BIOINFORMATICS ORIGINAL PAPER Vol. 22 no. 5 2006, pages 550–555
doi:10.1093/bioinformatics/bti782

Structural bioinformatics

Equivalent binding sites reveal convergently evolved interaction motifs

Andreas Henschel*, Wan Kyu Kim and Michael Schroeder

Bioinformatics Group, Biotechnological Centre, TU Dresden, Germany

Received on July 12, 2005; revised on November 10, 2005; accepted on November 12, 2005
Advance Access publication November 15, 2005
Associate Editor: Anna Tramontano

ABSTRACT

Motivation: Much research has been devoted to the characterization of interaction interfaces found in complexes with known structure. In this context, the interactions of non-homologous domains at equivalent binding sites are of particular interest, as they can reveal convergently evolved interface motifs. Such motifs are an important source of information to formulate rules for interaction specificity and to design ligands based on the common features shared among diverse partners.

Results: We develop a novel method to identify non-homologous structural domains which bind at equivalent sites when interacting with a common partner. We systematically apply this method to all pairs of interactions with known structure and derive a comprehensive database for these interactions. Of all non-homologous domains, which bind with a common interaction partner, 4.2% use the same interface of the common interaction partner (excluding immunoglobulins and proteases). This rises to 16% if immunoglobulin and proteases are included. We demonstrate two applications of our database: first, the systematic screening for viral protein interfaces, which can mimic native interfaces and thus interfere; and second, structural motifs in enzymes and its inhibitors. We highlight several cases of virus protein mimicry: viral M3 protein interferes with a chemokine dimer interface. The virus has evolved the motif SVSPLP, which mimics the native SSDTTP motif. A second example is the regulatory factor Nef in HIV which can mimic a kinase when interacting with SH3. Among others the virus has evolved the kinase's PxxP motif. Further, we elucidate motif resemblances in Baculovirus p35 and HIV capsid proteins. Finally, chymotrypsin is subject to scrutiny wrt. its structural similarity to subtilisin and wrt. its inhibitor's similar recognition sites.

Contact: ah@biotec.tu-dresden.de

Supplementary informaton: A database is online at scoppi.biotec.tu-dresden.de/abac/

INTRODUCTION

Protein interactions underlie all cellular processes and are important to reveal function. The interactions from known three-dimensional structures have been of particular interest in that they allow a number of detailed analyses of interfaces in terms of physico-chemical properties, shape and geometry [Jones and Thornton (1996); Conte *et al*. (1999); Bashton and Chothia (2002); Chakrabarti and Janin (2002); Nussinov *et al*. (1997); Ofran and Rost (2003)]. The rapid growth of multichain and multidomain structures in Protein Databank (PDB) [Berman *et al*. (2000)] enabled systematic analyses of domain–domain interactions and interfaces [Park *et al*. (2001); Bolser *et al*. (2003); Apic *et al*. (2001); Kim *et al*. (2004)] and several databases dedicated to the collection of structural domain–domain interactions are available [Finn *et al*. (2005); Stein *et al*. (2005); Davis and Sali (2005)]. Much work has concentrated on understanding under what circumstances homologous interactions are conserved [Pazos and Valencia (2001); Aloy *et al*. (2003); Tsai *et al*. (1996); Torrance *et al*. (2005)]. Aloy *et al*. (2003) did an extensive analysis on the relationship between sequence similarity and binding orientation and showed the geometry of interaction tends to be conserved between highly similar pairs.

An alternative approach is to investigate how non-homologous proteins bind at equivalent surfaces of homologous proteins [Tsai *et al*. (1996)]. Such interactions do not necessarily compete *in vivo*, but they reveal equivalent interaction sites. In some cases, the interactions may be truly competitive and regulated temporally by chemical modification or regulatory factors and spatially by compartmentalization. Independent of competitive or non-competitive binding, the identification of equivalent interfaces is a pointer to convergently evolved motifs. The motifs help to reveal key features which are necessary for the interaction.

