0.09	0.3	0.13
	3R12	C2	P21	CIT, GOL, PGO	3R13	C2	P21	ACT, GOL	0.7	0.06	0.17	0.3	0.14
46610;54720	1IXB	C2	P21	MH2	1IX9	C2	P21	MN	1.4	0.34	0.59	0.4	0.19
	1IXB	C2	P21	MH2	1VEW	C2	C2221	MN, OH	6.4	0.57	0.21	6.8	0.29
	1IX9	C2	P21	MN	1VEW	C2	C2221	MN, OH	6.7	0.68	0.22	6.9	0.45
	3MDS	D2	P41212	MN3	1MNG	D2	P41212	AZI, MN	0.2	0.07	0.07	0.2	0.13
75176;75446	1O8B	C2	P1	ABF	1KS2	C2	P1	free	0.6	0.06	0.10	0.4	0.23
	1O8B	C2	P1	ABF	1LKZ	C2	C2221	free	1.4	0.16	0.22	2.6	0.33
	1KS2	C2	P1	free	1LKZ	C2	C2221	free	2.1	0.19	0.31	2.5	0.28
	1LK5	D2	P212121	CL, NA	1LK7	D2	P212121	CL, DER, NA	0.5	0.17	0.19	0.7	0.16
51372	2FLIA	D3	P21	DX5, ZN	2FLIB	D3	P21	DX5, ZN	0.5	0.01	0.12	1.0	0.30
	1H1Y	C2	P21	SO4	1H1Z	C2	P21	SO4, ZN	1.3	0.14	0.23	0.6	0.30

Supplementary Table 5 Geometric differences between two homologues are larger than between different crystal structures (in different crystal forms and with different ligands) of the same protein. The table shows structural pairwise comparisons of all available crystal forms of the PyrR proteins from *B. caldolyticus* (in blue) and *B. subtilis* (in magenta). Pairs in the table have been ordered by increasing angle of rotation between subunits around the dimeric interface (sRot).

PDB code		Space group		Ligands		sTrans	sRot
Comparing different crystal structures of the same protein							
2GIB	1NON	$P2_12_12$	C2	EDO, U5P	free	**0.14**	**1.10**
2GIB	1XZ8	$P2_12_12$	H32	EDO, U5P	MG, 5GP, 3GP, U5P	**0.73**	**2.32**
1XZ8	1NON	H32	C2	MG, 5GP, 3GP, U5P	free	**0.88**	**2.40**
1A3C	1A4X	C2	H32	SO4, SM	SO4	**0.15**	**2.79**
1XZ8	1XZN	H32	$P4_12_12$	MG, 5GP, 3GP, U5P	MG, SO4	**0.33**	**2.89**
1NON	1XZN	C2	$P4_12_12$	free	MG, SO4	**0.57**	**2.92**
2GIB	1XZN	$P2_12_12$	$P4_12_12$	EDO, U5P	MG, SO4	**0.47**	**3.45**
Comparing structures of homologous proteins							
1XZ8	1A3C	H32	C2	MG, 5GP, 3GP, U5P	SO4, SM	**1.64**	**6.01**
1XZN	1A3C	$P4_12_12$	C2	MG, SO4	SO4, SM	**1.62**	**6.10**
2GIB	1A3C	$P2_12_12$	C2	EDO, U5P	SO4, SM	**1.90**	**7.50**
1NON	1A3C	C2	C2	free	SO4, SM	**2.28**	**8.18**
1XZN	1A4X	$P4_12_12$	H32	MG, SO4	SO4	**2.03**	**8.93**
1XZ8	1A4X	H32	H32	MG, 5GP, 3GP, U5P	SO4	**1.78**	**10.43**
1NON	1A4X	C2	H32	free	SO4	**2.34**	**11.13**
2GIB	1A4X	$P2_12_12$	H32	EDO, U5P	SO4, SM	**2.91**	**12.85**
BcPyrR structures				BsPyrR structures			

Supplementary Table 6 Relative accessible surface area (A_{rel}) of dimers and dimeric subcomplexes of all eleven families. A_{rel} is a ratio between a protein accessible surface area and the accessible surface area predicted for a protein of its molecular weight (9). Proteins with high (higher than 1) A_{rel} of bound conformations are predicted to have undergone conformational changes upon binding and proteins whose unbound conformations have high A_{rel} values are predicted to be more flexible than average.

