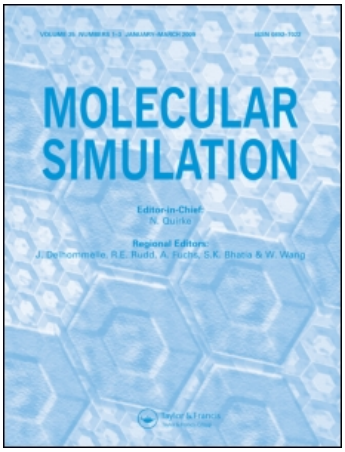

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Interfaces in Molecular Docking

Julie C. Mitchell[a]; Sharokina Shahbaz[a]; Lynn F. Ten Eyck[a]
[a] San Diego Supercomputer Center, University of California at San Diego, La Jolla, CA, USA

To cite this Article Mitchell, Julie C. , Shahbaz, Sharokina and Eyck, Lynn F. Ten(2004) 'Interfaces in Molecular Docking', Molecular Simulation, 30: 2, 97 — 106
To link to this Article: DOI: 10.1080/0892702031000152217
URL: http://dx.doi.org/10.1080/0892702031000152217

Molecular Simulation, VOL. 30 (2–3), 15 February–15 March 2004, pp. 97–106

Taylor & Francis
Taylor & Francis Group

Interfaces in Molecular Docking

JULIE C. MITCHELL*, SHAROKINA SHAHBAZ and LYNN F. TEN EYCK

San Diego Supercomputer Center, University of California at San Diego, 9500 Gilman Dr., MC 0527,La Jolla, CA 92093-0527, USA

(Received August 2002; In final form January 2003)

The Fast Atomic Density Evaluation (FADE) program analyzes shape complementarity using atomic density methods and fast Fourier transforms (FFT). Statistical results for 184 protein–protein and protein–DNA complexes are presented. Almost all of the interfaces studies were found to be complementary, and the average FADE complementarity score highlighted systems known to have strong shape complementarity at the binding interfaces. Also given is a detailed analysis of the interfaces for Fasciculin-Acetylcholinesterase and Barnase–Barstar to show how shape complementarity relates to site mutagenesis experiments. For these two cases, there was good agreement between interface points of highest complementarity and the location of residues known to be important for binding.

Keywords: Fast Atomic Density Evaluation; Fast Fourier transforms; Fasciculin-acetylcholinesterase; Barnase–Barstar; Mutagenesis experiments

INTRODUCTION

We present the results of a broad survey of shape complementarity in protein–protein and protein–DNA interfaces. Our Fast Atomic Density Evaluation (FADE) method rapidly produces a detailed analysis of macromolecular interfaces. The method performed well on a set of 184 distinct interfaces, producing consistent results that highlighted interfaces in which shape complementarity is known to play an important role, such as the Trypsin and Trypsin Inhibitor systems and Protein Kinase A with Protein Kinase Inhibitor.

In addition to providing a single scalar measure of the total complementarity of an interface, FADE can provide a detailed local analysis able to determine specific residues and atoms that are critical to the interaction. This will be demonstrated for the Acetylcholinesterase-Fasciculin and Barnase–Barstar interfaces. For these systems, the shape complementarity "hot spots" indicated residues that are known from mutagenesis experiments to significantly affect the binding affinity.

ATOMIC DENSITY AND SHAPE

The FADE method has previously been proposed as an effective means of analyzing molecular shape and shape complementarity in docking interfaces [1]. FADE uses fast Fourier transforms (FFT) to rapidly compute atomic density at points near the molecular surface. This use of FFT's allows FADE to return excellent shape complementarity results in a few seconds' time.

The relationship between atomic density and shape can be readily understood. Suppose a point x lies near the surface of a protein, and let $N(x, r)$ be a function that gives the number of atomic neighbors within a distance of r Å of the point x. Intuitively, if x lies within a crevice and is surrounded by atoms, $N(x, r)$ will increase rapidly. The opposite behavior occurs near a protrusion, and intermediate behavior is seen near flattish regions (Fig. 1).

To define a scalar value describing the shape, we take the slope, $\lambda(x)$, of a line that is the least-squares fit to $\log N$ against $\log r$, giving an approximation of the form

$$N(x, r) = r^{\lambda(x)} \tag{1}$$

The value of $\lambda(x)$ will generally be higher if x lies in a crevice than if it is near a protrusion. If x lies along a flat edge, $\lambda(x)$ will be approximately 2.8 [2]. In theory,

*Corresponding author. E-mail: mitchell@biochem.wisc.edu

ISSN 0892-7022 print/ISSN 1029-0435 online © 2004 Taylor & Francis Ltd
DOI: 10.1080/0892702031000152217

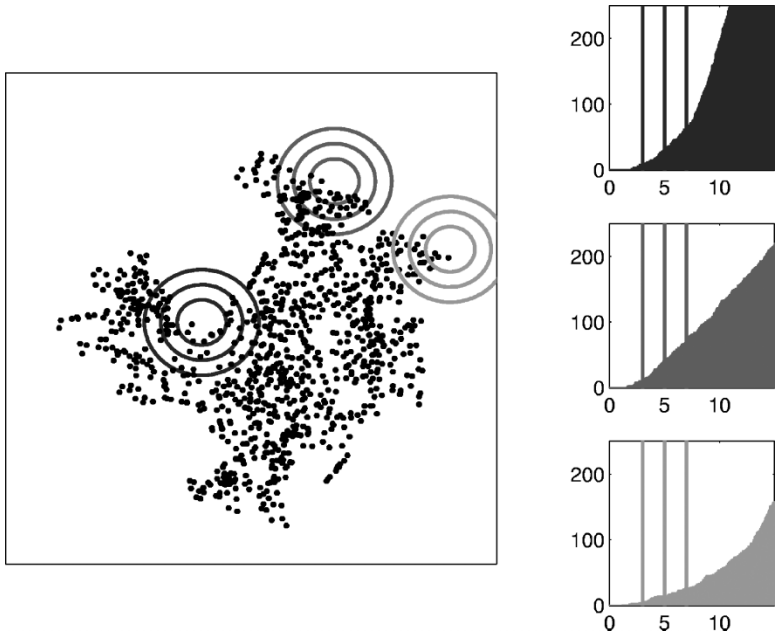

FIGURE 1 The black dots in the large image to the left represent atoms in a protein. The concentric spheres are taken at radii 3, 5 and 7 Å about three points outside the protein surface. Plots to the right indicate the rate of increase in the number of atoms (y-axis) as a function of the radius of the sphere (x-axis). The fast atomic density evaluation (FADE) method deduces molecular shape from the increase in atomic neighbors about a point outside the molecular surface. The concentric dark gray spheres encircle a point with high atomic Density, as can be seen by the rapid rate of increase in the dark gray density profile plot. Points with high atomic density tend to lie in crevices, where they are surrounded by atomic neighbors. The concentric light gray spheres enclose a point with low atomic density that lies near a protrusion. Near flat edges, the density profiles are intermediate between those having high and low density, as can be seen in the medium gray density plot.

this median value, λ_0, should equal 3.0, because the atoms are packed in a three-dimensional configuration. However, for atomic distributions encountered in proteins, this median value is smaller.

