Structural Modeling of Protein Interactions by Analogy: Application to PSD-95

Dmitry Korkin[1,2,3], Fred P. Davis[1,2,3], Frank Alber[1,2,3], Tinh Luong[4], Min-Yi Shen[1,2,3], Vladan Lucic[5], Mary B. Kennedy[4], Andrej Sali[1,2,3]*

1 Department of Biopharmaceutical Sciences, University of California San Francisco, San Francisco, California, United States of America, 2 Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, United States of America, 3 California Institute for Quantitative Biomedical Research, University of California San Francisco, San Francisco, California, United States of America, 4 Division of Biology, California Institute of Technology, Pasadena, California, United States of America, 5 Department of Structural Biology, Max Planck Institute of Biochemistry, Martinsried, Germany

We describe comparative patch analysis for modeling the structures of multidomain proteins and protein complexes, and apply it to the PSD-95 protein. Comparative patch analysis is a hybrid of comparative modeling based on a template complex and protein docking, with a greater applicability than comparative modeling and a higher accuracy than docking. It relies on structurally defined interactions of each of the complex components, or their homologs, with any other protein, irrespective of its fold. For each component, its known binding modes with other proteins of any fold are collected and expanded by the known binding modes of its homologs. These modes are then used to restrain conventional molecular docking, resulting in a set of binary domain complexes that are subsequently ranked by geometric complementarity and a statistical potential. The method is evaluated by predicting 20 binary complexes of known structure. It is able to correctly identify the binding mode in 70% of the benchmark complexes compared with 30% for protein docking. We applied comparative patch analysis to model the complex of the third PSD-95, DLG, and ZO-1 (PDZ) domain and the SH3-GK domains in the PSD-95 protein, whose structure is unknown. In the first predicted configuration of the domains, PDZ interacts with SH3, leaving both the GMP-binding site of guanylate kinase (GK) and the C-terminus binding cleft of PDZ accessible, while in the second configuration PDZ interacts with GK, burying both binding sites. We suggest that the two alternate configurations correspond to the different functional forms of PSD-95 and provide a possible structural description for the experimentally observed cooperative folding transitions in PSD-95 and its homologs. More generally, we expect that comparative patch analysis will provide useful spatial restraints for the structural characterization of an increasing number of binary and higher-order protein complexes.

Introduction

Protein–protein interactions play a key role in many cellular processes [1,2]. An important step towards a mechanistic description of these processes is a structural characterization of the proteins and their complexes [3–6]. Currently, there are two computational approaches to predict the structure of a protein complex given the structures of its components, comparative modeling [6–11] and protein–protein docking [12–15].

In the first approach to modelling a target complex, standard comparative modelling or threading methods build a model using the known structure of a homologous complex as a template [7,10]. The applicability of this approach is limited by the currently sparse structural coverage of binary interactions [6]. In the second approach, an atomic model is predicted by protein–protein docking, starting from the structures of the individual subunits without any consideration of homologous interactions [12–16]. This docking is usually achieved by maximizing the shape and physicochemical complementarity of two protein structures, through generating and scoring a large set of possible configurations [13,16]. Experimental information, such as that obtained from NMR chemical shift mapping, residual dipolar couplings, and cross-linking, can also be used to guide protein docking [17–20]. While docking is applicable to any two subunits whose structures are known or modeled, both the sampling of relevant configurations and the discrimination of native-like configurations from the large number of non-native alternatives remain challenging [15].

Comparative Patch Analysis

Here, we propose a third approach to modeling complexes between two structures (Figure 1). The approach, called comparative patch analysis, is a hybrid of protein docking and comparative modeling based on a template complex, with a greater applicability than comparative modeling and a higher accuracy than docking. Comparative patch analysis relies on our prior analysis of the location of binding sites within families of homologous domains [21]. This analysis

Abbreviations: DOPE, Discrete Optimized Protein Energy; GBS, GMP-binding site; GK, guanylate kinase–like; MALDI–TOF, matrix-assisted laser desorption ionization–time of flight); PDZ, PSD-95, DLG, and ZO-1; PDB, Protein Data Bank; PRBS, proline-rich binding site; PSD, postsynaptic density; SAXS, small angle X-ray scattering; SCOP, Structural Classification of Proteins; SH3, Src homology 3

* To whom correspondence should be addressed. E-mail: sali@salilab.org

Synopsis

Protein–protein interactions play a crucial role in many cellular processes. An important step towards a mechanistic description of these processes is a structural characterization of the proteins and their complexes. The authors developed a new approach to modeling the structure of protein complexes and multidomain proteins. The approach, called comparative patch analysis, complements the two currently existing approaches for structural modeling of protein complexes, comparative modeling, and protein docking. It limits the configurations refined by molecular docking to the structurally defined interactions of each of the complex components, or their homologs, with any other protein, irrespective of its fold; the final prediction corresponds to the best-scoring refined configuration. The authors applied comparative patch analysis to predict the structure of the core fragment of PSD-95, a five-domain protein that plays a major role in the postsynaptic density at neuronal synapses. The study suggests two alternate configurations of the core fragment that potentially correspond to the different functional forms of PSD-95. This finding provides a possible structural explanation for the experimentally observed cooperative folding transitions in PSD-95 and its homologs.

indicated that the locations of the binding sites are often conserved irrespective of the folds of their binding partners. The structure of the target complex can thus be modeled by restricting protein docking to only those binding sites that are employed by homologous domains. As a result, comparative patch analysis benefits from knowledge of all interactions involving either one of the two partners.

We find that comparative patch analysis increases the prediction accuracy relative to protein docking. It is able to correctly identify the binding mode in 70% of 20 benchmark complexes, predicting the overall structure with an average improvement in all-atom RMS error of 13.4 Å, compared with protein docking. In contrast, protein docking correctly identifies the binding mode in 30% of the complexes.

PSD-95 Protein

We apply comparative patch analysis to model the structure of PSD-95 protein. PSD-95 is abundant in the postsynaptic density (PSD), a cytosolic organelle that plays a pivotal role in neuronal signaling [22–26]. PSD-95 serves as a major scaffold for other signaling proteins, participates in receptor and channel clustering, and performs a range of other diverse functions [25–31].

PSD-95 is a member of the membrane associated guanylate kinase (MAGUK) family. It is composed of three PDZ (named after PSD-95, DLG, and ZO-1) domains followed by SH3 (Src homology 3) and GK (guanylate kinase-like) domains [32,33]. Isolated structures of all three PDZ domains as well as the structure of the SH3-GK domain complex have been solved [34–38]. The complete structure of PSD-95 has not been determined, but experiments suggest that it adopts multiple conformations [39,40]. The structures of these conformations are necessary for functional insight into the regulation of PSD-95 activity [40,41].

We apply comparative patch analysis to model the structure of the complex between the third PDZ (PDZ$_3$), SH3, and GK domains. These domains comprise 60% of the PSD-95 mass and are the defining domains of the membrane-associated guanylate kinase family. We propose two configurations that satisfied all imposed spatial restraints, including
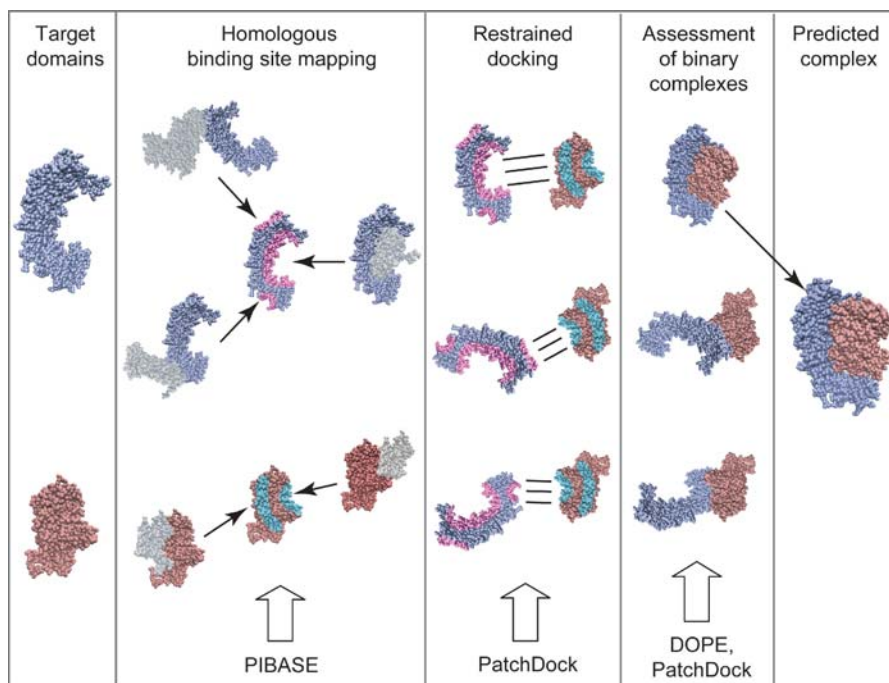
Figure 1. Basic Steps of Comparative Patch Analysis Approach
First, the binding sites of the homologs of each domain are extracted from PIBASE and superposed on its surface. Second, for each pair of the superposed binding sites, we apply a restrained docking of the domains with PatchDock to obtain a set of candidate binary domain complexes. Each of the binary complexes is then ranked using geometrical complementarity and statistical potential, and the top-ranked complex is selected to be a final prediction.

previously observed binding sites, consistency with the given linker length, and physicochemical complementarity of the interacting surfaces. In addition, the prediction is in concordance with and rationalizes available biochemical, structural, and evolutionary data.