	A_{rel}				Model
Family	mean	min	max	sd	
Triosephosphate isomerase	0.879	0.819	0.914	0.036	IV
D-ribulose-5-phosphate 3-epimerase	0.902	0.863	0.939	0.031	III/IV
Class I aldolase	0.903	0.852	0.961	0.054	IV
D-ribose-5-phosphate isomerase	0.925	0.889	0.965	0.038	III/IV
GABA-aminotransferase-like	0.925	0.852	0.989	0.069	III/IV
Fe,Mn superoxide dismutase	0.940	0.899	0.971	0.035	IV
Cofactor-dependent phosphoglycerate mutase	0.980	0.917	1.027	0.057	III
Tyrosine-dependent oxidoreductases	0.993	0.973	1.011	0.019	III
UPRTase	0.997	0.957	1.061	0.056	III/IV
PyrR	1.004	0.972	1.032	0.030	III
Interleukin 8-like chemokines	1.106	1.079	1.147	0.031	III

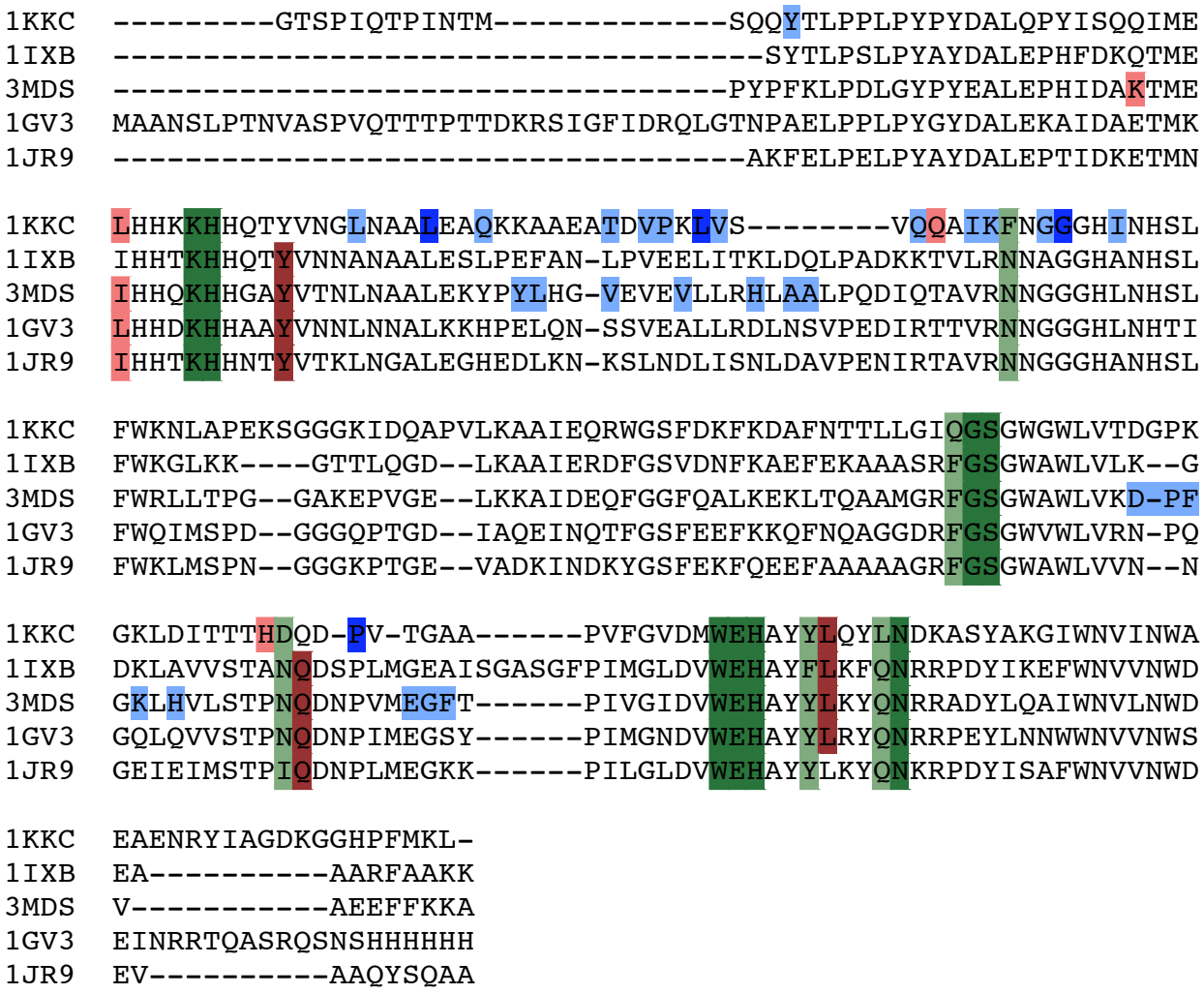
References

1. Murzin AG, Brenner SE, Hubbard T, & Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247(4):536-540.
2. Levy ED, Pereira-Leal JB, Chothia C, & Teichmann SA (2006) 3D complex: a structural classification of protein complexes. *PLoS Comput Biol* 2(11):e155.
3. Levy ED, Boeri Erba E, Robinson CV, & Teichmann SA (2008) Assembly reflects evolution of protein complexes. *Nature* 453(7199):1262-1265.
4. Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774-797.
5. Lesk AM ed (1986) *Integrated access to sequence and structural data. In Biosequences: perspectives and user services in Europe*), pp 23–28.
6. Tsai J, Taylor R, Chothia C, & Gerstein M (1999) The packing density in proteins: standard radii and volumes. *J Mol Biol* 290(1):253-266.

7. Hubbard SJT, J.M. (1993) 'NACCESS').
8. Levy ED (2010) A simple definition of structural regions in proteins and its use in analyzing interface evolution. *J Mol Biol* 403(4):660-670.
9. Marsh Joseph A & Teichmann Sarah A (2011) Relative Solvent Accessible Surface Area Predicts Protein Conformational Changes upon Binding. *Structure* 19(6):859-867.