For a protein complex, we would like to combine the shape information for separate molecules into a reliable measure of shape complementarity at the docking interface. Fortunately, a relatively simple formula has produced excellent results [1]. For a point x in the interface between two molecules, we define

$$S_c(x) = (\lambda_1(x) - \lambda_0) \cdot (\lambda_2(x) - \lambda_0) \qquad (2)$$

where $\lambda_1(x)$ is the shape measure relative to the first molecule and $\lambda_2(x)$ is computed relative to the second molecule. In the case where the shape features are complementary (i.e.: a crevice and a protrusion), the value of S_c will be negative. If nearby shape features are of the same type (i.e.: crevice–crevice or protrusion–protrusion), the value of S_c will be positive. For two flattish regions, the value of S_c will be close to zero.

The value of S_c gives a local shape complementarity measure. To deduce shape complementarity across an entire docking interface, FADE sums the S_c values over grid points found to be in the interface between the molecules. FADE also computes the average complementarity for all interface points, which is useful when comparing complementarity across different systems. We will see in the next section that the average complementarity lies within a relatively narrow range for 184 protein–protein and protein–DNA interfaces.

UNDERSTANDING SHAPE COMPLEMENTARITY

Atomic density provides a fundamentally different way of looking at shape and shape complementarity when compared with existing methodologies. For this reason, it may be good to provide an more detailed description of the shape measure based on atomic density exponents. Rather that compare numerical values resulting from different types of measures, we will describe them qualitatively and explain why they cannot have good correlations when viewed quantitatively.

Models of shape complementarity using buried surface area or Lennard-Jones potentials are abundant (c.f.: [3–8]). The information produced by the two measures is well-correlated, because the bulk of Lennard-Jones interactions will occur in the region where two molecules are in close contact. Larger interfaces tend to have more favorable Lennard-Jones energies, at least in the absence of atomic collisions. This proves not to be the case with atomic density measures of shape complementarity. The formula (2) returns a value of zero for locally flat–flat interfaces. Given two flat interfaces, one larger than the other, both would be viewed as having neutral affinity with regard to FADE's measure. FADE only returns a negative (favorable) score in the case where there are matches between protrusions and crevices. For this reason, FADE's measure cannot precisely correlate with surface area measures, although the total shape complementarity score is somewhat dependent on the size of the contact surface.

In this work, we are primarily concerned with the FADE average complementarity and local complementarity values. The FADE average complementarity is independent of the surface size, and it does not correlate at all with Lennard-Jones interactions or buried surface area. It likely has some relation to complementarity measures that look at the average "width" of the gap between molecules at their interface, in the sense that a large gap will lead to an unfavorable score with respect to both measures. Again, however, there are differences. The average width measure cannot distinguish between flat and curved interfaces, whereas FADE will view a curved interface more favorably.

FADE provides detailed local information on the geometric match between a target and ligand. It bears some similarity to a complementarity measure, developed in the Wolfson-Nussinov group, that deduces a match between protrusions and crevices based on surface curvature [9]. This technique uses a method for defining curvature across molecular surfaces that is due to Connolly [10]. Critical points of the curvature are aligned along surface normals to measure geometric complementarity. The FADE and Wolfson-Nussinov measures are similar in that they find the same types of interfaces favorable, namely those with lots of knobs-to-holes matches. One key difference between the methods is that the Wolfson-Nussinov measure is restricted to discrete points on molecular surfaces while FADE's measure is defined continuously throughout the entire 3D interface region. This is useful, as it defines a shape complementarity score at every point in the docking interface.

Finally, it is worth mentioning the differences between shape complementarity and overall binding affinity. Binding affinity is dependent both on electrostatics and shape fit, and for this reason it does not correlate precisely with any of the shape complementarity measures. However, shape features clearly relate to binding specificity. In particular, a binding site's unique geometry determines what can bind to it, and shape complementarity along with electrostatics dictates what will bind most tightly.

SURVEYING SHAPE COMPLEMENTARITY

This section presents FADE shape complementarity results for a survey of 184 protein–protein and protein–DNA interfaces. In addition, we will detail the 10 most complementary and 10 least complementary interfaces found in these studies. Some of the interfaces are defined by contacts between subunits of multidomain proteins, but most describe interactions between distinct molecules.

Table I lists the PDB codes for files we have used in the survey, along with the names and chain identifiers of the molecules in the crystal structures. From these, a set of 184 intermolecular and intramolecular domains have been analyzed. Table II gives the individual results for the number of interface points found for each interface, the total FADE shape complementarity score and the average complementarity score. Histograms of these results (Fig. 2) suggest that the size of the macromolecular interfaces varies widely across the systems studied. The total shape complementarity scores are thus distributed over a large range, as these scores depend on the number of interface points. The average shape complementarity scores, which are normalized by the number of interface points, fell within a relatively narrow range between -0.2 and -0.05. This consistency suggests that FADE is a good measure of shape complementarity in macromolecular systems.

The interdomain contacts in Feline Immunodeficiency Virus (2FIV.pdb) were found to be most complementary, with the ligand-bound form being slightly more so. Other systems having strong complementarity included three Trypsin and Trypsin Inhibitor systems (1BRC.pdb, 1AVW.pdb and 1TAW.pdb); PIT-1 and DNA (1AU7.pdb); a Tyrosine Kinase and Phosphopeptide (1LCK.pdb); Subtilisin BPN′ Prosegment and Subtilisin BPN′ (1SBP.pdb); Protein Kinase A and Protein Kinase Inhibitor (1CDK.pdb); and Triacylglycerol Acyl-Hydrolase and Colipase (1ETH.pdb). Many of these systems are known to be highly dependent on shape complementarity at their interfaces, and so it is not surprising that they would return the best FADE complementarity scores.