A well-known example of a convergently evolved motif is the catalytic triad (Ser, His, Asp) found in both chymotrypsin and subtilisin (e.g. Fig. 4a). The local features of the enzymes' catalytic sites are conserved in other enzymes [Torrance *et al*. (2005)]. Chymotrypsin and subtilisin do not share any sequence or structure similarity. Indeed, both belong to different classes with chymotrypsin consisting only of beta-sheets and subtilisin of beta–alpha–beta units. Despite this different architecture, there are various inhibitors, which inhibit both enzymes and which use the same interface to do so. Thus, despite non-homology of the enzymes, equivalent binding sites are used.

Consider Figure 1a. To elucidate such interfaces with convergently evolved motifs, we screen the known structures in PDB for pairs of interactions $A - B$ and $A' - C$, where B, C are from different superfamilies and A, A' from the same family. If B and C bind to equivalent sites of A and A', respectively, we label B and C as

*To whom correspondence should be addressed.

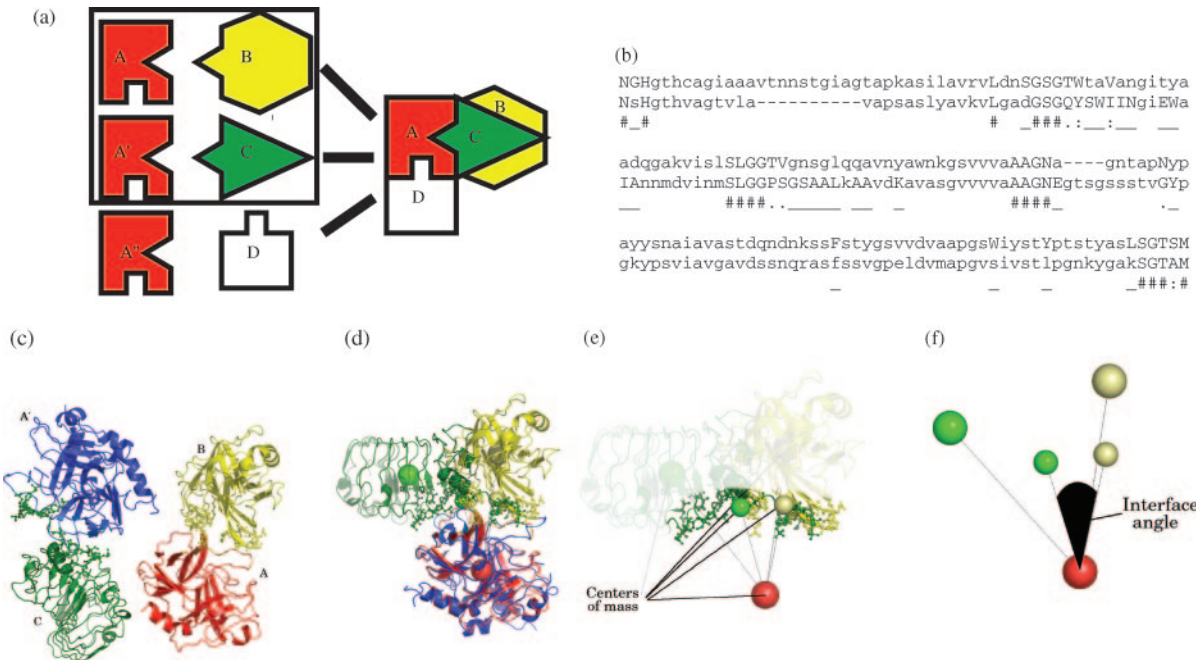


Fig. 1. (**a**) Overview: Family A, represented by A, A' and A'', interacts with B, C, D from different superfamilies. B and C bind to the same interface on A, whereas D does not. (**b**) A partial alignment for the sequences of two Subtilisin-like domains (SCOP id c.41.1.1) including a consensus line for interface residues. The respective binding partners are CI-2 family of serine protease inhibitors (d.40.1.1) and Protease propeptides/inhibitors (d.58.3.2). Capital letters indicate interacting residues. The consensus marks interface residues: identical (#), with positive (:) and negative (.) BLOSUM62 score. Positions with only one residue participating in the interface are marked _. (**c**) Structural comparison of interactions: two interacting pairs of domains ($A - B$ and $A' - C$) with sequential overlap in the interface of A. (**d**) Structural alignment of A and A' yields a (partial) spatial overlap of interfaces. (**e**) The centers of mass are calculated for $A/A', B$ and C (big spheres, in a similar shade as their domains), as well as for the interfaces (small spheres). (**f**) The centers of mass between the interfaces with regard to the shared partner's center of mass constitute an angle that reflects the proximity of the interfaces. We will refer to it as interface angle IA. The general binding partners' orientation—referred to as domain angle DA—is captured similarly.

interfaces with convergently evolved motifs. To define the equivalence of interfaces we use sequence and structure alignments of the shared domains A and A'. If there is sufficient overlap in the sequence alignment of A and A''s interface residues and if the angle between interfaces of B and C after superimposition of A and A' is sufficiently small, B and C bind at equivalent sites.