Family Alignments

Fe, Mn superoxide dismutase (MnSOD)



GABA-aminotransferase-like

```
1KKP  ----------------------MKYLPQQDPQVFAAIEQERKRQHAKIELIASENFVSRAV
1RV3  ATAVNGAPRDAALWSSHEQMLAQPLKDSDAEVYDIIKKESNRQRVGLELIASENFASRAV
1EJI  -----MADRDATLWASHEKMLSQPLKDSDAEVYSIIKKESNRQRVGLELIASENFASRAV

1KKP  MEAQGSVLTNKYAEGYPGRRYYGGCEYVDIVEELARERAKQLFGAEH----ANVQPHSGA
1RV3  LEALGSCLNNKYSLGYPGQRYYGGTEHIDELETLCQKRALQAYGLDPQCWGVNVQPYSGS
1EJI  LEALGSSLNNKYSEGYPGQRYYGGTEFIDELEMLCQKRALQAYHLDPQCWGVNVQPYSGS

1KKP  QANMAVYFTVLEHGDTVLGMNLSHGGHLTHG-----SPVNFSGVQYNFVAYGVDPETHVI
1RV3  PANFAVYTALVEPHGRIMGLDLPDGGHLTHGFMTDKKKISATSIFFESMAYKVNPDTGYI
1EJI  PANFAVYTALVEPHGRIMGLDLPDGGHLTHGFMTDKKKISATSIFFESMPYKVYPETGYI

1KKP  DYDDVREKARLHRPKLIVAAASAYPRIIDFAKFREIADEVGAYLMVDMAHIAGLVAAGLH
1RV3  DYDRLEENARLFHPKLIIAGTSCYSRNLDYGRLRKIADENGAYLMADMAHISGLVVAGVV
1EJI  NYDQLEENASLFHPKLIIAGTSCYSRNLDYARLRKIADDNGAYLMADMAHISGLVAAGVV

1KKP  PNPVPYAHFVTTTTHKTLRGPRGGMILCQE---------------QFAKQIDKAIFPGIQ
1RV3  PSPFEHCHVVTTTTHKTLRGCRAGMIFYRRGVRSVDPKTGKEILYNLESLINSAVFPGLQ
1EJI  PSPFEHCHVVTTTTHKTLRGCRAGMIFYRKGVRSVDPKTGKETYYELESLINSAVFPGLQ

1KKP  GGPLMHVIAAKAVAFGEALQDDFKAYAKRVVDNAKRLASALQNEGFTLVSGGTDNHLLLV
1RV3  GGPHNHAIAGVAVALKQAMTPEFKEYQRQVVANCRALSAALVELGYKIVTGGSDNHLILV
1EJI  GGPHNHAIAGVAVALKQAMTTEFKIYQLQVLANCRALSDALTELGYKIVTGGSDNHLILM

1KKP  DLRPQQLTGKTAEKVLDEVGITVNKNTIPYDPESPFVTSGIRIGTAAVTTRGFGLEEMDE
1RV3  DLRSKGTDGGRAEKVLEACSIACNKNTCPGD-KSALRPSGLRLGTPALTSRGLLEKDFQK
1EJI  DLRSKGTDGGRAEKVLEACSIACNKNTCPGD-KSALRPSGLRLGTPALTSRGLLEEDFQK

1KKP  IAAII--GLVL-----KNVGSEQALEEARQRVAALTDPTSRSAAGTMEFEA---------
1RV3  VAHFIHRGIELTVQIQDDTGPRATLKEFKEKLAG--DEKHQRAVRALRQEVESFAALFPL
1EJI  VAHFIHRGIELTLQIQSHMATKATLKEFKEKLAG--DEKIQSAVATLREEVENFASNFSL

1KKP  ------
1RV3  PGLPGF
1EJI  PGLPDF
```