The paper begins by comparing the performance of comparative patch analysis with protein docking on a benchmark set of 20 binary protein complexes (Results). Next the application of comparative patch analysis to predicting the structure of PDZ$_3$-SH3-GK complex is described (Results). We combine the predictions with existing experimental evidence to propose a mechanism for the intramolecular regulation of PSD-95 (Discussion). In addition, we discuss the advantages and disadvantages of comparative patch analysis and briefly outline future directions. Finally, we present the details of the method (Methods).

Results

To assess the method, we applied comparative patch analysis to a benchmark set of 20 binary complexes of known structure (Methods). We then used comparative patch analysis to predict the tertiary structure of the PSD-95 core fragment that contains PDZ$_3$, SH3, and GK domains.

Assessment of Comparative Patch Analysis

Comparative patch analysis may be applied to two scenarios where binding site information is available for both or just one of the interacting subunits. We compared their perform-

ance to that of protein docking (Methods). In both scenarios, comparative patch analysis was significantly more accurate than protein docking (Figure 2). Using both (one) binding site information, the overall structure was improved for 13 (8) of the 20 complexes, with an average improvement in the all-atom RMS error of 13.4 Å (6.1 Å). The interface coverage increased by 29% (6%), and the binding site coverage by 30% (10%), on average (Table 1). In 15 (8) complexes, comparative patch analysis produced models with all-atom RMS error <3 Å, while protein docking achieved this accuracy for only six complexes. Comparative patch analysis identified the interfaces correctly in 15 (9) complexes, including 8 (7) multi-domain proteins and 7 (2) protein complexes, while protein docking achieved this for 7 complexes, including 6 multi-domain proteins and 1 protein complex. In those 15 complexes, on average 71% of the predicted residue contacts were observed in the native structures (standard error is 5%). As expected, comparative patch analysis was more accurate using binding site information for both interacting domains compared with using only one.

Application to PSD-95

Next we modeled the tertiary structure of the core fragment of rat PSD-95, which includes the PDZ$_3$, SH3, and GK domains (Figure 3, see Figure 3A). As this fragment contains three independent domains, there are three possible domain–domain interactions. The interaction between SH3 and GK domains were known from X-ray crystallography [37,38]. Here
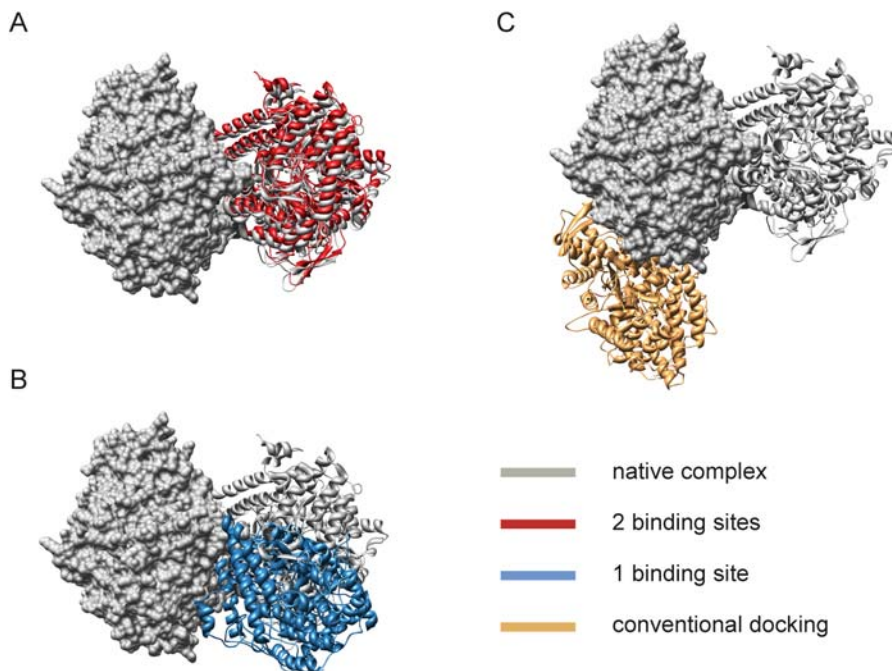

Figure 2. Examples of Predicted Protein Interface between Two Subunits for a Pyruvate Formate–Lyase Protein Complex from Our Benchmark Set

Shown are the structures of the native complex (grey) together with the best-scoring models that were predicted by comparative patch analysis using binding site information for (A) both, or (B) just one of the interacting subunits, and (C) by conventional protein docking, where no binding site information is provided. The predicted and native structures are superposed using one of the two subunits, which is represented by its accessible surface. The remaining subunits of the predicted structures are shown in the ribbon representation colored red, blue, and orange, correspondingly. In both scenarios, comparative patch analysis was significantly more accurate than protein docking. Using both binding sites, comparative patch analysis accurately predicted the protein interaction interface, including the relative orientation of subunits. The accuracy of interface prediction by our approach using only one binding site was significantly reduced, while it was still able to predict the binding sites near their native locations. The conventional protein docking failed to accurately predict either the relative orientation of subunits or the locations of their binding sites.
doi:10.1371/journal.pcbi.0020153.g002

Table 1. Assessment of Comparative Patch Analysis Approach

Number	PDB	Domain Chains and Boundaries		Restrained Docking (Two Binding Sites)			Restrained Docking (One Binding Site)			Unrestrained Docking		
		Domain1	Domain2	f_{BS}	f_{In}	RMSD	f_{BS}	f_{In}	RMSD	f_{BS}	f_{In}	RMSD
1	1l5k	A(4:349)	B(4:349)	0.96	0.88	0.64	0.23	0.00	20.45	0.05	0.00	34.35
2	1iew	A(1:388)	A(389:602)	0.92	0.85	0.64	0.92	0.83	0.71	0.79	0.48	4.20
3	1nfz	A(4:179)	B(4:183)	0.85	0.81	1.02	0.86	0.87	1.34	0.00	0.00	44.77
4	1orb	(1:149)	(150:293)	0.86	0.78	1.09	0.85	0.78	0.60	0.85	0.78	0.60
5	1cm5	A(1:759)	B(1:759)	0.86	0.73	1.10	0.37	0.00	30.41	0.00	0.00	63.34
6	1bwr	A(5:216)	B(5:216)	0.85	0.72	1.32	0.50	0.07	19.97	0.50	0.07	19.97
7	3pax	(662:796)	(797:1011)	0.77	0.66	2.07	0.77	0.66	2.07	0.77	0.66	2.07
8	1qck	A(1:89)	B(1:89)	0.91	0.68	2.47	0.10	0.00	47.82	0.02	0.00	46.01
9	1j93	A(10:352)	B(10:352)	0.43	0.02	33.02	0.25	0.00	33.02	0.08	0.00	33.02
10	1nsm	A(2:340)	B(2:347)	0.00	0.00	43.43	0.20	0.00	32.93	0.00	0.00	43.43
11	1g9a	A(862:1079)	A(1080:1289)	0.96	0.86	0.47	0.96	0.87	1.23	0.96	0.87	1.23
12	1fn7	A(9:135)	A(136:325)	0.95	0.78	0.85	0.84	0.78	0.81	0.84	0.78	0.81
13	1egu	A(171:540)	A(541:814)	0.86	0.74	0.98	0.78	0.61	3.12	0.78	0.61	3.12
14	1c97	A(2:528)	A(529:754)	0.88	0.73	1.11	0.93	0.82	0.91	0.93	0.82	0.91
15	2sic	E(1:275)	I(7:113)	0.83	0.80	1.33	0.27	0.00	26.86	0.30	0.00	13.19
16	1ltx	A(24:443)	B(6:331)	0.83	0.75	1.47	0.26	0.01	17.36	0.81	0.66	1.66
17	1pys	B(1:190)	B(191:399)	0.86	0.67	1.59	0.81	0.60	2.13	0.17	0.01	40.74
18	1e7v	A(525:725)	A(726:1094)	0.50	0.07	18.57	0.14	0.00	27.16	0.14	0.00	33.57
19	1tco	A(21:372)	C(1:107)	0.03	0.00	23.71	0.28	0.02	16.27	0.28	0.02	16.27
20	1ikn	A(19:191)	D(73:293)	0.17	0.01	39.73	0.04	0.00	36.08	0.00	0.00	40.76

The sample set of 20 binary protein complexes was used to evaluate our method. These complexes come from two groups. Each subunit of a complex in the first group is a member of a SCOP family that has been observed to interact with only one other SCOP family. In turn, each subunit from the second group of complexes comes from a SCOP family that has been observed to interact with multiple SCOP families. As expected, the accuracy of comparative patch analysis using two binding sites was higher for the first group of complexes ($\overline{\Delta RMSD} = 13.4$ Å, $\overline{\Delta f_{IN}} = 29\%$, $\overline{\Delta f_{BS}} = 30\%$) than for the second one ($\overline{\Delta RMSD} = 6.1$ Å, $\overline{\Delta f_{IN}} = 6\%$, $\overline{\Delta f_{BS}} = 10\%$).
doi:10.1371/journal.pcbi.0020153.t001

we focused on characterizing the other two putative interactions, namely between the PDZ$_3$ and SH3 as well as between PDZ$_3$ and GK domains. For both cases, we applied comparative patch analysis using two subunits, one containing PDZ$_3$ and the second one containing the interacting SH3 and GK domains. The first interaction was modeled using the binding site locations of the PDZ$_3$ and SH3 in all known homologs, while the second was modeled using those of the PDZ$_3$ and GK homologs (Methods). The results for both interactions are described next, followed by the comparison with results obtained by conventional protein docking.