In our analysis we use Structural Classification of Proteins (SCOP) domains [Murzin *et al*. (1995)]. We require A and A' to be of the same family, since interfaces are known to be more conserved in both sequence and structure within a family [Valdar and Thornton (2001)], but not across the families of a superfamily [Rekha *et al*. (2005)]. For B and C we require different superfamilies, which ensures that they are evolutionarily not related.

The method sketched above identifies all pairs of interfaces with convergently evolved motifs. One application of such a resource is the study of viral proteins, which mimic the interfaces of native proteins and can thus interfere accordingly. We discuss two such cases: the M3 protein, which mimics the chemokine homodimer interface and the regulatory factor Nef found in HIV, which mimics a kinase interface when interacting with SH3.

METHODS

To identify non-homologous domains binding at equivalent sites we proceed as illustrated in Figure 1a: we consider all pairs of interactions $A - B$ and $A' - C$, where A and A' belong to the same family and B and C to different superfamilies. If B and C bind at equivalent interfaces, we screen them for shared motifs. To define equivalent binding sites, we use a two-stage procedure: first, we scan for a significant interface residue overlap on the aligned sequences; second, the angle and the spatial overlap between the two interfaces are used to further refine the search for overlapping interfaces.

Sequence comparison of interfaces. Sequential analysis is based on a non-redundant (NR) multiple sequence alignment for each SCOP family. Each sequence contains information about its interacting residues. For all pairs of interactions $A - B$ and $A' - C$ we check to what extent the interfaces of A and A' agree. We define sequential interface overlaps in terms of positional agreement of interacting residues. The sequential overlap can be measured as follows:

Interface sequence overlap ISO counts how many interacting residues are aligned in the multiple sequence alignment of the family containing A and A'.

Interface sequence overlap ISO+ refines ISO in that only ISO residues with a positive BLOSUM62 substitution are counted.

As an example consider Figure 1b, where 59% of the interacting residues of A and A' are aligned, i.e. ISO = 59%, and where 50% of the aligned interacting residues have a positive BLOSUM score, i.e. ISO+ = 50%.

Structural comparison of interfaces. Consider Figure 1c–f. For two interactions $A - B$ and $A' - C$, the domain A and A' are structurally aligned using MultiProt [Shatsky *et al*. (2004)]. This is feasible since A and A' belong to the same family. Next, the centers of mass of the C_α atoms for the domains A, A', B, C as well as for the interfaces of B and C are computed. To measure the overlap on A's and A''s interaction sites, we compute:

The interface angle IA between the center of mass $A(A')$ and the centers of the masses of the interfaces of B and C.

The domain angle DA between the domains is also measured by taking the centers of masses for A/A', B and C, as the angle DA gives an information on the spatial arrangement of the domains in general. Note that IA is rather sensitive, while DA can be large.

The interface atom overlap IAO: We calculate the percentage of atoms in B's interface within 3 Å of C's interface and vice versa.

The motif match score MMS involves a residue–residue correspondence analysis of domains B and C on atom level. We detect correspondences based on pairwise distances between interface atoms. They fall into 4 categories: C-α atom pairs, C-β atom pairs, remaining side chain atoms and backbone atoms. Matching residue pairs are discovered by the amount and category of atom pair matches. The score for each residue pair is simply its BLOSUM score (if positive) + the sum of atom pair scores. The latter are between 0 and 1, with 0 for 3 Å distance or above and 1 for exact coordinate matches, linearly interpolated. A detailed listing of all matches and their respective score is given on the accompanying web site.