Phosphorybosyltransferases (UPRTase)

```
1BD3  AQVPASGKLLVDPRYSTNDQEESILQDIITRFPNVVLMKQTAQLRAMMTIIRDKETPKE
1O5O  ------------------MGSDKIHHHHHHMKNLVVVDH-PLIKHKLTIMRDKNTGPKE
1I5E  ------------------MG-----------KVYVFDH-PLIQHKLTYIRDKNTGTKE

1BD3  F FYADRLIR LI EALNELPFQKKEVTTPLDVSYHGVSFYSKICGVSIVRAGESMESGL
1O5O  FRELL EITLLLAYEATRHLKCEEVEVETPITKTIGYRINDKDIVVVPILRAGLVMADGI
1I5E  FRELVDEVATLMAFEITRDLPLEEVEIETPVSKARAKVIAGKKLGVIPILRAGIGMVDGI

      V     I  E
1BD3  RAVCRGVRIGKILIQRDETTAEPKLIYEKLPADIRERWVMLLDPMCATAGSVCKAIEVLL
1O5O  LELLPNASVGHIGIYRDPETLQAVEYYAKLPPLNDDKEVFLLDPMLATGVSSIKAIEILK
1I5E  LKLIPAAKVGHIGLYRDPQTLKPVEYYVKLPSDVEERDFIIVDPMLATGGSAVAAIDALK

1BD3  RLGVKEERIIFVNILAAPQGIERVFKEYPKVRMVTAAVDICLNSRYYIVPGIGDFGDRYF
1O5O  ENGAK--KITLVALIAAPEGVEAVEKKYEDVKIYVAALDERLNDHGYIIPGLGDAGDRLF
1I5E  KRGAK--SIKFMCLIAAPEGVKAVETAHPDVDIYIAALDERLNDHGYIVPGLGDAGDRLF

1BD3  GTM
1O5O  RTK
1I5E  GTK
```

Class I aldolase

```
1N7K  PSARDILQQGLDRLGS----------------------PEDLASRIDSTLLSPRATEED
1VCV  ----------------------------------------MIHLVDYALLKPYLTVDE
1MZH  ----------------------------------------MIDVRKYIDNAALKPHLSEKE
1O0Y  MGSDKIHHHHHHMIEYRIEEAVAKYREFYEFKPVRESAGIEDVKSAIEHTNLKPFATPDD
1J2W  ----------------------------------------MDLAAHIDHTLLKPTATLEE

1N7K  VRNLVREASDYGFRCAVLTPVYTVKISGLAEKLGVKLCSVIGFPLGQAPLEVKLVEAQTV
1VCV  AVAGARKAEELGVAAYCVNPIYAPVVRPL--LRKVKLCVVADFPFGALPTASRI-ALVSR
1MZH  IEEFVLKSEELGIYAVCVNPYHVKLASSI--AKKVKVCCVIGFPLGLNKTSVKVKEAVEA
1O0Y  IKKLCLEARENRFHGVCVNPCYVKLAREELEGTDVKVVTVVGFPLGANETRTKAHEAIFA
1J2W  VAKAAEEALEYGFYGLCIPPSYVAWVRARYPHAPFRLVTVVGFPLGYQEKEVKALEAALA

1N7K  LEAGATELDVVPHLSLGP----EAVYREVSGIVKLAKSYGAVVKVILEAPLWDDKTLSLL
1VCV  LAEVADEIDVVAPIGLVKSRRWAEVRRDLISVVGAAG--GRVVKVITEEPYLRDEERYTL
1MZH  VRDGAQELDIVWNLSAFKSEKYDFVVEELKEIFRETP--SAVHKVIVETPYLNEEEIKKA
1O0Y  VESGADEIDMVINVGMLKAKEWEYVYEDIRSVVESVK--GKVVKVIIETCYLDTEEKIAA
1J2W  CARGADEVDMVLHLGRAKAGDLDYLEAEVRAVREAVP--QAVLKVILETGYFSPEEIARL

1N7K  VDSSRRAGADIVKTSTGVY----------TKGGDPVTVFRLASLA--KPLGMGVKASGGI
1VCV  YDIIAEAGAHFIKSSTGFAEEAYAARQGNPVHSTPERAAAIARYIKEKGYRLGVKMAGGI
1MZH  VEICIEAGADFIKTSTGFA----------PRGTTLEEVRLIKSSA--KG-RIKVKASGGI
1O0Y  CVISKLAGAHFVKTSTGFG----------TGGATAEDVHLMKWIV---GDEMGVKASGGI
1J2W  AEAAIRGGADFLKTSTGFG----------PRGASLEDVALLVRVA--QG-RAQVKAAGGI

1N7K  RSGIDA---VLAVGAGAD----IIGTSSAVKVLESFKSLV--------
1VCV  RTREQAKAIVDAIGWGEDPARVRLGTSTPEALL--------------
1MZH  RDLETA---ISMIEAGAD----RIGTSSGISIAEEFLKRHLILEHHHH
1O0Y  RTFEDA---VKMIMYGAD----RIGTSSGVKIVQGGEERYGG------
1J2W  RDRETA---LRMLKAGAS----RLGTSSGVALVAGEGGTLGY------
```