PDZ$_3$–SH3 Interaction

The comparative patch analysis protocol was applied using the nonredundant sets of 49 PDZ$_3$ and 26 SH3 binding sites combined to give all 1,274 possible input pairs. The protocol resulted in an ensemble of 503 models of the PDZ$_3$–SH3 complex (Methods).

The interface of the best-scoring model (1493 Å^2) that satisfied the interdomain linker restraint consisted primarily of the C- and N-terminal residues of PDZ$_3$ as well as the residues of the proline-rich binding site (PRBS) and the first two beta strands of SH3 (Figure 3B). The PDZ$_3$ hydrophobic cleft, known to be essential for binding the C-termini of other proteins, remained accessible in this complex [42,43]. The N-terminus of PDZ$_3$ contains a PREP motif (P308, R309, E310, P311) which belongs to the canonical PXXP family of motifs known to interact with the PRBS of SH3 [44–47]. In the best-scoring model, this motif was in proximity to the PRBS (Figure 3B). Our confidence in this predicted binding mode was bolstered when its binding residues were found to occur in regions of high localization derived from the ten best

scoring models that satisfied the linker restraint (Figure 4). Ninety-four percent of the binding residues in the best-scoring model were found to occur in no less than 70% of the ten best-scoring models.

PDZ$_3$–GK Interaction

The comparative patch analysis protocol was applied to the PDZ$_3$–GK complex using 10,731 input pairs formed by combining the nonredundant sets of 49 PDZ$_3$ and 219 GK binding sites. The protocol resulted in an ensemble of 1,929 models (Methods).

The interface of the best-scoring model was extensive (2729 Å^2), and includes, among others, residues located at the C-terminus and near the hydrophobic cleft of PDZ$_3$ as well as a large groove of GK formed by the GMP-binding and LID regions [48–50] (Figure 3C). The analysis of the ten best-scoring models satisfying the interdomain linker restraints revealed high localization of the binding residues for both domains (Figure 4). The residues of PDZ$_3$ with the highest localization were located around the domain's hydrophobic cleft and the C-terminus (Figure 4). In addition, the entire GMP-binding site (GBS) of the GK domain and part of the hydrophobic cleft of the PDZ domain became inaccessible in most top-scoring models, including the best-scoring one. Forty-six percent of the binding residues in the best-scoring model were found to occur in no less than 70% of the ten best-scoring models.

Comparison with Protein Docking Results

To evaluate the effect of binding-site information on modeling the PDZ$_3$–SH3–GK complex, conventional protein docking of the PDZ$_3$ and SH3–GK domains was performed
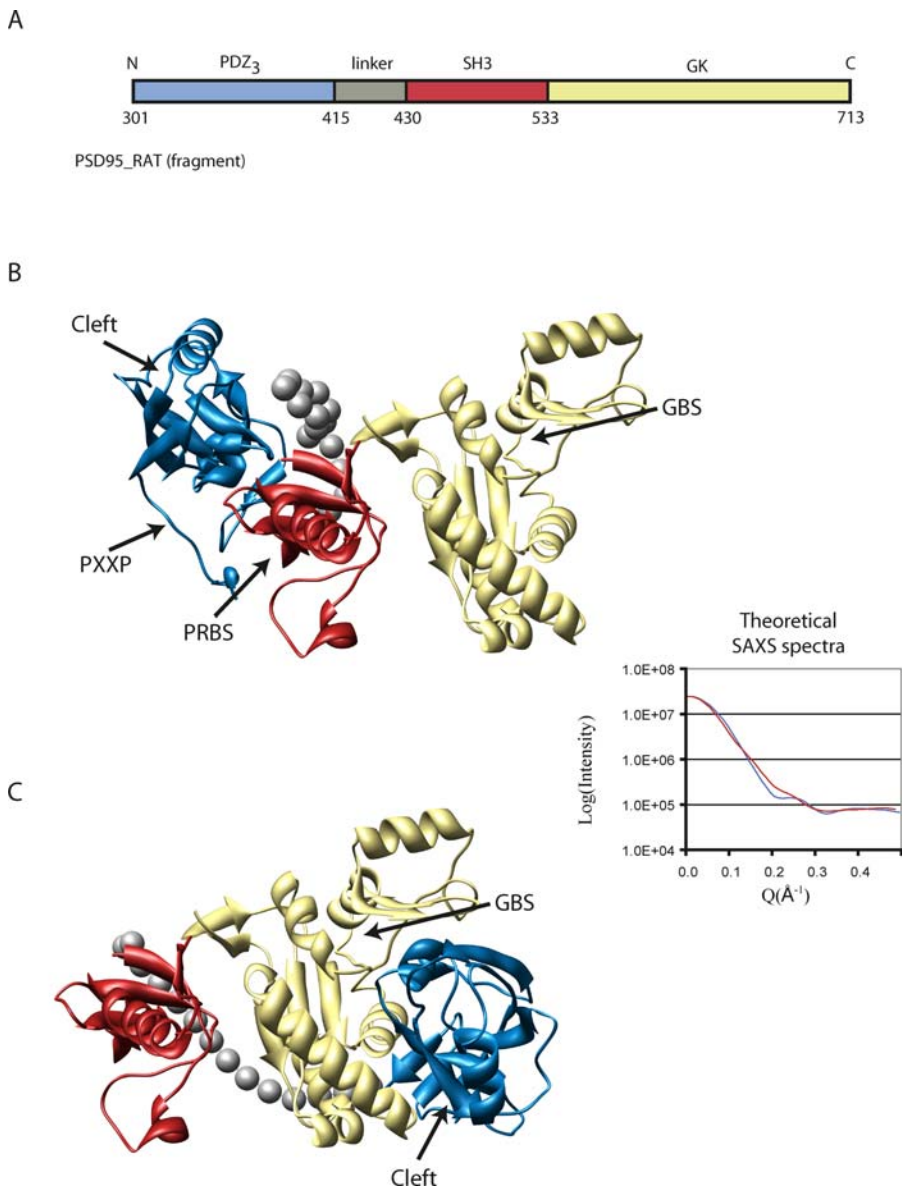
Figure 3. Two Binding Modes of the Core Fragment of Rat PSD-95

The PDZ$_3$ domain is shown in blue, SH3 in red, and GK in yellow. The grey spheres correspond to the residues of the interdomain linker between PDZ$_3$ and SH3. Locations of the hydrophobic cleft (Cleft) and PXXP motif (PXXP) in PDZ$_3$, PRBS in SH3, and the GBS in GK are shown by arrows. (A) The domain architecture of the core fragment. (B) The first predicted configuration. (C) The second predicted configuration. The difference between the theoretically calculated SAXS spectra of the first (red) and second (blue) configurations is significantly larger than the anticipated experimental error.
doi:10.1371/journal.pcbi.0020153.g003

(Methods). Analysis of the ten best-scoring models satisfying the interdomain linker restraint revealed that the binding sites of both PDZ$_3$ and SH3–GK domains were significantly delocalized compared with the comparative patch analysis models. Moreover, the binding residues of the top-scoring models almost completely covered the domain surfaces (93% and 81% of the PDZ$_3$ and SH3–GK domains) (Figure 4). The best-scoring model obtained using protein docking was different from both the best PDZ$_3$–SH3 and the PDZ$_3$–GK comparative patch analysis models (unpublished data).

PXXP Motif Conservation Analysis

The proximity of the PDZ$_3$ PXXP motif and the SH3 PRBS in the predicted model prompted a search for PXXP motifs in the sequences of six PSD-95 proteins and splice variants from four species to assess the significance of this observation. All sequences contained at least one form of a PXXP motif or noncanonical SH3-binding motif that could mimic the PXXP motif (Table 2) [47]. The human, rat, and mouse proteins all contained a PREP motif in PDZ$_3$; the zebrafish protein did not. Five other potential SH3 binding motifs were found outside of known domains; two at the N-terminus, one at the C-terminus, and two in the interdomain linker between PDZ$_2$ and PDZ$_3$. The conservation of the PREP sequence in PDZ$_3$ from the mammalian species suggests that its interaction with SH3 may be functionally significant.