As an indication of the structural alignment quality, we check the root mean square deviation (RMSD) between A and A' as well as the percentage of aligned residues. For instance, all the examples summarized in Figure 4 have an RMSD <1.3 Å and more than 90% of residues aligned.

```
1 Compute all domain–domain interaction pairs A-B
2 For all SCOP families Do:
3 Sequence Alignment of all family members
4 For all pairs A-B and A′-C Do
5 If MSA with A, A′ has ISO overlap > 30% Then
6 Structurally align A, A′
7 Compute centers of mass of interfaces
8 Compute interface angle IA
9 Compute interface atom overlap IAO
10 Compute motif match score MMS
11 Add ISO, IA, IAO, MMS to database
12 Sort database by MMS
```

Workflow. The workflow is summarized above as pseudo-code. In Step 1, we consider all domain–domain interactions in the PDB, using the SCOP domain definition. Domain pairs having at least 5 residue pairs within 5 Å are considered as interacting [Park *et al*. (2001); Dafas *et al*. (2004)].

In Step 2, the domain sequences are aligned using hidden markov models for each SCOP family in three steps. First, the seed sequences are aligned for each family after generating a series of NR sequences using Cluster Database at High Identity with Tolerance (CD-HIT). The cut-off for removing redundancy was varied from NR 98% to NR 70% to limit the number of seed sequences practical for multiple structural alignments. This limitation is needed particularly for large families such as immunoglobulin, which have more than a thousand member domains. The seed sequences are aligned by T-Coffee [Notredame *et al*. (2000)] based on the library of pairwise structural alignments. As T-Coffee makes the consensus alignments from the pool of pairwise alignment libraries, the resulting seed alignments are essentially multiple structural alignments. Second, hidden Markov models (HMM) are generated using the seed alignments. The influence of varied NR cutoff gets less critical because sequence weighting is applied in the course of building HMM models. Finally, all the member domain sequences in each family are aligned using the family-specific HMM model.

Using the multiple sequence alignment described above, the interface sequence overlap, ISO, is computed (Step 4). If the ISO is greater 30%, the structures of A and A' are aligned with MultiProt [Shatsky *et al*. (2004)] in Step 5. The interface angle, interface atom overlap and motif match score in Step 6–9 are computed by scripts using PyMOL functionality [Delano (2002), www.pymol.org]. All data characterizing a record of non-homologous binding is entered into a database (Step 10). Finally the database is sorted by motif match score (Step 11).

Family	Description	Binding to convergent interfaces	
b.1.1.2	Immunoglobulin	36.2%	(39494/108902)
b.1.1.1	Immunoglobulin	7.8%	(13236/168669)
b.47.1.2	Trypsin-like serine proteases	19.0%	(1523/8004)
c.37.1.8	P-loop	11.5%	(749/6496)
b.1.1.4	Immunoglobulin0	9.6%	(371/3840)
b.1.2.1	Fibronectin type III	7.0%	(363/5182)
d.19.1.1	MHC antigen-recognition d.	2.2%	(221/9886)
g.3.11.1	EGF/Laminin	17.1%	(166/966)
b.34.2.1	SH3-domain	20.1%	(110/546)
d.15.4.2	2Fe-2S ferredoxin-like	9.8%	(70/715)
d.2.1.2	Lysozyme-like	15.1%	(57/378)
b.42.2.1	Ricin B-like lectins	21.6%	(55/254)
d.144.1.7	Protein kinase-like (PK-like)	4.5%	(54/1184)

Fig. 2. The most frequent families A, which act as shared interaction partners. 'Binding to convergent interfaces' denotes the percentage of cases where B and C bind to A at the same site. In brackets is the total number of B/C binding at the same site divided by the total number of B/C binding to A.

RESULTS

We first present statistics, which show the most common families with convergently evolved motifs and the overall frequency of this phenomenon. Next, we discuss six examples in more detail.

Statistics. With over 40 000 domain interactions in the PDB, there is a combinatorial explosion of over 800 000 000 pairwise comparisons of domain interactions. Some 12 000 000 of these pairs are pairs with a common partner. After redundancy reduction, this number reduces dramatically to 70 000 pairs excluding immunoglobulins/proteases and 360 000 including them. Out of these pairs, 3000 (58 000 including immunoglobulins/proteases) have an interface sequence overlap of greater 30%. i.e. 4.2% (16% including immunoglobulins/proteases) of all non-homologous domains binding to a common partner do so at equivalent sites.