D-ribulose-5-phosphate 3-epimerase

```
1TQJ  ------MSKNIVVAPSILSADFSRLGEEIKAVDEAGADWIHVDVMDGRFVPNITIGPLIV
1RPX  SRVDKFSKSDIIVSPSILSANFSKLGEQVKAIEQAGCDWIHVDVMDGRFVPNITIGPLVV
1TQX  -----MGTLKAIIAPSVLASNISKLAEETQRMESLGAEWIHLDVMDMHFVPNLSFGPPVI
2FLI  -------MSTLKIAPSILAADYANFASELARIEETDAEYVHIDIMDGQFVPNISFGADVV
1H1Y  ----MAAAAAAKIAPSMLSSDFANLAAEADRMVRLGADWLHMDIMDGHFVPNLTIGAPVI

1TQJ  DAIRPLTK-KTLDVHLMIVEPEKYVEDFAKAGADIISVHVEHNASPHLHRTLC---QIRE
1RPX  DSLRPITD-LPLDVHLMIVEPDQRVPDFIKAGADIVSVHCEQSSTIHLHRTIN---QIKS
1TQX  NNLKKYTKSIFFDVHLMVEYPEKYV-PLLKT-SNQLTFHFE-ALNEDTERCIQLAKEIRD
2FLI  ASMRKHSK-LVFDCHLMVVDPERYVEAFAQAGADIMTIHTE--STRHIHGALQ---KIKA
1H1Y  QSLRKHTK-AYLDCHLMVTNPSDYVEPLAKAGASGFTFHIE-VSRDNWQELIQ---SIKA

1TQJ  LGKKAGAVLNPSTPLDFLEYVLPV---CDLILIMSVNPGFGGQSFIPEVLPKIRALRQMC
1RPX  LGAKAGVVLNPGTPLTAIEYVLDA---VDLVLIMSVNPGFGGQSFIESQVKKISDLRKIC
1TQX  NNLWCGISIKPKTDVQKLVPILDT-NLINTVLVMTVEPGFGGQSFMHDMMGKVSFLRK--
2FLI  AGMKAGVVINPGTPATALEPLLDL---VDQVLIMTVNPGFGGQAFIPECLEKVATVAKWR
1H1Y  KGMRPGVSLRPGTPVEEVFPLVEAENPVELVLVMTVEPGFGGQKFMPEMMEKVRALRK--

1TQJ  DERGLDPWIEVDGGLKPNNTWQVLEAGANAIVAGSAVFNAPNYAEAIAGVRNSKRPEPQL
1RPX  AERGLNPWIEVDGGVGPKNAYKVIEAGANALVAGSAVFGAPDYAEAIKGIKTSKRPE---
1TQX  --KYKNLNIQVDGGLNIETTEISASHGANIIVAGTSIFNAEDPKYVIDTMRVSVQKYLNN
2FLI  DEKGLSFDIEVDGGVDNKTIRACYEAGANVFVAGSYLFKASDLVSQVQTLRTALNV----
1H1Y  --KYPSLDIEVDGGLGPSTIDVAASAGANCIVAGSSIFGAAEPGEVISALRKSVEGSQNK

1TQJ  ATV
1RPX  ---
1TQX  ---
2FLI  ---
1H1Y  S--
```