Four systems returned non-negative shape complementarity scores, which indicates a lack of geometric match. Two of these were interdomain

TABLE I Shown are the PDB ID codes for each system used in our shape complementarity studies. For each system, the molecule name and chain identifiers are also given

PDB ID	*Molecule 1 Name*	*Chain(s)*	*Molecule 2 Name*	*Chain(s)*
1A0O	CHEY	ACEG	CHEA	BDFH
1AB9	GAMMA-CHYMOTRYPSIN	AC	PENTAPEPTIDE	BD
1ACB	ALPHA-CHYMOTRYPSIN	E	EGLIN C	I
1AFV	HUMAN IMMUNODEFICIENCY VIRUS CAPSID	AB	ANTIBODY FAB25.3	HKLM
1AGR	GUANINE NUCLEOTIDE-BINDING PROTEIN G(I)	AD	RGS4	EH
1AHW	IMMUNOGLOBULIN FAB 5G9	ABDE	TISSUE FACTOR	CF
1AN2	MAX PROTEIN	AC	DNA	BD
1AN4	USF	AB	DNA	CD
1AR1	CYTOCHROME C OXIDASE	AB	ANTIBODY FV FRAGMENT	CD
1ATN	DEOXYRIBONUCLEASE I	D	ACTIN	A
1AU7	PIT-1	AB	DNA	CD
1AUT	ACTIVATED PROTEIN C	CL	D-PHE-PRO-MAI	P
1AVW	TRYPSIN	A	TRYPSIN INHIBITOR	B
1AVZ	NEGATIVE FACTOR	AB	FYN TYROSINE KINASE	C
1AXI	GROWTH HORMONE	A	GROWTH HORMONE RECEPTOR	B
1BCR	SERINE CARBOXY-PEPTIDASE II	AB	ARGININE	D
1BND	BRAIN DERIVED NEUROTROPHIC FACTOR	A	NEUROTROPHIN 3	B
1BQL	HYHEL-5 FAB	LH	BOBWHITE QUAIL LYSOZYME	Y
1BRC	TRYPSIN VARIANT	E	AMYLOID BETA-PROTEIN PRECURSOR INHIBITOR DOMAIN	I
1BRS	BARNASE	ABC	BARSTAR MUTANT	DEF
1BVK	HULYS11	ABDE	LYSOZYME	CF
1CA0	BOVINE CHYMOTRYPSIN	ABCFGH	PROTEASE INHIBITOR	DI
1CBW	BOVINE CHYMOTRYPSIN	ABCFGH	BPTI	DI
1CDK	CAMP-DEPENDENT PROTEIN KINASE	AB	PROTEIN KINASE INHIBITOR	IJ
1CGI	ALPHA-CHYMOTRYPSINOGEN	E	PANCREATIC SECRETORY TRYPSIN INHIBITOR	I
1CHO	ALPHA-CHYMOTRYPSIN	E	OVOMUCOID THIRD DOMAIN	I
1CSE	SUBTILISIN CARLSBERG	E	EGLIN C	I
1CWE	P56LCK TYROSINE KINASE	AC	PHOSPHONOPEPTIDE	BD
1DFJ	RIBONUCLEASE A	E	RIBONUCLEASE INHIBITOR	I
1DKG	NUCLEOTIDE EXCHANGE FACTOR GRPE	AB	MOLECULAR CHAPERONE DNAK	D
1DQJ	ANTI-LYSOZYME ANTIBODY	AB	LYSOZYME	C
1DVF	FV D1.3	AB	FV E5.2	CD
1EBD	DIHYDROLIPOAMIDE DEHYDROGENASE	AB	DIHYDROLIPOAMIDE ACETYLTRANSFERASE	C
1EFN	FYN TYROSINE KINASE	AC	HIV-1 NEF PROTEIN	BD
1EFU	ELONGATION FACTOR TU	AC	ELONGATION FACTOR TS	BD
1EO8	HEMAGGLUTININ (HA1/HA2 CHAINs)	AB	ANTIBODY (LIGHT/HEAVY CHAINS)	LH
1ETH	TRIACYLGLYCEROL ACYL-HYDROLASE	AC	COLIPASE	BD
1FBI	FAB FRAGMENT	LHPQ	LYSOZYME	XY
1FIN	CYCLIN-DEPENDENT KINASE 2	AC	CYCLIN A	BD
1FJL	PAIRED PROTEIN	ABC	DNA	DEF
1FLE	ELASTASE	E	ELAFIN	I
1FSS	ACETYLCHOLINESTERASE	A	FASCICULIN II	B
1GDT	GAMMA-DELTA RESOLVASE	AB	SITE I OF RES DNA	CDEF
1GLA	GLYCEROL KINASE	G	GLUCOSE-SPECIFIC FACTOR III	F
1GUA	RAP1A	A	C-RAF1	B
1HCQ	HUMAN/CHICKEN ESTROGEN RECEPTOR	ABEF	DNA	CDGH
1HJA	ALPHA-CHYMOTRYPSIN	ABC	OVOMUCOID INHIBITOR	I
1HTT	HISTIDYL-TRNA SYNTHETASE	ABCD	HISTIDYL-ADENYLATE	EFGH
1HWG	GROWTH HORMONE	A	GROWTH HORMONE	BC
1IAI	IDIOTYPIC FAB 730.1.4 (IGG1) OF VIRUS NEUTRALIZING ANTIBODY	LH	ANTI-IDIOTYPIC	MI
1IGC	IGG1 FAB FRAGMENT	LH	PROTEIN G	A
1IHF	INTEGRATION HOST FACTOR	AB	DNA	CDE
1JCK	14.3. D T CELL ANTIGEN RECEPTOR	AC	STAPHYLOCOCCAL ENTEROTOXIN C3	BD
1JHL	FV FRAGMENT	LH	LYSOZYME	A
1JST	CYCLIN-DEPENDENT KINASE-2	AC	CYCLIN A	BD
1JXP	NS3 SERINE PROTEASE	AB	NS4A	CD
1KIP	MONOCLONAL ANTIBODY D1.3	AB	LYSOZYME	C
1LCK	P56LCK TYROSINE KINASE	A	TAIL PHOSPHOPEPTIDE	B
1LMW	UROKINASE-TYPE PLASMINOGEN ACTIVATOR	ABCD	GLU-GLY-ARG CHLOROMETHYL KETONE	IJ
1MAH	ACETYLCHOLINESTERASE	A	FASCICULIN 2	F
1MDA	METHYLAMINE DEHYDROGENASE	LHJM	AMICYANIN	AB
1MEL	VH SINGLE-DOMAIN ANTIBODY	AB	LYSOZYME	LM