These interactions cover 270 common families, which account for ∼15% of the total 1834 families. Immunoglobulin related families (b.1.1.1, b.1.1.2, b.1.1.4) with 93% constitute the majority. Some families of regulatory function are frequently found as common partner such as the P-loop (c.37.1.8), SH3 (b.34.2.1) and Protein kinase-like domain (d.144.1.7). The most abundant families are shown in Figure 2.

More than 63% of interfaces with ISO greater 30% have an interface angle of <25°. Only 8% have an interface angle of >60°, suggesting the criterion of ISO >30% is sufficient to filter out most spurious cases where the domains B and C do not bind to equivalent surfaces on the common family (Fig. 3).

Exceptional cases are Phycocyanin-like phycobilisome proteins (SCOP a.1.1.3, PDB: 1qgw and 1on7, not shown). They are able to bind non-globular alpha+beta subunits of globular proteins (SCOP d.184.1.1) or to build homodimers on equivalent binding sites. Despite the large sequential overlap of interface residues (37%), the interface centers are placed in distant locations including an angle of 85°.

A large interface sequence overlap ISO generally indicates a small interface angle IA. The opposite, however, does not hold,

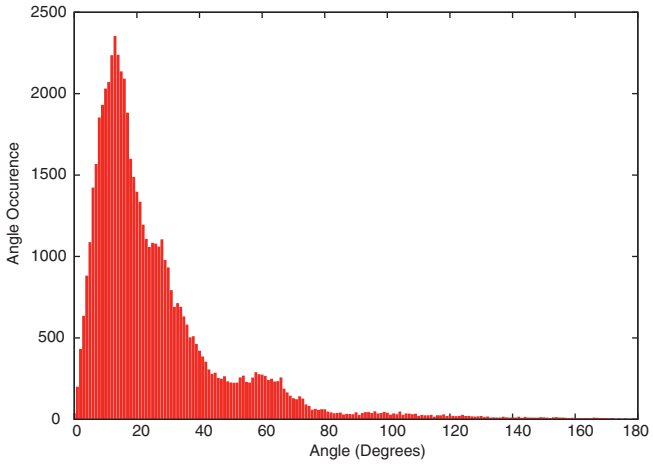


Fig. 3. Statistics of interface overlap parameters. The occurrence of all angles. Note that most angles of all detected interfaces with convergently evolved motifs are below 30° s.

as cases with small IA exist despite a small ISO only (see also Supplementary information).

The ratio of the overlapping interface residues with positive BLOSUM score (ISO+) is upper-bounded by ISO and widely distributed nearly over the whole range (0–100%), suggesting the various degrees of divergence for different partners. The families of the highest overlap (ISO) and conservation (ISO+) include trypsin (b.47.1.2) and its inhibitors from various superfamilies (Fig. 4a). The interface angle generally tends to decrease as the interface sequence overlap increases. Above 60% of ISO, the two partners associate within 30° of angle difference in most cases. ISO+ and interface angle show similar relationship as ISO and interface angle, but less correlation. It may be because the angle is only related to the geometry or the position on the alignments but not to the residue type.

Next, we discuss four examples of convergently evolved motifs. We start with a well-known example of subtilisin and trypsin-like serine proteases as a validation of our approach.

Shared partners of subtilisin and trypsin like serine proteases. Subtilisin (c.41.1.1) and trypsin-like serine proteases (chymotrypsin, b.47.1.2) have no obvious similarity wrt. sequence and structure. However, as it is known from Carter and Wells (1988), their catalytic triads comprise the same residues. While this common motif is impossible to detect by sequence or structure alignments, it is striking that there are as many as three inhibitors, which interact with both families: Plant proteinase inhibitors (g.15.1.2), Serine protease inhibitors (d.40.1.1), and Subtilisin inhibitors (d.84.1.1).

Using our method with the inhibitors acting as common interaction partner, we can superimpose the otherwise unalignable structures of subtilisin and chymotrypsin and the catalytic triads are indeed localized in immediate vicinity. Note that no a priori knowledge, such as structural templates for catalytic sites [as in Torrance *et al*. (2005)] is employed.

