Comparison of protein–protein interactions in serine protease-inhibitor and antibody-antigen complexes: Implications for the protein docking problem

RICHARD M. JACKSON

Biomolecular Structure and Modeling Unit, Department of Biochemistry and Molecular Biology,
University College London, Gower Street, London WC1E 6BT, United Kingdom

(RECEIVED July 1, 1998; ACCEPTED November 16, 1998)

Abstract

The protein–protein interaction energy of 12 nonhomologous serine protease-inhibitor and 15 antibody-antigen complexes is calculated using a molecular mechanics formalism and dissected in terms of the main-chain vs. side-chain contribution, nonrotameric side-chain contributions, and amino acid residue type involvement in the interface interaction.

There are major differences in the interactions of the two types of protein–protein complex. Protease-inhibitor complexes interact predominantly through a main-chain–main-chain mechanism while antibody-antigen complexes interact predominantly through a side-chain–side-chain or a side-chain–main-chain mechanism. However, there is no simple correlation between the main-chain–main-chain interaction energy and the percentage of main-chain surface area buried on binding. The interaction energy is equally effected by the presence of nonrotameric side-chain conformations, which constitute ~20% of the interaction energy. The ability to reproduce the interface interaction energy of the crystal structure if original side-chain conformations are removed from the calculation is much greater in the protease-inhibitor complexes than the antibody-antigen complexes. The success of a rotameric model for protein–protein docking appears dependent on the extent of the main-chain–main-chain contribution to binding.

Analysis of (1) residue type and (2) residue pair interactions at the interface show that antibody-antigen interactions are very restricted with over 70% of the antibody energy attributable to just six residue types (Tyr > Asp > Asn > Ser > Glu > Trp) in agreement with previous studies on residue propensity. However, it is found here that 50% of the antigen energy is attributable to just four residue types (Arg = Lys > Asn > Asp). On average just 12 residue pair interactions (6%) contribute over 40% of the favorable interaction energy in the antibody-antigen complexes, with charge-charge and charge/polar-tyrosine interactions being prominent. In contrast protease inhibitors use a diverse set of residue types and residue pair interactions.

Keywords: mean field optimization; multiple copy side-chain rotamer model; protein docking; protein–protein interactions

The effectiveness of protein–protein molecular docking methods depend critically on the computational model used to represent the molecular recognition process. The aim of such models is to provide a deeper understanding of the nature and mechanism of protein recognition. Here mean field refinement of a side-chain rotamer model is used to describe the interactions at the atomic level between two widely studied classes of protein–protein complex, the serine protease-inhibitor and antibody-antigen complexes. In addition to providing information about the nature and composition of the binding interface, the use of a rotameric model for side-

chain conformational states is critically assessed. Lastly, differences in the characteristics of binding are placed in the wider context of other known protein–protein complexes.

Jackson et al. (1998) described a model that uses a molecular mechanics representation including a discrete solvation model and a conformational search procedure to optimize side-chain conformations in the protein–protein interface. The proteins are described by a multiple copy representation of side-chains modeled according to a rotamer library (Tuffery et al., 1991) on a fixed peptide main-chain. Given the initial rigid-body docking of the unbound protein molecules (Gabb et al., 1997), it was found that the predicted side-chain conformations reproduced the experimental ones for a series of serine protease-inhibitor complexes. However, the results for antibody-antigen complexes were not so encouraging, suggesting that the results represent the somewhat different bind-

Reprint requests to: Richard M. Jackson, Biomolecular Structure and Modeling Unit, Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, United Kingdom; e-mail: jackson@biochem.ucl.ac.uk.

ing mechanisms dominant in the two types of protein complex. Energetic features are, however, common to both systems. In particular, it was found that the intermolecular nonbonded interaction energy was the primary determinate for successful ranking of structures. This comprises the van der Waals and electrostatic potential energy across the protein–protein interface.