Proteolysis of PSD-95

Limited proteolysis of recombinant PSD-95 with Proteinase K produces a prominent ~48 kDa band at 30 min (Figure 5).
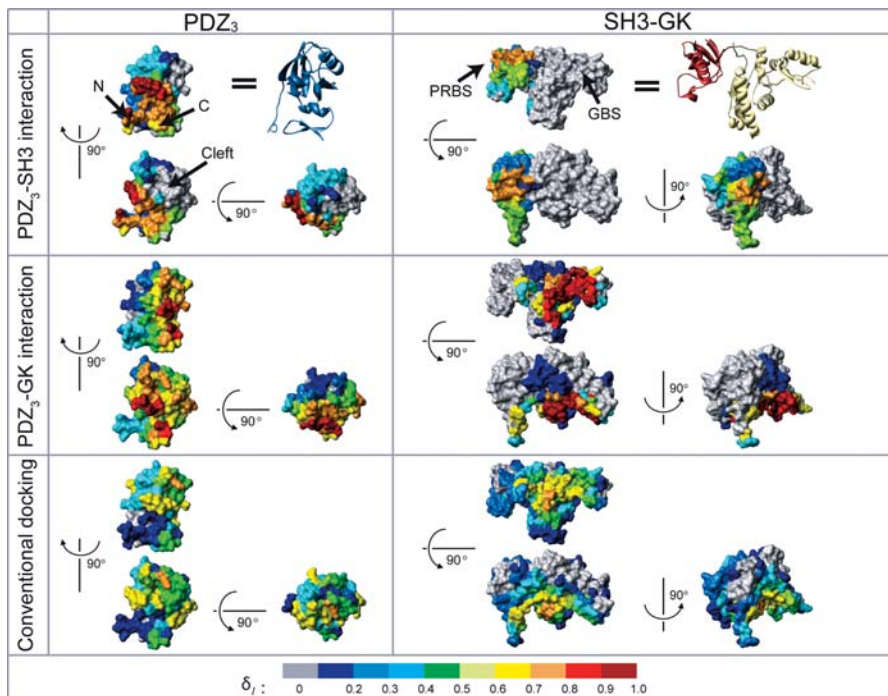
Figure 4. The Localization of Binding Sites for Both Modeled Configurations of the PDZ$_3$–SH3–GK Core Fragment Compared with Protein Docking

Top ten scoring models were selected for both interactions (PDZ$_3$–SH3, PDZ$_3$–GK) obtained using comparative patch analysis and using conventional protein docking. The localization index δ_I of a residue defines the relative frequency of its participation in the interaction interface. The residues that are colored grey do not participate in the interface of any of the top ten models. The PRBS in SH3 and the GBS in GK are shown by arrows.

doi:10.1371/journal.pcbi.0020153.g004

Matrix-assisted laser desorption ionization (MALDI) analysis of peptides generated by tryptic digestion of this band indicates that it represents the sequence from residues 300 to 721, which corresponds to the PDZ$_3$ and SH3–GK domains (mass accuracy, Δppm $\leq$ 13). Further digestion leads to the disappearance of the PDZ$_3$–SH3–GK entity and the appearance of a stable ~34 kDa fragment. The 34-kDa band was identified by MALDI analysis as the SH3–GK domains, encompassing residues 429 to 721 (Δppm $\leq$ 10 for all detected peptides). Cleavage with

thermolysin, another nonspecific protease, generates similarly sized stable fragments (unpublished data).

Discussion

We have introduced comparative patch analysis, an approach to the modeling of a complex between two subunit structures, and applied it to the protein PSD-95, a key neural-signaling scaffold. The approach relies on structurally defined interactions of each of the complex components, or their homologs, with any other subunit, irrespective of its fold (Figure 1). We assessed comparative patch analysis for its increased applicability relative to comparative modeling as well as increased accuracy relative to conventional protein docking (Figure 2, Table 1). Next, comparative patch analysis was applied to model the structure of a core fragment of rat PSD-95, containing the PDZ$_3$, SH3, and GK domains, resulting in two predicted configurations (Figures 3 and 4). The model was experimentally supported by limited proteolysis (Figure 5). In addition, the prediction is in concordance with and rationalizes available biochemical, structural, and evolutionary data (Figures 3 and 4, Table 2).

Comparative Patch Analysis

By limiting the configurational search to the known binding modes of the homologous subunits and applying a physical assessment of candidate complex structures, comparative patch analysis benefits from the advantages of both homology-driven and physics-driven docking. Its coverage is larger than that of comparative modeling and its accuracy is higher than that of protein docking (Figure 2), although the

Table 2. Cross-Species Analysis of PXXP Motif in PSD-95 Proteins

Motif	Position in PSD-95 Sequence	Accession ID	Species	Splice Variant?
PREP	PDZ$_3$ domain	P78352	Human	No
		P31016	Rat	No
		Q62108	Mouse	No
PRAP	N-terminal region	P78352–2	Human	Isoform 2
		Q62108–2	Mouse	Isoform 2
PAKP	N-terminal region	Q6R005	Zebrafish	No
PTSP	A linker between PDZ$_2$ and PDZ$_3$	P78352	Human	No
		Q62108	Mouse	No
		P31016	Rat	No
PSSP	A linker between PDZ$_2$ and PDZ$_3$	Q6R005	Zebrafish	No
RVAK	C-terminal region	Q6R005	Zebrafish	No

The human, rat, and mouse proteins all contained a PREP motif in PDZ$_3$; the zebrafish protein did not. Five other potential SH3 binding motifs were found outside of known domains; two at the N-terminus, one at the C-terminus, and two in the interdomain linker between PDZ$_2$ and PDZ$_3$. The conservation of the PREP sequence in PDZ$_3$ from mammalian species suggests that its interaction with SH3 may be functionally significant.
doi:10.1371/journal.pcbi.0020153.t002

A

B
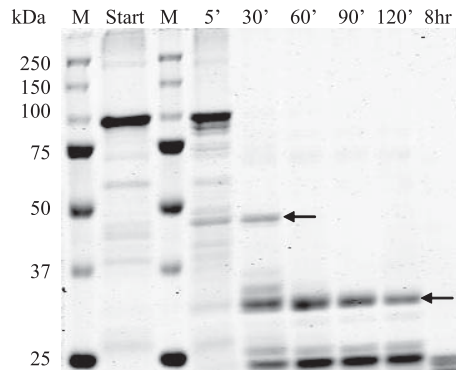
```
  1 mdclcivttk kyryqdedtp plehspahlp nqansppviv ntdtleapgy elqvngtege
 61 meyeeitler gnsglgfsia ggtdnphigd dpsifitkii pggaaaqdgr lrvndsilfv
121 nevdvrevth saavealkea gsivrlyvmr rkppaekvme iklikgpkgl gfsiaggvgn
181 qhipgdnsiy vtkiieggaa hkdgrlqigd kilavnsvgl edvmhedava alkntydvvy
241 lkvakpsnay lsdsyappdi ttsysqhldn eishssylgt dyptamtpts prryspvakd
301 llgeedipre prrivihrgs tglgfnivgg edgegifisf ilaggpadls gelrkgdqil
361 svngvdlrna sheqaaialk nagqtvtiia qykpeeysrf eakihdlreq lmnsslgsgt
421 aslrsnpkrg fyiralfdyd ktkdcgflsq alsfrfgdvl hvidagdeew wqarrvhsds
481 etddigfips krrverrews rlkakdwgss sgsqgredsv lsyetvtqme vhyarpiiil
541 gptkdrandd llsefpdkfg scvphttrpk reyeidgrdy hfvssrekme kdiqahkfie
601 agqynshlyg tsvqsvreva eqgkhcildv sanavrrlqa ahlhpiaifi rprslenvle
661 inkriteeqa rkafdratkl eqeftecfsa ivegdsfeei yhkvkrvied lsgpyiwvpa
721 rerl
```

Figure 5. The PDZ₃–SH3–GK and SH3–GK Domains Are Stable Fragments

(A) Coomassie-stained gel (10% acrylamide) of aliquots from limited proteolysis of PSD-95 by Subtilisin proteinase: panels 1 and 3, Precision Plus Protein molecular weight marker (Bio-Rad, http://www.bio-rad.com); panel 2, starting sample prior to proteinase addition; panels 4, Lanes 4–9, Aliquots at 5, 30, 60, 90, 120 min, and 8 h after protease addition (as labeled). Arrows point to stable fragments that were excised from the gel and analyzed by mass spectrometry as described in Methods.

(B) Sequence of Rat PSD-95: underlined are the peptide sequences identified by mass spectrometry from the ~34 kDa stable fragment corresponding to residues 429–721 (33,944 kDa). In bold are the sequences derived from the ~48 kDa stable fragment comprising residues 300–721 (47,796 kDa).

doi:10.1371/journal.pcbi.0020153.g005

coverage and accuracy are lower than those of protein docking and comparative modeling, respectively.

At least one binding site is available for 1,989 of the 3,114 total Structural Classification of Proteins (SCOP) domain families (release 1.69, July 2005). Eight hundred fifty of these families contain between ten and 100 binding sites, allowing the exhaustive pairwise docking that is currently required. Thus, the applicability of comparative patch analysis extends to approximately 41%, and in the current implementation is computationally feasible for 8%, of the ~4,850,000 theoretically possible binary domain–domain interactions. The coverage of conventional protein docking is 100%, while the comparative modeling approach is applicable to only 2,126 pairs of families, which constitutes 0.06% of the theoretically possible interactions.