TABLE I – *continued*

PDB ID	*Molecule 1 Name*	*Chain(s)*	*Molecule 2 Name*	*Chain(s)*
1MEY	DNA	ABDE	CONSENSUS ZINC FINGER PROTEIN	CFG
1MHC	MHC CLASS I ANTIGEN H2-M3	ABDE	NONAPEPTIDE FROM RAT	CF
1MLC	MONOCLONAL ANTIBODY FAB	ABCD	LYSOZYME	EF
1MPA	MN12H2 IGG2A-KAPPA	LH	PORA P1.16 PEPTIDE	P
1N2C	NITROGENASE MOLYBDENUM-IRON PROTEIN	ABCD	NITROGENASE IRON PROTEIN	EFGH
1NCA	N9 NEURAMINIDASE-NC41	N	FAB	LH
1NFD	N15 ALPHA-BETA T-CELL RECEPTOR	ABCD	H57 FABEF	GH
1NFK	NUCLEAR FACTOR KAPPA-B	AB	"KB SITE, DNA (5′-D(TGAGAATTCCC)-3′)"	CD
1NMA	N9 NEURAMINIDASE	N	FAB NC10	LH
1NMB	N9 NEURAMINIDASE	N	FAB NC10	LH
1NPO	NEUROPHYSIN II	AC	OXYTOCIN	BD
1NSN	"IGG FAB (IGG1, KAPPA)"	LH	STAPHYLOCOCCAL NUCLEASE	S
1OSP	FAB 184.1	LH	OUTER SURFACE PROTEIN A	O
1PAU	APOPAIN	AB	ACE-ASP-GLU-VAL-ASP-CHO	C
1PFX	FACTOR IXA	CL	D-PHE-PRO-ARG	I
1PPE	TRYPSIN	E	TRYPSIN INHIBITOR (CMTI-I)	I
1QFU	HEMAGGLUTININ (HA1/HA2 CHAINS)	AB	IMMUNOGLOBULIN IGG1-KAPPA ANT	LH
1RLB	TRANSTHYRETIN	ABCD	RETINOL BINDING PROTEIN	EF
1RMH	CYCLOPHILIN A	AB	AAPF PEPTIDE SUBSTRATE	CD
1RUN	DNA	CDEF	CATABOLITE GENE ACTIVATOR	AB
1RVF	HUMAN RHINOVIRUS 14 COAT PROTEIN	1234	FAB 17-IA	LH
1SGP	STREPTOMYCES GRISEUS PROTEINASE B	E	TURKEY OVOMUCOID INHIBITOR	I
1SPB	SUBTILISIN BPN′ PROSEGMENT	P	SUBTILISIN BPN′	S
1SRS	SERUM RESPONSE FACTOR	AB	SRE SPECIFIC DNA	WC
1STF	PAPAIN	E	STEFIN B (CYSTATIN B) MUTANT	I
1TAB	TRYPSIN	E	BOWMAN-*BIRK INHIBITOR	I
1TAW	TRYPSIN	A	PROTEASE INHIBITOR DOMAINOF ALZHEIMER'S AMYLOID BETA-PROTEIN	B
1TBG	TRANSDUCIN	ABCD	TRANSDUCIN	EFGH
1TBQ	THROMBIN	LHJK	RHODNIIN	RS
1TCO	SERINE/THREONINE PHOSPHATASE B2	AB	FK506-BINDING PROTEIN	C
1TFX	TRYPSIN	AB	TISSUE FACTOR PATHWAY INHIBITOR	CD
1TGS	TRYPSINOGEN	Z	PANCREATIC SECRETORY TRYPSIN INHIBITOR	I
1TSR	P53 TUMOR SUPPRESSOR	ABC	DNA	EF
1UDI	URACIL-DNA GLYCOSYLASE	E	URACIL-DNA GLYCOSYLASE INHIBI	I
1UGH	URACIL-DNA GLYCOSYLASE	E	URACIL-DNA GLYCOSYLASE INHIBI	I
1URN	U1A SPLICEOSOMAL PROTEIN	ABC	RNA 21MER HAIRPIN	PQR
1VLT	ASPARTATE RECEPTOR	AB	ASPARTATE	CD
1WEJ	E8 ANTIBODY	LH	CYTOCHROME C	F
1WQ1	H-RAS	R	P120GAP	G
1XBR	T PROTEIN	AB	DNA	CD
2BTF	BETA-ACTIN	A	PROFILIN	P
2FIV	FELINE IMMUNODEFICIENCY VIRUS PROTEASE	AB	ACE-NAL-VAL-STA-GLU-NAM	IJ
2JEL	JEL42 FAB FRAGMENT	LH	HISTIDINE-CONTAINING PROTEIN	P
2KAI	KALLIKREIN A	AB	BOVINE PANCREATIC TRYPSIN INHIBITOR	I
2PCC	YEAST CYTOCHROME C PEROXIDASE	AC	YEAST ISO-1-CYTOCHROME C	BD
2PTC	BETA-TRYPSIN	E	PANCREATIC TRYPSIN INHIBITOR	I
2SIC	SUBTILISIN /BPN	E	STREPTOMYCES SUBTILISIN INHIBITOR	I
2SNI	SUBTILISIN NOVO	E	CHYMOTRYPSIN INHIBITOR	I
2TEC	THERMITASE	E	EGLIN C	I
2VIR	IMMUNOGLOBULIN (IGG1LAMBDA)	AB	HEMAGGLUTININ	C
4HTC	ALPHA-THROMBIN	LH	RECOMBINANT HIRUDIN	I

contacts in USF and Max Protein, proteins that interact with DNA (1AN2.pdb and 1AN4.pdb). The monomers of USF and Max Protein have long, skinny regions that do not form a tight interface, so FADE did not view them as being complementary. Tyrosine Kinase and a Negative Factor (1AVZ.pdb) and Transthyretin and Retinol Binding Protein (1RLB.pdb) also returned non-negative complementarity scores. These systems were found to have gaps in their interfaces. The dimerization of Chymotrypsin-BPTI (1CBW.pdb) and FYN Tyrosine Kinase with HIV-1 NEF Protein (1EFN.pdb) was not tight enough to result in strong complementarity. Additionally, several systems were found to have

TABLE II The results of running FADE on 184 protein-protein and protein-DNA interfaces are displayed