Local structure conservation in chymotrypsin's binding partners. The screen shows that a large number of chymotrypsin's inhibitors belong to such diverse superfamilies as ecotin (b.16.1), STI-like (b.42.1) and ovomucoid PCI1-like inhibitor (g.15.1). A common feature is the binding to a pocket adjacent to trypsin's catalytic triad, Figure 4a. This keyhole binds to side chains from loop regions with high but local structural similarity. The motif derived from the structurally aligned residues can serve as a template to search for chymotrypsin binding sites.

M3 mimicry of chemokine binding. Chemokines play a key role in leukocyte recruitment and migration. Alexander *et al*. (2002) report that viral protein M3 sequestrates chemokine with high affinity due to conformational flexibility and electrostatic complementation. Figure 4b shows an additional feature of M3: an optimal fitting to a binding site that is utilized by chemokines to form homodimers. To achieve this, the virus has evolved the SVSPLP motif which can play the role of the native motif SSDTTP.

Nef mimicry of SH3 binding. Regulatory factor Nef (d.102.1.1, PDB: 1efn D) and Protein-Kinase like (PKL) (d.144.1.7, 2chk B) exhibit similarities in their way of binding SH3. As shown in Figure 4c, several residues are in relative proximity in Nef/PKL: Arg71/Lys241, Pro72/Pro250, Gln73/Gln251, Pro75/Pro253 (PxxP motif), Phe90/His289, which are part of a hydrophobic pocket. Note that all residue pairs are of similar/equal physico–chemical properties.

Baculovirus p35 protein resembles apoptosis inhibitor motif. The viral p35 protein is known to be an effective broad-spectrum inhibitor for caspase thus preventing apoptosis. [Xu *et al*. (2003)]. The caspase recognition sequence and adjacent residues are found to be similar to residues of inhibitor of apoptosis (IAP) repeat at corresponding positions. Key residues in this resembly are revealed by the similarity screen (step 8): P35's Asp84 matches IAP's Asp148 in both position and orientation (C-α, C-β, backbone and side chain atoms all correspond). Together with their respective sequence neighbours (backbone matches) they shape a convergently evolved motif.

Moreover, caspase offers a second binding site, which both p35 and IAP make use of (Ser252–Asn234).

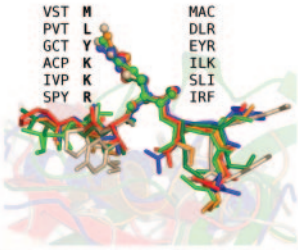
HIV-capsid protein and Cyclophilin interfere. HIV capsid protein (HIV-CA) blocks the cyclophilins homodimerisation binding site by forming a backbone stretch (Gly94–Ile91) that resembles a loop region of cyclophilin (Arg143–Met146). Together with Val86–Thr88, HIV-CA corresponds to residues in cyclophilin peripherical to the active site.

Summary of examples. A summary of interfaces with significant commonalities is presented in Figure 4. The complete set of NR examples (wrt. SCOP family combinations) is presented in great detail on the Supplementary information web site.

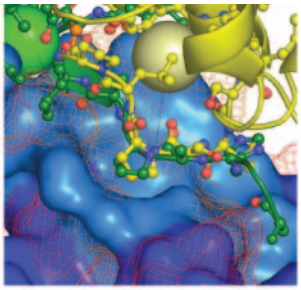
CONCLUSION

In this work we extract instances belonging to the same family that bind to completely different binding partners through equivalent interfaces. As pointed out by Tsai *et al*. (1996), these cases may provide particular insight into biomolecular recognition: the study of most diverse binding partners and the extraction of their commonalities suggest key principles for domain–domain binding.

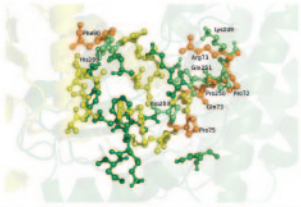
To this end we designed a set of both sequential and structural criteria to allow for an exhaustive screen. Our method shows that out of all interaction pairs with a common partner (excluding immunoglobulins and proteases) and two non-homologous domains



(a) Many non-homologous chymotrypsin inhibitors bind at the S_1 pocket featuring a residue with a long side chain, typically preceded by a threonine or proline.