When compared with protein docking, comparative patch analysis was able to correctly identify the binding mode in 40% more benchmark complexes, predicting the overall structure of the complexes with an average improvement in all-atom RMS error of 13.4 Å. The method also exhibits robustness to small errors in the locations of the specified binding sites, due to the configurational search performed by the docking procedure. In the benchmark set of complexes with known structures, a minimal threshold of 75% overlap between the initially specified and resulting refined binding sites captured all but one of the good models (LRMS error less than 3 Å), while allowing no false positives.

PSD-95 Protein: Predicting the Structure of the Core Fragment by Analogy

Evolutionary and experimental evidence for intermolecular interaction between PDZ₃ and SH3–GK domains. When modeling the structure of the PDZ₃–SH3–GK fragment, we assumed an interaction between the PDZ₃ and SH–GK domains. PDZ₃ is a good candidate for interaction with the SH3–GK domains because it is immediately upstream of SH3, separated by a relatively short 14-residue linker. To investigate whether or not PDZ₃ interacts with SH3–GK, the analysis of domain co-occurrence, as well as limited proteolysis, were applied.

A survey of the domain architectures of proteins that contain both SH3 and GK domains revealed that the proteins either do not have other domains or also contain at least one PDZ domain always preceding the SH3–GK tandem domain. The minimal architecture that contains at least one PDZ, SH3, and GK domain consists of only these three domains. This pattern strongly suggests a physical interaction between the SH3–GK tandem and the preceding PDZ domain [51,52].

The stable fragments resulting from limited proteolysis of PSD-95 by nonspecific proteases reflect the cleavage of accessible loops, rather than cleavage at a particular substrate sequence. We identified stable PDZ₃–SH3–GK and SH3–GK fragments by mass spectrometry, demonstrating susceptibility of PSD-95 to protease cleavage at sites between the PDZ₂ and

PDZ$_3$ domains and between the PDZ$_3$ and SH3–GK domains. Limited proteolysis with trypsin (unpublished data) also supports the conclusion that the PDZ$_3$ and SH3–GK domains are stable protein structures. These data are consistent with intramolecular interactions between the PDZ$_3$ and the SH3–GK domains of PSD-95.

Application of comparative patch analysis. Modeling the structure of the core PSD-95 fragment is challenging for a number of reasons. First, the structures of neither PDZ–SH3 nor PDZ–GK complexes are available, rendering comparative modeling inapplicable in this case. Moreover, conventional protein docking results were ambiguous, generating a varied ensemble of PDZ$_3$ and SH3–GK complexes without a predominant binding mode (Figure 3C). On the other hand, each of the domain families is known to repeatedly utilize a small number of binding sites for different protein interactions. For instance, PDZ domains bind the C-termini of several different proteins through its hydrophobic cleft [42,53]. Similarly, the PRBS of SH3 domains recognizes PXXP-sequence motifs in a variety of proteins [45,46]. These observations suggest that comparative patch analysis is suited for modeling the PSD-95 core fragment.

Functional roles of the predicted configurations. Comparative patch analysis of the PDZ$_3$–SH3–GK fragment found two possible configurations that satisfied all imposed spatial restraints, including previously observed binding sites, consistency with the given linker length, and physicochemical complementarity of the interacting surfaces. In addition, the ensemble of models produced by comparative patch analysis for each interaction type (PDZ$_3$–SH3, PDZ$_3$–GK) exhibited a single predominant binding mode. The binding sites forming the interaction interfaces of these models are located at the same or similar regions of the protein surface (Figure 4). Therefore, the binding modes are predicted with relative confidence. Multiple stable configurations of PSD-95 and its close homologs have recently been suggested independently based on biochemical studies [40,54] and single-particle electron microscopy experiments [41]. As we describe below, we suggest the two binding modes have clear functional implications.

The two predicted configurations exhibit structural properties that suggest unique functional roles. In the first configuration, the hydrophobic cleft of the PDZ domain and the GBS of the GK domain are both accessible, suggesting that this configuration corresponds to an active state in which binding of other proteins at these two sites can occur (Figure 3B). These binding sites are thought to mediate intermolecular interactions essential for the scaffolding role of PSD-95 [42,49,55–57]. In contrast, both binding sites are buried in the second configuration, by the interface between the PDZ$_3$ and GK domains (Figure 3C), which is suggestive of an alternative functional state. This second configuration points to an efficient intramolecular regulatory mechanism for switching the functional state with a single interaction. Similar regulatory mechanisms have been observed in other signaling networks, such as the TCR and MAPK systems [58,59], indicating this regulation may be a general feature of signaling pathways.

This two-state model also provides a structural explanation for the change in binding affinity between the GK domain and MAP1A protein in the presence of the PDZ$_3$ domain [60]. It has been shown that the GK domain alone is able to bind MAP1A. In the presence of PDZ$_3$, this binding affinity is dramatically reduced. The affinity is recovered upon titration of a C-terminal peptide of CRIPT known to specifically interact with the hydrophobic cleft of PDZ$_3$. This competitive binding suggests that binding to MAP1A and binding to PDZ$_3$ are mediated by the same GK binding site. Our model is in complete agreement with this hypothesis and provides a structural explanation for these observations.

It is known that SH3 domains bind proteins with PXXP sequence motifs through their proline-rich binding regions. The proximity of the PDZ$_3$ PXXP motif to the SH3 PRBS in the first configuration proposed by comparative patch analysis is consistent with the classical SH3–PXXP motif recognition. A similar PXXP-mediated intermolecular PDZ–SH3 interaction has been previously suggested to occur in syntenin [61]. Sequence analysis of PSD-95 from different species indicates that PXXP motifs are not found in its other two PDZ domains, although such motifs are found in the PDZ$_2$–PDZ$_3$ linkers and the flexible N-terminus (Table 2). Recent studies have demonstrated the importance of disordered regions in binding events [62], suggesting that future investigation of interactions of these PXXP motifs using recently developed flexible docking algorithms [63] should prove fruitful.

The limited proteolysis experiment (Figure 5) is a first step to verifying the intramolecular interactions suggested by comparative patch analysis. The two functional states hypothesis, outlined in the Discussion, points to a number of experiments that could shed light on the structure and function of PSD-95. First, the proposed regulation of the PSD-95 activity by PDZ$_3$–specific C-terminal peptides can be further tested using immunoprecipitation and yeast two-hybrid experiments similar to those performed for other GK-mediated interactions [60] (e.g., with the GKAP protein [57]). If the proposed regulation mechanism is verified, experimental control of the PSD-95 activity may become possible, enabling detailed study of the functional differences between the two states. Next, the intramolecular interactions proposed here can be tested by a variety of experimental techniques [64], including NMR spectroscopy [65], site-directed mutagenesis [66], hydrogen/deuterium exchange combined with mass spectrometry [67], and small angle X-ray scattering (SAXS) [68]. In particular, site-directed mutagenesis [66] of the interface residues in the first proposed state (see Datasets S1 and S2) could be used with pull-down assays to validate the predicted interaction interface [69]. In addition, the lack of accessibility of the GBS in the second state could be tested using nucleotide-binding assays [70,71]. Finally, the shapes of the calculated SAXS spectra for the best-scoring models in both conformations are substantially different (Figure 3). Thus, we expect the experimentally obtained SAXS spectra to be helpful in distinguishing between the two PSD-95 states.

Comparative patch analysis for characterizing the quaternary structure of protein assemblies provides a framework for combining data from known protein structures with a physical assessment of protein interactions. This framework will benefit from future developments in protein–protein docking, such as the explicit treatment of flexibility and more accurate scoring functions. We are currently developing an automated comparative patch analysis pipeline for large-scale modeling of protein complexes via a Web server. In closing, we expect that comparative patch analysis will

provide useful spatial restraints for the structural characterization of an increasing number of binary and higher order protein complexes, as it did for PSD-95.

Methods

Comparative patch analysis protocol. We start by outlining the steps in comparative patch analysis, followed by a more detailed description. First, for each partner domain in a binary complex, a set of protein binding sites of its homologs represented in PIBASE was identified [72]. Second, these binding sites were mapped onto the partner domain surface using structure-based alignments between the domain and each of its homologs. Third, all pairs of the mapped binding sites were converted by restrained docking to obtain candidate models of the binary complex. This ensemble of models was then ranked using a measure of geometric complementarity and a statistical potential score.

Extracting and mapping binding sites of domain homologs. For each of the two partner domains, we first defined a family of its homologs. Several schemes both dissect proteins into domains and cluster them into families, based on sequence, structure, and/or function [73–76]. We used the family definitions in SCOP [73]. Domains that belong to the same SCOP family usually share at least 30% sequence identity or the same biological function.

For a given SCOP family, the set of binary domain interfaces between its members and other domains was obtained from PIBASE, our comprehensive relational database of all structurally characterized interfaces between pairs of protein domains [72]. The domain–domain interfaces in PIBASE were extracted from protein structures in the Protein Data Bank (PDB) [77] and Protein Quaternary Structure (PQS) server [78] using domain definitions from the SCOP and CATH domain classification systems [73,74]. An interface is defined by a list of pairs of residues, one from each protein, that are in contact with each other. Each binding site consists of the residues that are within 6.05 Å of its partner domain, where the threshold is defined between any two nonhydrogen atoms.