Interleukin 8-like chemokines

```
3IL8  -----SAKELRCQCIKTYSKPFHPKFIKELRVIESGPHCANTEIIVKLSDGRELCLDPKE
1TVX  --DSDLYAELRCLCIKTTSG-IHPKNIQSLEVIGKGTHCNQVEVIATLKDGRKICLDPDA
1PLF  DSEGGEDEDLQCVCLKTTSG-INPRHISSLEVIGAGLHCPSPQLIATLKTGRKICLDQQN
1F9S  --EAEEDGDLQCLCVKTTSQ-VRPRHITSLEVIKAGPHCPTAQLIATLKNGSKICLDLQA

3IL8  NWVQRVVEKFLKRAENS-
1TVX  PRIKKIVQKKLAGDESAD
1PLF  PLYKKIIKRLLKS-----
1F9S  PLYKKIIKKLLES-----
```

Tyrosine-dependent oxidoreductases

```
1N7G  MASENNGSRSDSESITAPKADSTVVEPRKIALITGITGQDGSYLTEFLLGKGYEVHGLIR
1T2A  ----MGSSHHHHHHSSGRENKYFQGHMRNVALITGITGQDGSYLAEFLLEKGYEVHGIVR
1DB3  --------------------------SKVALITGVTGQDGSYLAEFLLEKGYEVHGIKR

1N7G  RSSNFNTQRINHIYIDPHNVNKALMKLHYADLTDASSLRRWIDVIKPDEVYNLAAQSHVA
1T2A  RSSSFNTGRIEHLYKNPQAHIEGNMKLHYGDLTDSTCLVKIINEVKPTEIYNLGAQSHVK
1DB3  RASSFNTERVDHIYQDPHTCNP-KFHLHYGDLSDTSNLTRILREVQPDEVYNLGAMSHVA

1N7G  VSFEIPDYTADVVATGALRLLEAVRSHTIDSGRTVKYYQAGSSEMFGSTPP-PQSETTPF
1T2A  ISFDLAEYTADVDGVGTLRLLDAVKTCGLI--NSVKFYQASTSELYGKVQEIPQKETTPF
1DB3  VSFESPEYTADVDAMGTLRLLEAIRFLGLE--KKTRFYQASTSELYGLVQEIPQKETTPF

1N7G  HPRSPYAASKCAAHWYTVNYREAYGLFACNGILFNHESPRRGENFVTRKITRALGRIKVG
1T2A  YPRSPYGAAKLYAYWIVVNFREAYNLFAVNGILFNHESPRRGANFVTRKISRSVAKIYLG
1DB3  YPRSPYAVAKLYAYWITVNYRESYGMYACNGILFNHESPRRGETFVTRKITRAIANIAQG

1N7G  LQTKLFLGNLQASRDWGFAGDYVEAMWLMLQQEKPDDYVVATEEGHTVEEFLDVSFGYLG
1T2A  QLECFSLGNLDAKRDWGHAKDYVEAMWLMLQNDEPEDFVIATGEVHSVREFVEKSFLHIG
1DB3  LESCLYLGNMDSLRDWGHAKDYVKMQWMMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLG

1N7G  --LNWKD-------------------------YVEIDQRYFREVDNLQGDASKAKEVL
1T2A  KTIVWEGKNENEVGRC----------KETGKVHVTVDLKYYRPTEVDFLQGDCTKAKQKL
1DB3  IKLRFEGTGVEEKGIVVSVTGHDAPGVKPGDVIIAVDPRYFRPAEVETLLGDPTKAHEKL

1N7G  GWKPQVGFEKLVKMMVDEDLELAKREKVLVDAGYMDAKQQPLEHHHHHH
1T2A  NWKPRVAFDELVREMVHADVELMRTNP---NAGS---------------
1DB3  GWKPEITLREMVSEMVANDLEAAKKHSLLKSHGYDVAIALES-------
```