System	*Chain*	*Chain*	*Points*	S_c *total*	S_c *avg*	*System*	*Chain*	*Chain*	*Points*	S_c *total*	S_c *avg*
1A0O	A	B	882	− 60.086	− 0.068	1JCK	A	C	877	− 63.536	− 0.072
1AB9	AB	CD	5213	− 712.606	− 0.137	1JCK	B	C	508	− 54.081	− 0.106
1ACB	E	I	1251	− 186.432	− 0.149	1JHL	LH	A	1012	− 61.826	− 0.061
1AFV	A	HL	1223	− 129.066	− 0.106	1JST	AB	CD	1474	− 90.191	− 0.061
1AGR	A	E	1287	− 98.256	− 0.076	1JST	A	B	2733	− 250.379	− 0.092
1AHW	AB	C	1701	− 118.370	− 0.070	1JXP	AC	BD	1250	− 86.492	− 0.069
1AN2	AC	BD	1864	− 105.871	− 0.057	1KIP	AB	C	1066	− 58.222	− 0.055
1AN2	A	C	1777	54.241	0.031	1KIP	A	B	1433	− 104.912	− 0.073
1AN2	C	D	714	− 63.107	− 0.088	1KIP	A	C	492	− 4.547	− 0.009
1AN4	AB	CD	2202	− 142.666	− 0.065	1KIP	B	C	617	− 50.397	− 0.082
1AN4	A	B	1008	16.447	0.016	1LCK	A	B	889	− 175.028	− 0.197
1AN4	C	D	1941	− 115.184	− 0.059	1LMW	ABI	CDJ	723	− 40.342	− 0.056
1AR1	AB	CD	1144	− 56.039	− 0.049	1MAH	A	F	1707	− 244.588	− 0.143
1AR1	A	B	5908	− 387.549	− 0.066	1MDA	LH	A	669	− 24.877	− 0.037
1AR1	C	D	1621	− 160.206	− 0.099	1MDA	LH	JM	4664	− 452.349	− 0.097
1ATN	D	A	1463	− 195.324	− 0.134	1MEL	A	L	1377	− 207.840	− 0.151
1AU7	AB	CD	4706	− 373.353	− 0.079	1MEL	L	MI	638	− 49.879	− 0.078
1AU7	A	B	910	− 181.194	− 0.199	1MEY	AB	C	2066	− 117.959	− 0.057
1AU7	C	D	2060	− 123.361	− 0.060	1MEY	AB	CG	2125	− 128.195	− 0.060
1AUT	C	L	1508	− 208.258	− 0.138	1MHC	AB	DE	862	− 80.798	− 0.094
1AVW	A	B	1551	− 293.416	− 0.189	1MHC	A	B	2158	− 177.269	− 0.082
1AVZ	AB	C	1002	− 88.192	− 0.088	1MLC	ABE	CDF	591	− 14.700	− 0.025
1AVZ	A	B	676	1.159	0.002	1MLC	AB	E	1207	− 101.169	− 0.084
1AVZ	B	C	1002	− 88.735	− 0.089	1MPA	LH	P	1085	− 162.284	− 0.150
1AXI	A	B	1962	− 136.511	− 0.070	1N2C	AB	EF	2905	− 173.658	− 0.060
1BCR	A	B	7082	− 694.929	− 0.098	1N2C	A	B	7919	− 691.211	− 0.087
1BND	A	B	2254	− 135.043	− 0.060	1N2C	E	F	4215	− 338.460	− 0.080
1BQL	LH	Y	1312	− 105.591	− 0.080	1NCA	N	LH	1619	− 129.854	− 0.080
1BQL	L	H	2592	− 107.792	− 0.042	1NFD	AB	EF	1507	− 239.492	− 0.159
1BRC	E	I	983	− 204.292	− 0.208	1NFK	AC	BD	2695	− 174.219	− 0.065
1BRS	A	D	1360	− 139.414	− 0.103	1NFK	A	C	1062	− 101.857	− 0.096
1BVK	ABDE	CF	2184	− 152.808	− 0.070	1NMA	N	LH	1276	− 59.790	− 0.047
1BVK	A	B	1257	− 103.018	− 0.082	1NMB	N	LH	1200	− 70.866	− 0.059
1BVK	B	C	649	− 52.680	− 0.081	1NPO	A	B	783	− 139.866	− 0.179
1CA0	ABC	D	1162	− 183.163	− 0.158	1NPO	A	C	840	− 51.705	− 0.062
1CA0	B	C	5143	− 615.621	− 0.120	1NSN	LH	S	1378	− 73.363	− 0.053
1CBW	ABC	D	1266	− 207.292	− 0.164	1OSP	LH	O	1209	− 72.472	− 0.060
1CBW	FGH	I	1145	− 171.323	− 0.150	1PAU	A	BD	3535	− 264.566	− 0.075
1CBW	ABCD	FGHI	1252	− 16.229	− 0.013	1PFX	C	LI	1929	− 336.534	− 0.174
1CDK	A	I	1768	− 341.876	− 0.193	1PPE	E	I	1442	− 246.693	− 0.171
1CGI	E	I	1797	− 301.258	− 0.168	1QFU	AB	LH	1321	− 58.625	− 0.044
1CHO	E	I	1195	− 165.305	− 0.138	1RLB	ABCD	E	1165	− 97.163	− 0.083
1CSE	E	I	1310	− 209.990	− 0.160	1RLB	AB	CD	1672	29.834	0.018
1CWD	L	P	650	− 100.483	− 0.155	1RLB	A	B	1573	− 166.750	− 0.106
1CWE	AB	CD	449	− 36.609	− 0.082	1RMH	AC	BD	584	− 38.898	− 0.067
1DFJ	E	I	1885	− 138.547	− 0.073	1RUN	CDEF	AB	2175	− 242.264	− 0.111
1DKG	AB	D	1457	− 53.579	− 0.037	1RVF	1234	LH	1557	− 172.683	− 0.111
1DKG	A	B	3241	− 111.522	− 0.034	1RVF	L	H	1263	− 90.613	− 0.072
1DQJ	AB	C	1570	− 147.817	− 0.094	1RVF	1	3	6197	− 777.478	− 0.125
1DQJ	A	B	2849	− 153.402	− 0.054	1RVF	2	3	2451	− 116.352	− 0.047
1DVF	AB	CD	1437	− 33.029	− 0.023	1SGP	E	I	987	− 101.573	− 0.103
1DVF	A	B	1370	− 114.794	− 0.084	1SPB	P	S	2143	− 417.844	− 0.195
1EBD	AB	C	1046	− 92.076	− 0.088	1SRS	AB	WC	3424	− 475.338	− 0.139
1EBD	A	B	4882	− 424.146	− 0.087	1SRS	A	B	2831	− 164.773	− 0.058
1EFN	AB	CD	582	− 5.187	− 0.009	1STF	E	I	1345	− 228.056	− 0.170
1EFN	A	B	1098	− 91.813	− 0.084	1TAB	E	I	1113	− 172.952	− 0.155
1EFU	AB	CD	2591	− 76.847	− 0.030	1TAW	A	B	1152	− 215.433	− 0.187
1EFU	A	B	2908	− 344.612	− 0.119	1TBG	ABEF	CDGH	2165	− 65.282	− 0.030
1EO8	AB	LH	1204	− 111.248	− 0.092	1TBG	AE	BF	1395	− 104.186	− 0.075
1EO8	A	B	4789	− 795.312	− 0.166	1TBG	A	E	3747	− 521.337	− 0.139
1EO8	L	H	2779	− 104.482	− 0.038	1TBQ	L	H	3235	− 408.850	− 0.126
1ETH	AB	CD	1231	− 232.054	− 0.189	1TBQ	H	R	2920	− 274.676	− 0.094
1ETH	A	B	1317	− 127.873	− 0.097	1TCO	AB	C	1384	− 111.675	− 0.081
1FBI	LH	X	1309	− 90.568	− 0.069	1TCO	A	B	2950	− 285.529	− 0.097
1FBI	L	H	2928	− 176.580	− 0.060	1TFX	A	C	1158	− 208.454	− 0.180
1FIN	AB	CD	1242	− 81.862	− 0.066	1TGS	Z	I	1521	− 260.851	− 0.171
1FIN	A	B	2819	− 258.163	− 0.092	1TSR	ABC	EF	1446	− 149.826	− 0.104
1FJL	ABC	DEF	3956	− 361.484	− 0.091	1TSR	A	B	1186	− 108.894	− 0.092
1FLE	E	I	1411	− 216.899	− 0.154	1UDI	E	I	1498	− 131.317	− 0.088
1FSS	A	B	1628	− 251.065	− 0.154	1UGH	E	I	1680	− 95.083	− 0.057
1GDT	AB	CDEF	4747	− 646.398	− 0.136	1URN	AP	BCQR	1604	− 94.251	− 0.059
1GDT	A	B	1532	− 94.659	− 0.062	1URN	AP	BQ	732	− 44.639	− 0.061