The binding site residues from all domain family members were then mapped onto the partner domain using structure-based alignments obtained by DaliLite. DaliLite uses a Monte Carlo procedure to find the best alignment by optimizing a similarity score defined in terms of equivalent intramolecular distances [79].

Modeling protein complexes. The structures of binary protein complexes were predicted by restrained docking using the PatchDock software [80,81]. PatchDock uses an algorithm for rigid body docking that searches for the maximal geometric complementarity between two protein structures, optionally restrained by having to match two user-specified binding sites. Here, we provided all pairs of mapped binding sites, one from each target domain, as input for individual PatchDock runs. When a resulting refined model was inconsistent with the specified binding sites, it was discarded. More specifically, a model was considered not to correspond to a specified binding site interaction if the binding sites predicted by docking had less than 75% of their residues in common with the specified binding sites (the normalization is based on the size of the smaller of the specified and predicted binding sites).

The resulting binary complexes were scored using a combination of two independent scores, the geometric complementarity function of PatchDock and DOPE (Discrete Optimized Protein Energy) score. DOPE is a distance-dependent pairwise statistical potential calculated from known protein structures and available through the MODELLER program [82,83]. The configurations in the ensemble of models were ranked by a sum of the PatchDock and DOPE scores, first scaled to lie in the range between 0 and 1.

Assessment of comparative patch analysis. A benchmark set of 20 binary domain complexes was used to evaluate comparative patch analysis (Table 1). These complexes were divided into two groups. Each subunit of a complex in the first group is a member of a SCOP family that has been observed to interact with only one other SCOP family. In contrast, each subunit from the second group of complexes comes from a SCOP family that has been observed to interact with multiple SCOP families. The complexes were randomly selected from PIBASE such that the number of interactions available for the families of each component ranged between ten and 100. In total, there are 11 protein complexes (noncovalently linked domains) and nine multidomain proteins (covalently linked domains) in the benchmark set.

As in previous data-dependent approaches for modeling the structures of protein interactions [18,84,85], we have tested our method using a benchmark set designed within its scope of

applicability. Our method is applicable only to protein complexes for which structures of the subunits or their homologs interacting with other proteins are available. This constraint on applicability also applies to the benchmark structures used to test the method. For this reason, we did not use the two benchmark sets that are generally used for protein docking methods, the set of CAPRI targets [16,86] and a benchmark set developed by Weng and coworkers [87]. The set of 19 CAPRI targets, whose structures are publicly available, was not an appropriate benchmark for our method because the majority of the structures either (i) contain subunits consisting of multiple SCOP domains ($n = 7$: T02–T07, T19), (ii) are not annotated by SCOP ($n = 4$: T09, T13, T20, T21), or (iii) there are no observed binding sites available for patch analysis ($n = 4$: T11, T12, T15, T19). This leaves five structures (T01, T08, T10, T14, T18) on which comparative patch analysis can be tested. Similarly, of the 63 rigid-body docking targets in the Weng benchmark set, 37 contain subunits with multiple SCOP domains and two contain subunits for which there are no observed binding sites available for comparative patch analysis. The remaining 24 targets contain subunits for which there is an average of 850 binding sites available for our method. This number of binding sites makes comparative patch analysis computationally very expensive, requiring on average more than two million localized docking calculations per target. There are only five targets in the Weng set that require no more than ten thousand calculations, the threshold we used in selecting our benchmark set. We are currently developing a method to cluster binding sites that would allow a significant reduction in the number of docking calculations required for a target structure, enabling the use of a more comprehensive benchmark set.

Adapting existing benchmarks to assess our method required ad hoc processing such as assigning domain boundaries and classifications, dissecting multidomain complexes into binary domain interactions, and reducing the number of input binding sites. Instead, we developed a benchmark set that is applicable to our method in an automated fashion. In addition, our benchmark set was designed to assess the performance of comparative patch analysis for domain–domain interactions in both multidomain proteins and protein complexes. The targets in the CAPRI and Weng benchmark sets are exclusively protein–protein interaction structures.

To quantify the amount of additional information provided by comparative patch analysis relative to docking, the structure of each protein complex was modeled using three independent protocols, relying on the docking program PatchDock (Methods). In the first protocol, known binding sites for the homologs of both subunits were used to restrain the docking. In the second protocol, known binding sites for the homologs of only one subunit were used to restrain the docking. In the final protocol, no binding site information was used, and conventional protein docking was applied.

Distance metrics. To evaluate the accuracy of comparative patch analysis in predicting the interaction interface and relative orientation of two structurally defined protein domains, the following three measures were used: binding site overlap, interface overlap, and RMS error.

First, we calculate the binding site overlap (O_B), which we define as the percentage of correctly predicted binding site residues:

$$O_B = \frac{1}{2}\left[\frac{N(B_1^{pred} \cap B_1^{exp})}{N(B_1^{pred} \cup B_1^{exp})} + \frac{N(B_2^{pred} \cap B_2^{exp})}{N(B_2^{pred} \cup B_2^{exp})}\right],$$

where $N(B_i^{pred} \cap B_i^{exp})$ is the number of residues in common between the predicted and actual binding sites, and $N(B_i^{pred} \cup B_i^{exp})$ is the total number of contact residues in both binding sites.

Next, we used the interface overlap (O_I), as a measure to assess the predicted interface between the binding sites:

$$O_I = \frac{N(I_{pred} \cap I_{native})}{N(I_{pred} \cup I_{native})},$$

where $N(I_{pred} \cap I_{native})$ is the number of residue contacts in common between the predicted (I_{pred}) and native (I_{native}) interfaces, and $N(I_{pred} \cup I_{native})$ is the total number of residue contacts. Interfaces were deemed to be correct when at least half of the residue contacts were identified.

Finally, we calculated the all-atom RMS error between the predicted and native complexes using the L_RMS measure defined in CAPRI [88]. The predicted and native structures were superposed using the larger of the two domains, and the RMS error was calculated for the other domain.

Modeling the PDZ$_3$–SH3–GK complex of rat PSD-95. *Comparative patch analysis application.* Comparative patch analysis was used to predict the tertiary structure of the rat PSD-95 core fragment that

contains the PDZ₃, SH3, and GK domains. From PIBASE, 126, 298, and 517 protein binding sites were obtained for the PDZ₃, SH3, and GK domains, respectively. The binding sites were mapped onto the target structures. Redundant binding sites were removed so that no pair of binding sites shared more than 95% of their residues, leaving 49, 26, and 219 binding sites for the PDZ₃, SH3, and GK, respectively. The comparative patch analysis protocol was then applied.

We then assessed whether the models were compatible with the 14-residue linker length between the PDZ₃ and SH3 domains. To do so, the linker was modeled as a flexible chain of 14 spheres with 1.9 Å radii and a maximum distance of 3.8 A between consecutive spheres, to mimic the excluded volume of the linker and restrict the maximum spatial separation of the domains. Each model was assessed using the following protocol in MODELLER [83]. First, the positions of the 14 linker residues were placed at random coordinates and then optimized using simulated annealing molecular dynamics and conjugate gradient minimizations. The scoring function consists of terms equal to $\frac{(f-f_0)^2}{\sigma^2}$, where f is the restrained distance and σ is the parameter that regulates the strength of the term. Linker distances are restrained if $f > f_0$, where $f_0 = 3.8$ and $\sigma = 0.05$. Excluded volume restraints between the protein and the linker are imposed if $f > f_0$, where f_0 is the sum of the atomic and linker radii and $\sigma = 0.01$. The optimization of the scoring function was performed in 20 independent trials for each model, and the optimized coordinates of the linker residues with the lowest score were added to the model. As a result of assessment, those models that violated the imposed linker restraints and thus could not have an interdomain linker of such length between PDZ₃ and SH3 domains were removed from the ensemble.

Exhaustive docking. The PDZ₃–SH3–GK models built by comparative patch analysis were compared with those built by exhaustive docking using PatchDock without prior information about the potential binding site [80,81]. The model with the best PatchDock–DOPE score that satisfied the interdomain linker restraint was selected.

Sequence analysis. The SMART domain annotation tool was used to search for proteins containing the PDZ, SH3, and GK domains [89,90]. Proteins and splice variants annotated as PSD-95 proteins were obtained from the UniProt sequence database [91]. The sequences were scanned for known SH3 binding motifs (PXXP, PXXDY, RXXK [47]) using grep regular expression search.