(b) Viral protein M3 (yellow) adapting to chemokine (blue), mimicking chemokines homodimer (green). The backbone of M3 between Ser269 to Pro274 (motif SVSPLP) resembles Ser4 to Pro9 (motif SSDTTP) in chemokine.



(c) The viral negative factor Nef and a protein kinase-like aligned via their shared partner SH3. Corresponding residues are labelled. E.g. Pro72/Pro250 and Pro75/Pro253 match.

A	B	C	*ISO* (%)	Angle *IA*	Overlap *IAO*	Motif score	RMSD (aligned %)	Motif
Serine protease inhibitor (d.40.1.1) Leech/Barley	Chymotrypsin (b.47.1.2) Cow	Subtilisin (c.41.1.1) amyloliquef.	75%	10.8°	96%	79	0.8 (99%)	Catalytic triad H,D,S well aligned Gly193-Gly219, Phe41-Asn218 411 atom correspondences
Chymotrypsin (b.47.1.2) Cow	Inhibitors (e.g. b.16.1.1) E.coli	Inhibitors (e.g. g.4.1.1) Locust	77%	21.4°	97%	110	1.0 (92%)	long side chain (often K,R), ± 3 residues, high backbone similarity
SH3 (b.34.2.1) Human	Nef (d.102.1.1) HIV-1	P-kinase like (d.144.1.7) Human	74%	15.5°	92%	56	0.8 (98%)	PxxP motif in spacial proximity, Arg71-Lys241, Phe90-His289
Caspase (c.17.1.1) Human	Baculovirus p35 protein (b.28.1.1) AcMNPV	Inhibitor of apoptosis repeat (g.52.1.1) Human	56%	19.0°	80%	37	1.0 (91%)	SDQM (83-86) vs IDVV (149-146) consecutive backbone match with positive blosum score, Ser253–Asn234.
Cyclophilin (c.17.1.1) Human/E.coli	Retrovirus capsid protein (b.28.1.1) HIV-1	Cyclophilin (g.52.1.1) E.coli	52%	23.0°	52%	22	1.0 (91%)	consecutive backbone match: viral Ile91–Gly94 matches Met146–Arg143
Chemokine (d.9.1.1) Human	M3 viral protein (b.116.1.1) Muhv-4	Chemokine (d.9.1.1) Human	67%	37.1°	70%	64	1.3 (91%)	Residues 270–274 in M3 (VSPLP) match chemokine's residues 4–9 (SDTTP)

Fig. 4. Examples of non-homologous domains binding at equivalent sites together with a description of their convergently evolved motifs.

4.2% bind at equivalent sites and feature convergently evolved motifs.

The degree of ligands' structural conservation varies strongly. Sometimes, pairwise interface residue correspondences are possible, either when clefts or enzyme pockets allow for very little degree of freedom (chymotrypsin's S_1 pocket adjacent to its catalytic triad), or in the case of flat interfaces as illustrated in the example of SH3 interacting with Nef and PKL.

Generally our results suggests that the majority of alternative ligands (B and C) show some local sequence and shape similarity. For the remaining cases it will be interesting to explore how patches with neither sequential nor structural similarity can mimic each others surface conditions. We showed in the example of chymotrypsin's binding partners, how our method generates sequential patterns from local structural alignments. An adequate statistic representation (such as position-specific scoring matrices) will help to predict potential binding sites.

As the examples—particularly the subtilisin-chymotrypsin comparison—demonstrate, information about functional sites can be inferred using our method. The incorporation of functional site information {e.g. from the catalytic site atlas [Porter *et al*. (2004)]} will help to localize functionally important residues. Finally, viral proteins' mimicry of native interfaces are detected (see section on chemokines, SH3, caspase and cyclophilin). More examples are online in the Supplementary database: scoppi.biotec.tu-dresden.de/abac/

All in all, our novel method allows for a first comprehensive overview on how non-homologous domains can evolve similar motifs, which allow them to bind to the same partner at equivalent sites.