TABLE II – *continued*

System	*Chain*	*Chain*	*Points*	S_c *total*	S_c *avg*	*System*	*Chain*	*Chain*	*Points*	S_c *total*	S_c *avg*
1GDT	C	F	1315	− 94.499	− 0.072	1URN	A	P	1648	− 214.473	− 0.130
1GDT	D	E	1207	− 78.878	− 0.065	1VLT	AD	BC	1529	− 26.196	− 0.017
1GLA	G	F	933	− 22.842	− 0.024	1WEJ	LH	F	1113	− 74.177	− 0.067
1GUA	A	B	1114	− 87.514	− 0.079	1WQ1	R	G	2120	− 182.231	− 0.086
1HCQ	A	B	660	− 52.545	− 0.080	1XBR	AB	CD	3318	− 411.125	− 0.124
1HCQ	AB	C	1036	− 119.535	− 0.115	1XBR	A	B	301	− 17.718	− 0.059
1HJA	ABC	I	1368	− 248.402	− 0.182	2BTF	A	P	1633	− 149.237	− 0.091
1HJA	B	C	5047	− 591.337	− 0.117	2FIV	AB	IJ	1870	− 432.930	− 0.232
1HTT	A	B	6111	− 620.748	− 0.102	2FIV	A	B	3169	− 667.148	− 0.211
1HWG	A	BC	3338	− 223.620	− 0.067	2JEL	LH	P	1297	− 98.928	− 0.076
1HWG	A	B	2002	− 125.530	− 0.063	2KAI	AB	I	1250	− 196.611	− 0.157
1HWG	A	C	1336	− 116.425	− 0.087	2KAI	A	B	5337	− 870.145	− 0.163
1HWG	B	C	806	− 95.869	− 0.119	2PCC	A	B	821	− 16.008	− 0.019
1IAI	LH	MI	1488	− 81.015	− 0.054	2PCC	AB	CD	962	− 28.074	− 0.029
1IGC	LH	A	1183	− 98.563	− 0.083	2PTC	E	I	1243	− 206.419	− 0.166
1IHF	AB	CDE	3815	− 504.427	− 0.132	2SIC	E	I	1437	− 228.502	− 0.159
1IHF	A	B	3824	− 191.961	− 0.050	2SNI	E	I	1352	− 229.771	− 0.170
1IHF	C	D	1389	− 72.480	− 0.052	2TEC	E	I	1307	− 237.485	− 0.182
1JCK	AC	BD	2852	− 164.086	− 0.058	2VIR	AB	C	1093	− 101.751	− 0.093
1JCK	A	B	885	− 26.556	− 0.030	4HTC	LH	I	2612	− 327.139	− 0.125

[a] For each interface, the PDB ID code for the system is given, along with the specific chains used to define the interface, the number of interface points, the total complementarity score and the average complementarity score. Histograms displaying the distribution of scores are given in Fig. 2.

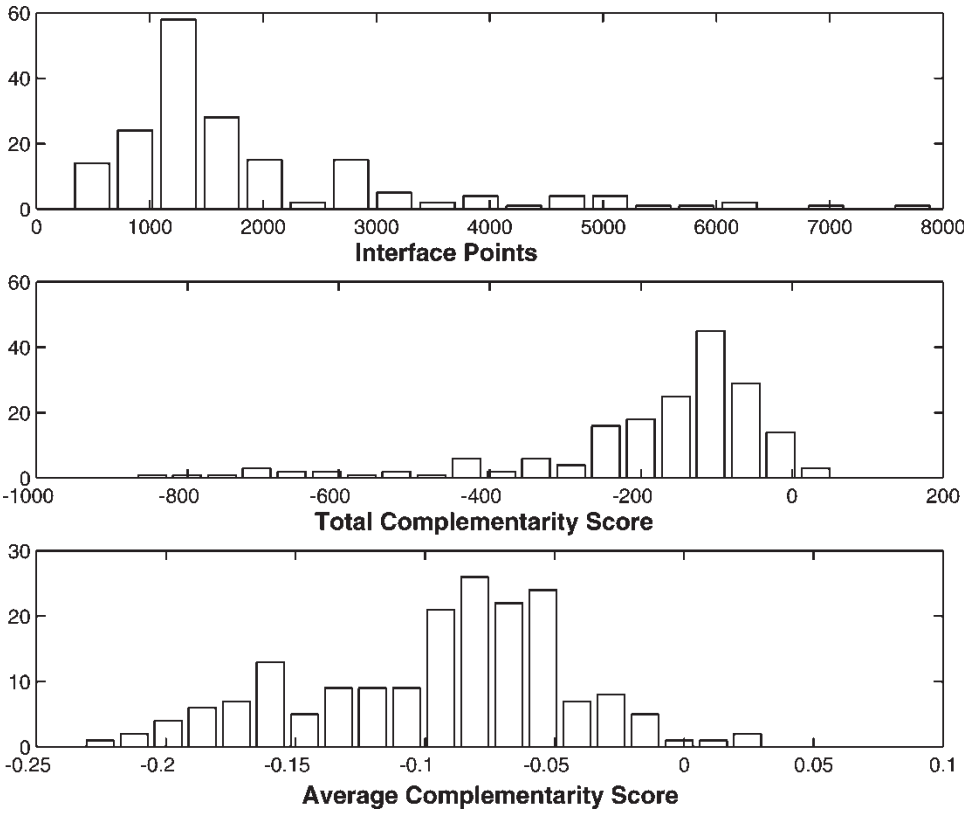

FIGURE 2 The histograms give statistical data on the FADE results for 184 protein–protein and protein–DNA systems. The first histogram is based on the number of interface points found for each system, and we see that the sizes of the interfaces vary widely. Similarly, the total complementarity scores, which depend on the size of the interface, are also distributed across a range of values. However, the average complementarity scores fell within a relatively narrow range, with the vast majority between −0.2 and −0.05. Systems at the lower end of the spectrum are ones in which shape complementarity is important to the interaction, whereas those at the upper end have a lack of shape complementarity that is apparent up on visual inspection.

large voids in their interfaces that were viewed unfavorably by FADE's shape complementarity measures. Two systems of this type were Aspartate Receptor and Aspartate (1VLT.pdb); and a monoclonal antibody and lysozyme (1KIP.pdb). Finally, it is interesting to note that the controversial yeast cytochrome C Peroxidase and cytochrome *c* structure (2PCC.pdb) was found to have very weak complementarity. FADE's results for a related structure (2PCB.pdb) not used in the survey returned an average complementarity of -0.067, which lies within the typical range.