Proteolysis of PSD-95. Rat PSD-95 was cloned into pET47b (+) and expressed as a His-tagged fusion protein (~83.4 kDa) in BL21 (DE3) pLysS cells at 37 °C. Cells were harvested 3–3.5 h after induction by 0.4 mM IPTG. The cell lysate was centrifuged at 17K RPM, and the supernatant was loaded onto a nickel NTA column (Qiagen, http://www1.qiagen.com) and eluted with an imidazole gradient (20 mM to 500 mM). The purest fractions were exchanged (using PD10 columns, Amersham Biosciences, http://www.amersham.com) to: 20 mM Tris (pH 8), 150 mM NaCl, 5 mM DTT, 10% glycerol for limited proteolysis (protocol based on that of Stroh et al. [92]). Digests of PSD-95 were initiated by adding protease to the following final concentrations: 0.83 μg/ml sequencing grade modified Trypsin (Roche, http://www.amersham.com), 0.1 μg/ml of proteinase (Fluka, http://www.sigmaaldrich.com), or 8.3 μg/ml of thermolysin (Sigma, http://www.sigmaaldrich.com). The thermolysin reaction was also supplemented with 5 mM CaCl₂. Digests were incubated at 37 °C and stopped with 5 mM PMSF for trypsin and proteinase, and 10 mM EDTA for thermolysin. Aliquots were taken at 5, 30, 60, 90, 120 min, and 8 h after addition of protease and flash frozen in liquid nitrogen until analysis by SDS-PAGE. Stable fragments were excised from Coomassie-stained gels and subjected to tryptic digestion in the gel piece after reduction with DTT and alkylation with iodoacetamide [92,93]. The tryptic peptides were extracted from gel slices with 5% formic acid in 50% acetonitrile, concentrated in a SpeedVac (Savant Instruments, http://www.combichemlab.com), and desalted with the

use of a Zip Tip (Millipore, http://www.millipore.com) before analysis by MALDI–TOF (matrix-assisted laser desorption ionization–time of flight) mass spectrometry. Samples were mixed with either α-cyanohydroxycinnamic acid or a "Universal" MALDI matrix from Fluka. Analyses were performed with a Voyager DE-PRO MALDI–TOF mass spectrometer (Applied Biosystems, http://www.appliedbiosystems.com) that was first externally calibrated using a calibration mix supplied by the manufacturer. The MALDI spectra were recalibrated internally with known peptide masses, e.g., trypsin autolysis peaks or expected masses obtained from in silico digests of the known protein. The software, Prospector MSFIT (University of California San Francisco), was used to identify the tryptic fragments.

The representations of proteins in Figures 1–3 were obtained using Chimera [94], and in Figure 4 with the help of MolMol software [95].

Supporting Information

Dataset S1. The Best Model of the First Configuration of PSD-95 Core Fragment

Found at doi:10.1371/journal.pcbi.0020153.sd001 (229 KB TXT).

Dataset S2. The Best Model of the Second Configuration of PSD-95 Core Fragment

Found at doi:10.1371/journal.pcbi.0020153.sd002 (229 KB TXT).

Accession Numbers

Accession numbers from the Protein Data Bank (http://rcsb.org) for the proteins mentioned in this paper are: rat PSD-95 GK (1JXM), rat PSD-95 PDZ₃ (1BE9), rat PSD-95 PDZ₃ (1BFE), rat PSD-95 SH3 and GK (1JXO), pyruvate formate–lyase protein complex (1CM5).
Accession numbers from the European Bioinformatics Institute SMART database (http://www.ebi.ac.uk/interpro/) for proteins mentioned in this paper are: PSD-95 PDZ (SM00228), PSD-95 SH3 (SM00326), and PSD-05 GK (SM00072).

Acknowledgments

References

1. Pawson T, Nash P (2003) Assembly of cell regulatory systems through protein interaction domains. Science 300: 445–452.
2. Alberts B, Miake-Lye R (1992) Unscrambling the puzzle of biological machines: The importance of the details. Cell 68: 415–420.
3. Park J, Lappe M, Teichmann SA (2001) Mapping protein family interactions: Intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast. J Mol Biol 307: 929–938.
4. Edwards AM, Kus B, Jansen R, Greenbaum D, Greenblatt J, et al. (2002) Bridging structural biology and genomics: Assessing protein interaction data with known complexes. Trends Genet 18: 529–536.
5. Sali A, Glaeser R, Earnest T, Baumeister W (2003) From words to literature in structural proteomics. Nature 422: 216–225.
6. Aloy P, Bottcher B, Ceulemans H, Leutwein C, Mellwig C, et al. (2004) Structure-based assembly of protein complexes in yeast. Science 303: 2026–2029.
7. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, et al. (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29: 291–325.
8. Aloy P, Russell RB (2003) InterPreTS: Protein interaction prediction through tertiary structure. Bioinformatics 19: 161–162.
9. Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, et al. (2004) MODBASE, a database of annotated comparative protein structure models, and associated resources. Nucleic Acids Res 32 (Database issue): D217–D222.
10. Lu L, Lu H, Skolnick J (2002) MULTIPROSPECTOR: An algorithm for the prediction of protein–protein interactions by multimeric threading. Proteins 49: 350–364.
11. Lu L, Arakaki AK, Lu H, Skolnick J (2003) Multimeric threading-based

prediction of protein–protein interactions on a genomic scale: Application to the *Saccharomyces cerevisiae* proteome. Genome Res 13: 1146–1154.

12. Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: An overview of search algorithms and a guide to scoring functions. Proteins 47: 409–443.

13. Smith GR, Sternberg MJ (2002) Prediction of protein–protein interactions by docking methods. Curr Opin Struct Biol 12: 28–35.

14. Wodak SJ, Janin J (2002) Structural basis of macromolecular recognition. Adv Protein Chem 61: 9–73.

15. Wodak SJ, Mendez R (2004) Prediction of protein–protein interactions: The CAPRI experiment, its evaluation and implications. Curr Opin Struct Biol 14: 242–249.

16. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, et al. (2003) CAPRI: A critical assessment of predicted interactions. Proteins 52: 2–9.

17. Clore GM, Schwieters CD (2003) Docking of protein–protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from 1H/15N chemical shift mapping and backbone 15N–1H residual dipolar couplings using conjoined rigid body/torsion angle dynamics. J Am Chem Soc 125: 2902–2912.

18. Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: A protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125: 1731–1737.

19. Schulz DM, Ihling C, Clore GM, Sinz A (2004) Mapping the topology and determination of a low-resolution three-dimensional structure of the calmodulin–melittin complex by chemical cross-linking and high-resolution FTICRMS: Direct demonstration of multiple binding modes. Biochemistry 43: 4703–4715.

20. van Dijk AD, Boelens R, Bonvin AM (2005) Data-driven docking for the study of biomolecular complexes. FEBS J 272: 293–312.

21. Korkin D, Davis FP, Sali A (2005) Localization of protein-binding sites within families of proteins. Protein Sci 14: 2350–2360.

22. Cho KO, Hunt CA, Kennedy MB (1992) The rat brain postsynaptic density fraction contains a homolog of the *Drosophila* discs-large tumor suppressor protein. Neuron 9: 929–942.

23. Hunt CA, Schenker LJ, Kennedy MB (1996) PSD-95 is associated with the postsynaptic density and not with the presynaptic membrane at forebrain synapses. J Neurosci 16: 1380–1388.

24. Kennedy MB (1997) The postsynaptic density at glutamatergic synapses. Trends Neurosci 20: 264–268.

25. Hata Y, Takai Y (1999) Roles of postsynaptic density-95/synapse-associated protein 90 and its interacting proteins in the organization of synapses. Cell Mol Life Sci 56: 461–472.

26. Roche KW (2004) The expanding role of PSD-95: A new link to addiction. Trends Neurosci 27: 699–700.

27. Koulen P, Fletcher EL, Craven SE, Bredt DS, Wassle H (1998) Immunocytochemical localization of the postsynaptic density protein PSD-95 in the mammalian retina. J Neurosci 18: 10136–10149.

28. Kennedy MB (2000) Signal-processing machines at the postsynaptic density. Science 290: 750–754.

29. Romorini S, Piccoli G, Jiang M, Grossano P, Tonna N, et al. (2004) A functional role of postsynaptic density-95-guanylate kinase-associated protein complex in regulating Shank assembly and stability to synapses. J Neurosci 24: 9391–9404.

30. van Zundert B, Yoshii A, Constantine-Paton M (2004) Receptor compartmentalization and trafficking at glutamate synapses: A developmental proposal. Trends Neurosci 27: 428–437.

31. Cubelos B, Gonzalez-Gonzalez IM, Gimenez C, Zafra F (2005) The scaffolding protein PSD-95 interacts with the glycine transporter GLYT1 and impairs its internalization. J Neurochem 95: 1047–1058.

32. Anderson JM (1996) Cell signalling: MAGUK magic. Curr Biol 6: 382–384.

33. Gonzalez-Mariscal L, Betanzos A, Avila-Flores A (2000) MAGUK proteins: Structure and role in the tight junction. Semin Cell Dev Biol 11: 315–324.

34. Long JF, Tochio H, Wang P, Fan JS, Sala C, et al. (2003) Supramodular structure and synergistic target binding of the N-terminal tandem PDZ domains of PSD-95. J Mol Biol 327: 203–214.

35. Tochio H, Hung F, Li M, Bredt DS, Zhang M (2000) Solution structure and backbone dynamics of the second PDZ domain of postsynaptic density-95. J Mol Biol 295: 225–237.

36. Doyle DA, Lee A, Lewis J, Kim E, Sheng M, et al. (1996) Crystal structures of a complexed and peptide-free membrane protein-binding domain: Molecular basis of peptide recognition by PDZ. Cell 85: 1067–1076.

37. McGee AW, Dakoji SR, Olsen O, Bredt DS, Lim WA, et al. (2001) Structure of the SH3-guanylate kinase module from PSD-95 suggests a mechanism for regulated assembly of MAGUK scaffolding proteins. Mol Cell 8: 1291–1301.