In a biological sense canonical serine protease inhibitors and antibodies represent very different classes of binding protein. However, biological interest in these systems has resulted in an extensive body of knowledge about these two classes of heterogeneous protein–protein complex. In particular, the considerable extent of three-dimensional structural information allows their study in terms of topological and physicochemical characteristics (Wilson & Stanfield, 1993; Webster et al., 1994; MacCallum et al., 1996; Jones & Thornton, 1997) as well as energetic ones (Novotny et al., 1989; Krystek et al., 1993). Similarly, the preponderance of structural information has made them the most widely studied test systems in macromolecular docking (for reviews see Cherfils & Janin, 1993; Janin, 1995; Sternberg et al., 1998). Biologically, there are few common features between the two classes of structure. Antibodies can recognize a wide spectrum of antigenic compounds that include proteins, polysaccharides, nucleic acids, and even lipids (Sela & Pecht, 1996). This enormous diversity is accommodated by a common structural motif (an immunoglobulin-like β -sandwich), in which sequence and structural diversity is largely confined to six loops, the complementarity determining regions (CDRs). These loops are modified in shape and chemical nature to match the corresponding features of the antigen epitope. In contrast the endogenous canonical serine protease inhibitors are small proteins that form tight substrate-like interactions with the serine endopeptidases. Inhibition involves a common mechanism whereby an exposed binding loop of a characteristic canonical (substrate-like) conformation interacts with the active site of the protein (Bode & Huber, 1992). In spite of a common binding loop conformation, the remaining structure is unrelated in the different inhibitor families. The inhibitors are important for blocking the activity of their target proteases in emergencies, or for receptor signaling, and have developed in an evolutionary sense to bind tightly to their target enzyme. The fundamental difference in binding affinity is reflected in the binding energetics of the antibodies and canonical inhibitors, involving up to six orders of magnitude (nanomolar as opposed to femtomolar) in the binding constants.

This paper attempts to dissect the energetics of the interface interaction in the two major classes of heterogeneous protein–protein complexes studied by molecular docking methodologies. This study analyzes the constitution of the interaction enthalpy ($\Delta E_{\text{interaction}}$) in protein–protein binding because of its importance in ranking docked solutions (Jackson et al., 1998). The overall free energy of binding can be expressed as a series of additive terms in a “master equation” (Ajay & Murko, 1995) suitable for use with force field methods:

$$\Delta G_{\text{bind}} = \Delta G_{\text{interaction}} + \Delta G_{\text{solvent}} + \Delta G_{\text{internal}} + \Delta G_{\text{motion}} \quad (1)$$

The free energy of binding is expressed in terms of change in (1) the specific protein–protein interactions, (2) the protein–solvent energy, (3) the internal free energy of the proteins themselves, and (4) the conformational degrees of freedom of side-chain/main-chain, and the molecular translational/rotational and vibrational free energies. The interaction free energy $\Delta G_{\text{interaction}}$ is the change in free

energy due to specific interactions (van der Waals and electrostatic) on binding and is almost entirely enthalpic ($\Delta G_{\text{interaction}} \approx \Delta E_{\text{interaction}} = \Delta E_{\text{vdW}} + \Delta E_{\text{electrostatic}}$). Unlike the other energy components involved in an empirical estimation of binding free energy, the interaction energy can be calculated given the atomic coordinates of the complex and a specified energy function. In contrast, the change in solvation (or hydration) free energy, the change in internal energy and the loss in conformational entropy (loss of side-chain/main-chain degrees of freedom) on binding are more difficult to estimate using empirical models (Jackson & Sternberg, 1995; Weng et al., 1996; Novotny et al., 1997). This is due to the fact that in many cases (i.e., where a rigid-body approximation is not valid), unbound (reference) states of the molecules need to exist in addition to the complex to calculate realistic estimates. However, in the case of all but a few complexes, the unbound structures are not available.

This paper is based on the experimentally determined crystal complexes, and subsequent analysis concentrates on the relative contributions of different interaction energy components to molecular recognition. This allows the comparison of different protein–protein complexes albeit at a qualitative rather than quantitative level. A direct relationship between the interaction energy and the macroscopic state functions (ΔC , ΔH , ΔS , and ΔG) would not be expected for a series of nonhomologous complexes where other contributions to ΔG_{bind} are of significant magnitude and remain undetermined. The concept of the “energetic epitope” model of Novotny et al. (1989) is used to define the binding interface in an energetic sense. In this study, the interaction is described in terms of the interaction energy component.