ACKNOWLEDGEMENTS

We gratefully acknowledge support of the EFRE Project CODI. We would like to thank Nataraj Dongre, Lee Cheung and Christof Winter. Last but not least, we are thankful to the fruitful comments by the anonymous reviewers, which helped to improve the paper. Funding to pay the Open Access publication charges for this article was provided by EFRE project CODI no. 4212/04-07.

Conflict of Interest: none declared.

REFERENCES

Alexander,J. *et al*. (2002) Structural basis of chemokine sequestration by a herpesvirus decoy receptor. *Cell*, **111** (3), 343–356.

Aloy,P. *et al.* (2003) The relationship between sequence and interaction divergence in proteins. *J. Mol. Biol.*, **332** (5), 989–998.

Apic,G. *et al.* (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.*, **310**, 311–325.

Bashton,M. and Chothia,C. (2002) The geometry of domain combination in proteins. *J. Mol. Biol.*, **315**, 927–939.

Berman,H. *et al.* (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.

Bolser,D. *et al.* (2003) Visualisation and graph-theoretic analysis of a large-scale protein structural interactome. *BMC Bioinformatics*, **4**, 45.

Carter,P. and Wells,J. (1988) Dissecting the catalytic triad of a serine protease. *Nature*, **332**, 564–568.

Chakrabarti,P. and Janin,J. (2002) Dissecting protein–protein recognition sites. *Proteins*, **47**, 334–343.

Conte,L.L. *et al.* (1999) The atomic structure of protein–protein recognition sites. *J. Mol. Biol.*, **285**, 2177–2198.

Dafas,P. *et al.* (2004) Using convex hulls to extract interaction interfaces from known structures. *Bioinformatics*, **20**, 1486–1490.

Davis,F. and Sali,A. (2005) Pibase: a comprehensive database of structurally defined protein interfaces. *Bioinformatics*, **21**, 1901–1907.

Delano,W. (2002) The PyMOL molecular graphics system. www.pymol.org.

Finn,R. *et al.* (2005) ipfam: visualization of protein–protein interactions in pdb at domain and amino acid resolutions. *Bioinformatics*, **21**, 410–412.

Jones,S. and Thornton,J. (1996) Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13–20.

Kim,W. *et al.* (2004) Large-scale co-evolution analysis of protein structural interlogues using the global protein structural interactome map (PSIMAP). *Bioinformatics*, **20**, 1138–1150.

Murzin,A. *et al.* (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536.

Notredame,C. *et al.* (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–217.

Nussinov,R. *et al.* (1997) Hydrogen bonds and salt bridges across protein–protein interfaces. *Protein Eng.*, **10**, 999–1012.

Ofran,Y. and Rost,B. (2003) Analyzing six types of protein–protein interfaces. *J. Mol. Biol.*, **325**, 377–387.

Park,J. *et al.* (2001) Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the pdb and yeast. *J. Mol. Biol.*, **307**, 929–938.

Pazos,F. and Valencia,A. (2001) Similarity of phylogenetic trees as indicator of protein–protein interaction. *Protein Eng.*, **14**, 609–614.

Porter,C. *et al.* (2004) The catalytic site atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res.*, **32**, D129–D133.

Rekha,N. *et al.* (2005) Interaction interfaces of protein domains are not topologically equivalent across families within superfamilies: implications for metabolic and signaling pathways. *Proteins*, **58**, 339–353.

Shatsky,M. *et al.* (2004) A method for simultaneous alignment of multiple protein structures. *Proteins*, **56**, 143–156.

Stein,A. *et al.* (2005) 3did: interacting protein domains of known three-dimensional structure. *Nucleic Acids Res.*, **33**, D413–D417.

Torrance,J. *et al.* (2005) Using a library of structural templates to recognise catalytic sites and explore their evolution in homologous families. *J. Mol. Biol.*, **347**, 565–581.

Tsai,C. *et al.* (1996) A dataset of protein–protein interfaces generated with a sequence-order-independent comparison technique. *J. Mol. Biol.*, **260**, 604–620.

Valdar,W. and Thornton,J. (2001) Protein–protein interfaces: analysis of amino acid conservation in homodimers. *Proteins*, **42**, 108–124.

Xu,G. *et al.* (2003) Mutational analyses of the p35-caspase interaction. A bowstring kinetic model of caspase inhibition by p35. *J. Biol. Chem.*, **278**, 5455–5461.