38. Tavares GA, Panepucci EH, Brunger AT (2001) Structural characterization of the intramolecular interaction between the SH3 and guanylate kinase domains of PSD-95. Mol Cell 8: 1313–1325.

39. Yaffe MB (2002) MAGUK SH3 domains—Swapped and stranded by their kinases? Structure 10: 3–5.

40. Fukunaga Y, Matsubara M, Nagai R, Miyazawa A (2005) The interaction between PSD-95 and Ca2+/calmodulin is enhanced by PDZ-binding proteins. J Biochem (Tokyo) 138: 177–182.

41. Nakagawa T, Futai K, Lashuel HA, Lo I, Okamoto K, et al. (2004) Quaternary structure, protein dynamics, and synaptic function of SAP97 controlled by L27 domain interactions. Neuron 44: 453–467.

42. Kornau HC, Schenker LT, Kennedy MB, Seeburg PH (1995) Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. Science 269: 1737–1740.

43. Niethammer M, Kim E, Sheng M (1996) Interaction between the C terminus of NMDA receptor subunits and multiple members of the PSD-95 family of membrane-associated guanylate kinases. J Neurosci 16: 2157–2163.

44. Ren R, Mayer BJ, Cicchetti P, Baltimore D (1993) Identification of a ten-amino acid proline-rich SH3 binding site. Science 259: 1157–1161.

45. Mayer BJ (2001) SH3 domains: Complexity in moderation. J Cell Sci 114: 1253–1263.

46. Agrawal V, Kishan KV (2002) Promiscuous binding nature of SH3 domains to their target proteins. Protein Pept Lett 9: 185–193.

47. Zarrinpar A, Bhattacharyya RP, Lim WA (2003) The structure and function of proline recognition domains. Sci STKE 179: RE8.

48. Blaszczyk J, Li Y, Yan H, Ji X (2001) Crystal structure of unligated guanylate kinase from yeast reveals GMP-induced conformational changes. J Mol Biol 307: 247–257.

49. Li Y, Spangenberg O, Paarmann I, Konrad M, Lavie A (2002) Structural basis for nucleotide-dependent regulation of membrane-associated guanylate kinase-like domains. J Biol Chem 277: 4159–4165.

50. Sekulic N, Shuvalova L, Spangenberg O, Konrad M, Lavie A (2002) Structural characterization of the closed conformation of mouse guanylate kinase. J Biol Chem 277: 30236–30243.

51. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, et al. (1999) Detecting protein function and protein–protein interactions from genome sequences. Science 285: 751–753.

52. Vogel C, Berzuini C, Bashton M, Gough J, Teichmann SA (2004) Supradomains: Evolutionary units larger than single protein domains. J Mol Biol 336: 809–823.

53. Nourry C, Grant SG, Borg JP (2003) PDZ domain proteins: Plug and play! Sci STKE 179: RE7.

54. Wu H, Reissner C, Kuhlendahl S, Coblentz B, Reuver S, et al. (2000) Intramolecular interactions regulate SAP97 binding to GKAP. EMBO J 19: 5740–5751.

55. Zhang M, Wang W (2003) Organization of signaling complexes by PDZ-domain scaffold proteins. Acc Chem Res 36: 530–538.

56. Kim E, Sheng M (2004) PDZ domain proteins of synapses. Nat Rev Neurosci 5: 771–781.

57. Kim E, Naisbitt S, Hsueh YP, Rao A, Rothschild A, et al. (1997) GKAP, a novel synaptic protein that interacts with the guanylate kinase-like domain of the PSD-95/SAP90 family of channel clustering molecules. J Cell Biol 136: 669–678.

58. Andreotti AH, Bunnell SC, Feng S, Berg LJ, Schreiber SL (1997) Regulatory intramolecular association in a tyrosine kinase of the Tec family. Nature 385: 93–97.

59. Dueber JE, Yeh BJ, Bhattacharyya RP, Lim WA (2004) Rewiring cell signaling: The logic and plasticity of eukaryotic protein circuitry. Curr Opin Struct Biol 14: 690–699.

60. Brenman JE, Topinka JR, Cooper EC, McGee AW, Rosen J, et al. (1998) Localization of postsynaptic density-93 to dendritic microtubules and interaction with microtubule-associated protein 1A. J Neurosci 18: 8805–8813.

61. Grootjans JJ, Zimmermann P, Reekmans G, Smets A, Degeest G, et al. (1997) Syntenin, a PDZ protein that binds syndecan cytoplasmic domains. Proc Natl Acad Sci U S A 94: 13683–13688.

62. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 272: 5129–5148.

63. Bonvin AM (2006) Flexible protein–protein docking. Curr Opin Struct Biol 16: 194–200.

64. Russell RB, Alber F, Aloy P, Davis FP, Korkin D, et al. (2004) A structural perspective on protein–protein interactions. Curr Opin Struct Biol 14: 313–324.

65. Burz DS, Dutta K, Cowburn D, Shekhtman A (2006) Mapping structural interactions using in-cell NMR spectroscopy (STINT–NMR). Nat Methods 3: 91–93.

66. Kube E, Becker T, Weber K, Gerke V (1992) Protein–protein interaction studied by site-directed mutagenesis. Characterization of the annexin II–binding site on p11, a member of the S100 protein family. J Biol Chem 267: 14175–14182.

67. Maier CS, Deinzer ML (2005) Protein conformations, interactions, and H/D exchange. Methods Enzymol 402: 312–360.

68. Rosenberg OS, Deindl S, Sung RJ, Nairn AC, Kuriyan J (2005) Structure of the autoinhibited kinase domain of CaMKII and SAXS analysis of the holoenzyme. Cell 123: 849–860.

69. Uchino S, Wada H, Honda S, Nakamura Y, Ondo Y, et al. (2006) Direct interaction of post-synaptic density-95/Dlg/ZO-1 domain-containing synaptic molecule Shank3 with GluR1 alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptor. J Neurochem 97: 1203–1214.

70. Kistner U, Garner CC, Linial M (1995) Nucleotide binding by the synapse associated protein SAP90. FEBS Lett 359: 159–163.

71. Song H, Endow SA (1998) Decoupling of nucleotide- and microtubule-binding sites in a kinesin mutant. Nature 396: 587–590.

72. Davis FP, Sali A (2005) PIBASE: A comprehensive database of structurally defined protein interfaces. Bioinformatics 21: 1901–1907.

73. 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: 536–540.

74. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, et al. (1997) CATH—A hierarchic classification of protein domain structures. Structure 5: 1093–1108.

75. Holm L, Sander C (1998) Touring protein fold space with Dali/FSSP. Nucleic Acids Res 26: 316–319.

76. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Barrell D, et al. (2003) The InterPro Database 2003 brings increased coverage and new features. Nucleic Acids Res 31: 315–318.

77. Westbrook J, Feng Z, Jain S, Bhat TN, Thanki N, et al. (2002) The Protein Data Bank: Unifying the archive. Nucleic Acids Res 30: 245–248.

78. Henrick K, Thornton JM (1998) PQS: A protein quaternary structure file server. Trends Biochem Sci 23: 358–361.

79. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. Bioinformatics 16: 566–567.

80. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: Servers for rigid and symmetric docking. Nucleic Acids Res 33: W363–W367.

81. Duhovny D, Nussinov R, Wolfson HJ (2002) Efficient unbound docking of rigid molecules. In: Guigó R, Gusfield D, editors. Lecture notes in computer science. Volume 2452. London: Springer-Verlag. pp. 185–200.

82. Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein. Protein Sci 15: 1–18. doi:10.1110ps.062416606

83. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234: 779–815.

84. Dobrodumov A, Gronenborn AM (2003) Filtering and selection of structural models: Combining docking and NMR. Proteins 53: 18–32.

85. Morelli XJ, Palma PN, Guerlesquin F, Rigby AC (2001) A novel approach for assessing macromolecular complexes combining soft-docking calculations with NMR data. Protein Sci 10: 2131–2137.

86. Janin J (2005) The targets of CAPRI rounds 3–5. Proteins 60: 170–175.

87. Mintseris J, Wiehe K, Pierce B, Anderson R, Chen R, et al. (2005) Protein–Protein Docking Benchmark 2.0: An update. Proteins 60: 214–216.

88. Mendez R, Leplae R, De Maria L, Wodak SJ (2003) Assessment of blind predictions of protein–protein interactions: Current status of docking methods. Proteins 52: 51–67.

89. Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. Proc Natl Acad Sci U S A 95: 5857–5864.

90. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, et al. (2004) SMART 4.0: Towards genomic data integration. Nucleic Acids Res 32: D142–D144.

91. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, et al. (2005) The Universal Protein Resource (UniProt). Nucleic Acids Res 33: D154–D159.

92. Stroh JG, Loulakis P, Lanzetti AJ, Xie J (2005) LC-mass spectrometry analysis of N- and C-terminal boundary sequences of polypeptide fragments by limited proteolysis. J Am Soc Mass Spectrom 16: 38–45.

93. Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 68: 850–858.

94. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem 25: 1605–1612.

95. Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: A program for display and analysis of macromolecular structures. J Mol Graph 14: 51–55, 29–32.