In general, the FADE results returned good scores in cases where significant shape complementarity is expected, such as for the Trypsin and Trypsin Inhibitor systems. A lack of shape complementarity could be visually verified for those systems in which the FADE score indicated weak complementarity or mismatch. The authors were pleasantly surprised to see that FADE performed well on interfaces between proteins and DNA. The radial methods employed by FADE do not deduce directional shape information, and it was unclear how they would perform on the saddle-like interfaces often found between proteins and DNA. However, the average shape complementarity scores were distributed throughout the same range as those of the protein–protein interfaces, and visual analysis of the results for the saddle interface between DNA and a zinc finger domain (1MEY.pdb) indicated that FADE's shape measures highlighted the most complementary regions of the interface.

It is worth noting some limitations to the use of FADE in determining shape complementarity. A radial counting scheme does not make sense when the number of atoms in the ligand is small, and we have found by trial and error that FADE's methods require a ligand with at least 100 non-hydrogen atoms in order to produce reliable shape complementarity results. Similarly, the results for small interfaces, even between large proteins, may not be valid. FADE scores for interfaces that produce less than 500 interface points should be regarded with caution.

COMPLEMENTARITY HOT SPOTS

Here, we will demonstrate how complementarity markers in docking interfaces relate to experimental results. The Fasciculin-Acetylcholinesterase and Barnase–Barstar systems were chosen for study because shape complementarity is known to play an important role in the interaction and because there are abundant experimental results detailing the effects of mutation. It is hoped that shape complementarity analysis of existing crystal structures can guide experimental analysis, and these examples provide evidence of FADE's value in this regard.

The Fasciculin-Acetylcholinesterase Complex

Acetylcholinesterase is an enzyme that hydrolyzes the neurotransmitter acetylcholine into its components, acetate and choline. The venom of the mamba snake contains the "three fingered" toxin Fasciculin, which inhibits Acetylcholinesterase by blocking its active site gorge. When the action of Acetylcholinesterase is blocked, the nervous system becomes overloaded with acetylcholine, and death occurs.

Complementarity analysis by FADE of the interface between Fasciculin and Acetylcholinesterase (1MAH.pdb) reveals five regions of strong shape complementarity (Fig. 3). Mutations of Fasciculin in four of these regions, specifically ARG^{11}, THR^{8-9}, ARG^{27}–PRO^{31} and MET^{33}, are known to significantly affect its ability to inhibit Acetylcholinesterase [11]. The effects were varied, with mutations in ARG^{27}, PRO^{30} and PRO^{31} resulting in a two orders of magnitude loss of inhibitory activity and mutations in MET^{33} producing a one order of magnitude loss. These residues all lie on the second of Fasciculin's three fingers. The mutation of ARG^{11} and THR^{8-9}, in the first finger of Fasciculin, resulted in an activity increase. This may be due to the fact that these residues, particularly ARG^{11}, must change conformation in order to achieve an induced fit with Acetylcholinesterase.

The Barnase–Barstar Complex

Barnase is an extracellular ribonuclease, and Barstar is an inhibitor of Barnase that is roughly comparable in size. Barnase and Barstar are known to form a very tight complex in which shape complementarity plays an important role. Our analysis of the Barnase–Barstar interface uncovered numerous complementarity hot spots. Most, although not all, can be correlated with published experimental data. It is well-known that mutations in Barnase HIS^{102} affect Barstar's ability to bind to it [12–16], and this is also true of Barnase ARG^{59} [14–16]. Both these residues were found within shape complementarity hot spots (Fig. 4). For ARG^{59}, the CZ atom and NH1 group appeared key to the shape complementarity. For HIS^{102}, there are two sets of complementarity markers, one of which lies near the CO_2^- group and the other near the CE1 and NE2 atoms of the imidazole ring.

Mutagenesis studies indicate that TYR^{29} and ASP^{35} are two of three Barstar residues that contribute most significantly to the tight complex between Barnase and Barstar [12,13,15,17]. The OD1, OD2 and CG atoms of ASP^{35} play an important role

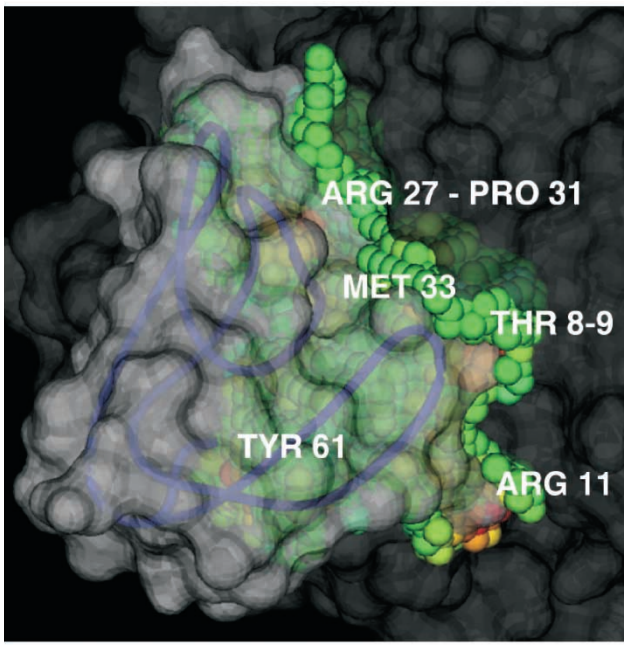

FIGURE 3 Acetylcholinesterase is shown in dark grey, and Fasciculin in lighter gray. Interface points are colored by local shape complementarity measures computed by FADE, with red highlighting complementarity "hot spots" and blue suggesting a mismatch. Beneath the translucent surface can be seen the Fasciculin backbone and several regions of strong complementarity. It is most evident near THR^{8-9} and ARG^{11} at the tip of the first of Fasciculin's three "fingers". Regions of complementarity are also seen near the tip of the second finger (from ARG^{27} to PRO^{31} and MET^{33}) and at the terminal residue, TYR^{61}. Mutations in four of these regions are known to affect the inhibitory activity of fasciculin.

in the interaction [17], and in TYR^{29}, the CD1, CE1 and CE2 atoms are regarded as critical. For ASP^{35}, one can find orange and red complementarity markers directly adjacent to the OD1, OD2, CB and CG atoms; and with TYR^{29}, it is clear that the CE2 atom lies at the heart of the complementarity hot spot (Fig. 4).