Cofactor-dependent phosphoglycerate mutase

```
2HHJ  MSKYKLIMLRHGEGAWNKENRFCSWVDQKLNSEGMEEARNCGKQLKALNFEFDLVFTSVL
1E58  -AVTKLVLVRHGESQWNKENRFTGWYDVDLSEKGVSEAKAAGKLLKEEGYSFDFAYTSVL
1QHF  ---PKLVLVRHGQSEWNEKNLFTGWVDVKLSAKGQQEAARAGELLKEKKVYPDVLYTSKL

2HHJ  RSHTAWLILEELGQEWVPVESSWRLNERHYGALIGLNREQMALNHGEEQVRLWRRSYN
1E58  KRAIHTLWNVLDELDQAWLPVEKSWKLNERHYGALQGLNKAETAEKYGDEQVKQWRRGFA
1QHF  SRAIQTANIALEKADRLWIPVNRSWRLNERHYGDLQGKDKAETLKKFGEEKFNTYRRSFD

2HHJ  VTPPPIEESHPYYQEIYNDRRYKVCDVPLDQLPRSESLKDVLERLLPYWNERIAPEVLRG
1E58  VTPPELTKDDERYPG--HDPRYAKLSE--KELPLTESLALTIDRVIPYWNETILPRMKSG
1QHF  VPPPPIDASSPFSQK--GDERYKYVDP--NVLPETESLALVIDRLLPYWQDVIAKDLLSG

2HHJ  KTILISAHGNSSRALLKHLEGISDEDIINITLPTGVPILLELDENLRAVGPHQFLGDQEA
1E58  ERVIIAAHGNSLRALVKYLDNMSEEEILELNIPTGVPLVYEFDENFKPLKRY-YLGNADE
1QHF  KTVMIAAHGNSLRGLVKHLEGISDADIAKLNIPTGIPLVFELDENLKPSKPS-YYLDPEA

2HHJ  IQAAIKKVEDQGKVKQAKKLEHHHHHH
1E58  IAAKAAAVANQGKAK-----------
1QHF  AAAGAAAV-----------------
```

Phosphorybosyltransferases (PyrR)

```
1W30  MGAAGDAAIGRESRELMSAANVGRTISRIAHQIIEKTALDDPVGPDAPRVVLLGIPTRGV
1NON  ----------MQKAVVMDEQAIRRALTRIAHEIIERN-------KGIDGCVLVGIKTRGI
1A3C  ---------MNQKAVILDEQAIRRALTRIAHEMIERN-------KGMNNCILVGIKTRGI

1W30  TLANRLAGNITEYSGIHVGHGALDITLYRDDLMIKPP--RPLA-STSIPAGGIDDALVIL
1NON  YLARRLAERIEQIEGASVPVGELDITLYRRDDLTVKTDDHEPLVKGTNVPF-PVTERNVIL
1A3C  YLAKRLAERIEQIEGNPVTVGEIDITLYRDDLSKKTSNDEPLVKGADIPV-DITDQKVIL

1W30  VDDVLYSGRSVRSALDALRDVGRPRAVQLAVLVDRGHRELPLRADYVGKNVPTSRSESVH
1NON  VDDVLFTGRTVRAAMDAVMDLGRPARIQLAVLVDRGHRELPIRADFVGKNVPTSRSELIV
1A3C  VDDVLYTGRTVRAGMDALVDVGRPSSIQLAVLVDRGHRELPIRADYIGKNIPTSKSEKVM

1W30  VRLREHDGRDGVVISRGSHHHHHH
1NON  VELSEVDGIDQVSI------HEK-
1A3C  VQLDEVDQNDLVAI------YENE
```

D-ribose-5-phosphate isomerase

```
1LK5  MNVEEMKKIAAKEALKFIEDDMVIGLGTGSTTAYFIKLLGEKLKRGEISDIVGVPTSYQA
1O8B  MTQDELKKAVGWAALQYVQPGTIVGVGTGSTAAHFIDALGT--MKGQIEGAVSSSDASTE
1M0S  MNQLEMKKLAAQAALQYVKADRIVGVGSGSTVNCFIEALGT--IKDKIQGAVAASKESEE

1LK5  KLLAIEHDIPIASLDQVDAIDVAVDGADEVDPNLNLIKGRGAALTMEKIIEYRAGTFIVL
1O8B  KLKSL--GIHVFDLNEVDSLGIYVDGADEINGHMQMIKGGGAALTREKIIASVAEKFICI
1M0S  LLRKQ--GIEVFNANDVSSLDIYVDGADEINPQKMMIKGGGAALTREKIVAALAKKFICI

1LK5  VDERKLVDYLCQKMPVPIEVIPQAWKAIIEELSIFNAKAELRMGVNKDGPVITDNGNFII
1O8B  ADASKQVDILG-KFPLPVEVIPMARSAVARQLVKLGGRPEYRQG------VVTDNGNVIL
1M0S  VDSSKQVDVLGSTFPLPVEVIPMARSQVGRKLAALGGSPEYREG------VVTDNGNVIL

1LK5  DAKFPRIDDPLDMEIELNTIPGVIENGIFADI-ADIVIVGTREGVKKLER
1O8B  DVHGMEILDPIAMENAINAIPGVVTVGLFANRGADVALIGTPDGVKTIVK
1M0S  DVHNFSILNPVEIEKELNNVAGVVTNGIFALRGADVVIVGTPEGAKVID-
```