CONCLUSIONS

Here, we have demonstrated the value of the Fast Atomic Density Evaluator in determining shape complementarity for protein–protein and protein–DNA complexes. Our results indicate that the average complementarity scores produced by FADE lie within a narrow range for a collection of 184 protein–protein and protein–DNA systems with interfaces of widely varying size and shape. Interfaces found to be most complementary by FADE are ones in which the geometric match is of known importance to the interaction. Finally, our detailed analysis of the Acetylcholinesterase-Fasciculin and Barnase–Barstar systems indicates that shape complementarity hot spots are well-correlated with residues in which mutation is known to have a measurable impact on the binding affnity. For the Barnase–Barstar system, FADE's results highlighted not only the critical residues but also specific atoms that are known to play an important role in binding.

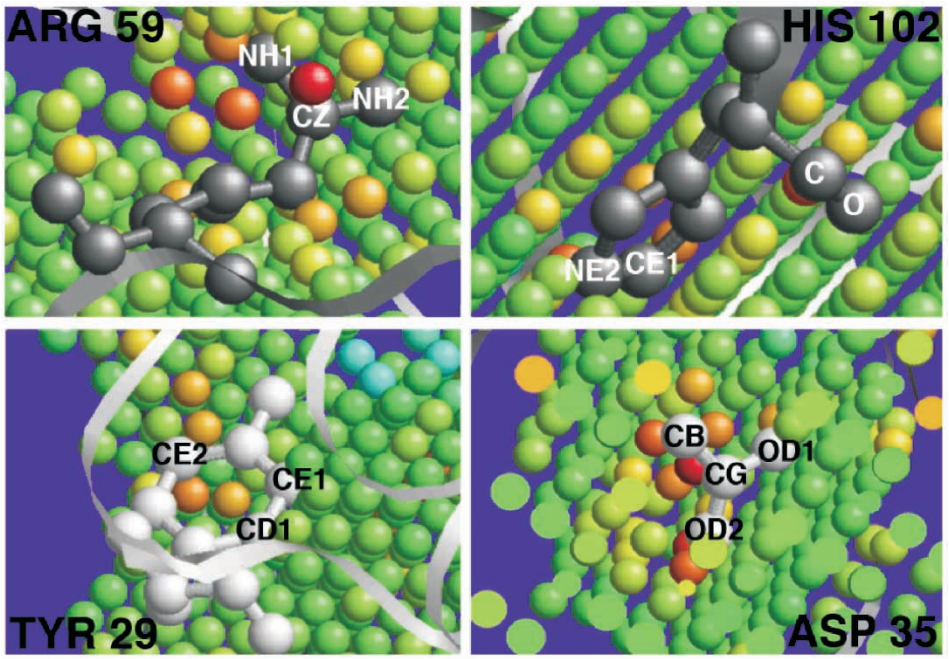

FIGURE 4 The complementarity markers in the Barnase-Barstar interface can be correlated with experimental mutagenesis data. Mutation of Barnase residues ARG^{59} and HIS^{102} are known to affect Barstar's ability to bind tightly. The TYR^{29} and ASP^{35} residues of Barstar play an important role in the interaction, and the complementarity markers highlight specific atoms that are critical to the interaction.

Acknowledgements

Figures 1 and 2 were created using Matlab, Fig. 3 was created using Molscript, and Fig. 4 was created using Rasmol. Various enhancements were made to the figures using Graphic Converter for the Macintosh. All computations were performed on a 500 MHz G4 Titanium Powerbook from Apple Computers.

References

[1] Mitchell, J.C., Kerr, R. and Ten Eyck, L.F. (2001) "Rapid atomic density methods for molecular shape characterization", *J. Mol. Graph. Model.* **19**, 324.
[2] Kuhn, L.A., Siani, M.A., Pique, M.E., Fisher, C.L., Getzoff, E.D. and Tainer, J.A. (1992) "The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures", *J. Mol. Biol.* **228**, 13.
[3] Mandell, J.G., Roberts, V.A., Pique, M.E., Kotlovyi, V., Mitchell, J.C., Nelson, E., Tsigelny, I. and Ten Eyck, L.F. (2001) "Protein docking using continuum electrostatics and geometric fit", *Protein Eng.* **14**, 105.
[4] Gabb, H.A., Jackson, R.M. and Sternberg, M.J. (1997) "Modeling protein docking using shape complementarity, electrostatics and biochemical information", *J. Mol. Biol.* **272**, 106.
[5] Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C. and Vakser, I.A. (1992) "Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques", *Proc. Natl Acad. Sci. USA* **89**, 2195.
[6] Chen, R. and Weng, Z. (2002) "Docking unbound proteins using shape complementarity, desolvation, and electrostatics", *Prot. Struct. Funct. Gen.* **47**, 281.
[7] Chen R. and Weng Z. (2002) "A novel shape complementarity function for scoring protein–protein docking", *Prot. Struct. Funct. Gen.* **51**, 397.
[8] Vakser, I.A. (1995) "Protein docking for low-resolution structures", *Protein Eng.* **8**, 371.
[9] Norel, R., Lin, S.L., Wolfson, H.J. and Nussinov, R. (1994) "Shape complementarity at protein–protein interfaces", *Biopolymers* **34**, 933.
[10] Connolly, M.L. (1986) "Measurement of protein surface shape by solid angles", *J. Mol. Graphics* **4**, 3.
[11] Marchot, P., Prowse, C.N., Kanter, J., Camp, S., Ackermann, E.J., Radić, Z., Bougis, P.E. and Taylor, P. (1997) "Expression and activity of mutants of fasciculin, a peptidic acetylcholinesterase inhibitor from mamba venom", *J. Biol. Chem.* **272**, 3502.
[12] Vaughan, C.K., Buckle, A.M. and Fersht, A.R. (1999) "Structural response to mutation at a protein–protein interface", *J. Mol. Biol.* **286**, 1487.
[13] Martin, C., Hartley, R. and Mauguen, Y. (1999) "X-ray structural analysis of compensating mutations at the Barnase–Barstar interface", *FEBS Lett.* **452**, 128.
[14] Jucovic, M. and Hartley, R.W. (1996) "Protein–protein interaction: A genetic selection for compensating mutations at the barnase-barstar interface", *Proc. Natl Acad. Sci. USA* **93**, 2343.
[15] Hartley, R.W. (1993) "Directed mutagenesis and Barnase–Barstar recognition", *Biochemistry* **32**, 5978.
[16] Schreiber, G. and Fersht, A.R. (1993) "Interaction of Barnase with its polypeptide inhibitor Barstar studied by protein engineering", *Biochemistry* **32**, 5145.
[17] Liang, S., Liu, Z., Li, W., Ni, L. and Lai, L. (2000) "Construction of binding sites in scaffold proteins", *Biopolymers* **54**, 515.