Triosephosphate isomerase

```
1R2R  -APSRKFFVGGNWKMNGRKKNLGELITTLNAAKVPAD-TEVVCAPPTAYIDFARQKLD-P
1N55  MSAKPQPIAAANWKCNGTTASIEKLVQVFNEHTISHD-VQCVVAPTFVHIPLVQAKLRNP
1KV5  -MSKPQPIAAANWKCNGSQQSLSELIDLFNSTSINHD-VQCVVASTFVHLAMTKERLSHP
1B9B  --ITRKLILAGNWKMHKTISEAKKFVSLLVNELHDVKEFEIVVCPPFTALSEVGEILSGR
2BTM  ----RKPIIAGNWKMNGTLAEAVQFVEDVKGHVPPADEVISVVCAPFLFLDRLVQAADGT

1R2R  KIAVAAQNCYKVTNGAFTGEISPGMIKDCGATWVVLGHSERRHVFGESDELIGQKVAHAL
1N55  KYVISAQNAI-AKSGAFTGEVSMPILKDIGVHWVILGHSERRTYYGETDEIVAQKVSEAC
1KV5  KFVIAAQNAI-AKSGAFTGEVSLPILKDFGVNWIVLGHSERRAYYGETNEIVADKVAAAV
1B9B  NIKLGAQNVFYEDQGAFTGEISPLMLQEIGVEYVIVGHSERRRIFKEDDEFINRKVKAVL
2BTM  DLKIGAQTMHFADQGAYTGEVSPVMLKDLGVTYVILGHSERRQMFAETDETVNKKVLAAF

1R2R  SEGLGVIACIGEKLDEREAGITEKVVFEQTKVIADNVK--DWSKVVLAYEPVWAIGTGKT
1N55  KQGFMVIACIGETLQQREANQTAKVVLSQTSAIAAKLTKDAWNQVVLAYEPVWAIGTGKV
1KV5  ASGFMVIACIGETLQERESGRTAVVVLTQIAAIAKKLKKADWAKVVIAYEPVWAIGTGKV
1B9B  EKGMTPILCVGETLEEREKGLTFCVVEKQVREGFYGLDKEEAKRVVIAYEPVWAIGTGRV
2BTM  TRGLIPIICCGESLEEREAGQTNAVVASQVEKALAGLTPEQVKQAVIAYEPIWAIGTGKS

1R2R  ATPQQAQEVHEKLRGWLKSNVSDAVAQSTRIIYGGSVTGATCKELASQPDVDGFLVGGAS
1N55  ATPEQAQEVHLLLRKWVSENIGTDVAAKLRILYGGSVNAANAATLYAKPDINGFLVGGAS
1KV5  ATPQQAQEAHALISSWVSSKIGADVAGELRILYGGSVNGKNARTLYQQRDVNGFLVGGAS
1B9B  ATPQQAQEVHAFIRKLLSEMYDEETAGSIRILYGGSIKPDNFLGLIVQKDIDGGLVGGAS
2BTM  STPEDANSVCGHIRSVVSRLFGPEAAEAIRIQYGGSVKPDNIRDFLAQQQIDGALVGGAS

1R2R  LKPE-FVDIINAKQ----
1N55  LKPE-FRDIIDATR----
1KV5  LKPE-FVDIIKATQ----
1B9B  LKES-FIELARIMRGVIS
2BTM  LEPASFLQLVEAGRHE--
```

LEGEND

- Dimeric interface - common and conserved residues
- Dimeric interface - common and not conserved residues
- Dimeric interface - variable and conserved residues
- Dimeric interface - variable and not conserved residues
- Tetrameric/hexameric interface - conserved residues
- Tetrameric/hexameric interface - not conserved residues