Results

The protein–protein interaction energies were computed for the available nonhomologous complexes (see Methods). These include the 12 serine endopeptidase–protein inhibitor and 15 antibody–antigen complexes listed in Table 1. To facilitate comparison in all cases, the relevant component of the interaction energy is expressed as a percentage of the total interaction energy, and not as an absolute value.

Analysis of the main-chain and side-chain interactions

Figure 1 shows the percentage contribution of different components to the interaction energy, in terms of (1) main-chain–main-chain, (2) main-chain (of enzyme/antibody)–side-chain (of inhibitor/antigen), (3) main-chain (of inhibitor/antigen)–side-chain (of enzyme/antibody), and (4) side-chain–side-chain interaction energies. It should be noted that while the main-chain–main-chain and side-chain–side-chain contributions are by definition symmetric (i.e., both molecules contribute equally to the interaction) those of the main-chain–side-chain are asymmetric. The two separate groups of protein complex are sorted in ascending order according to (1) the main-chain–main-chain percentage contributions to the total energy. The two groups of protein complex show very different characteristics in the predominance of these different contributions. The interaction enthalpies of the serine protease-inhibitor complexes are dominated by the main-chain–main-chain contribution representing on average 41% of the total interaction energy. In all cases this is the single largest contribution to the interaction, representing between 36–53% of the total interaction energy. The

Table 1. Serine endopeptidase-protein inhibitor and antibody-antigen complexes^a

PDB code	Protein complex	Resolution (Å)
Antibody-antigen complexes		
1fbi	F9.13.7—Guinea fowl egg lysozyme	3.0
1jhl	D11.15—Pheasant egg lysozyme	2.4
1mel	VH fragment CAB-LYS3—Hen egg lysozyme	2.5
1mlc	D44.1—Hen egg lysozyme	2.1
2hfl	HyHEL5—Hen egg lysozyme	2.5
3hfm	HyHEL10—Hen egg lysozyme	3.0
1vfb	D1.3—Hen egg lysozyme	1.8
1dvf	Idiotopic D1.3—Anti-idiotopic E5.2	1.9
1iai	Idiotopic 730.1.4—Anti-idiotopic 409.5.3	2.9
1jel	JE142—Histidine containing protein	2.8
1nma	NC10—N9 neuraminidase (influenza virus)	3.0
1nca	NC41—N9 neuraminidase (influenza virus)	2.5
1osp	184.1—Outer surface protein A (lipoprotein)	2.0
1ar1	7E2C50S—Cytochrome <i>c</i> oxidase	2.7
1ahw	5G9—Extracellular domain of coagulation factor III	3.0
Serine endopeptidase-inhibitor complexes		
(1) BPTI (Kunitz, Kunin) family		
2ptc	β -Trypsin—Bovine pancreatic trypsin inhibitor (BPTI)	1.9
(2) Kazal family		
1cho	α -Chymotrypsin—Overmucoid turkey third domain (OMTKY3)	1.8
1tgs	Trypsinogen—Pancreatic secretory trypsin inhibitor	1.8
(3) Potato-inhibitor-1 family		
1cse	Subtilisin Carlsberg—Eglin C	1.2
1acb	α -Chymotrypsin—Eglin C	2.0
(4) Potato-inhibitor-2 family		
4sgb	Streptomyces griseus proteinase B—Potato inhibitor PCI-1	2.1
(5) Bowman-Birk family		
1tab	β -Trypsin—Bowman-Birk inhibitor (AB-I)	2.3
(6) Squash seed inhibitors		
1mct	β -Trypsin—Bitter melon inhibitor	1.6
(7) Streptomyces subtilisin inhibitor (SSI) family		
2sic	Subtilisin BPN'—Streptomyces subtilisin inhibitor (SSI)	1.8
(8) Ecotin, trypsin inhibitor family		
1azz	Collagenase—Ecotin	2.3
(9) Elafin-like family		
1fle	Elastase—Elafin	1.9
(10) Hirustasin family		
1hia	Kallikrein—Hirustasin	2.4

^aSee Methods for further details.

next largest energy component averaging 24% is the main-chain (of enzyme)–side-chain (of inhibitor) representing between 12–30% of the interaction energy. This represents a very marked dominance of the main-chain component of the target enzyme with both main-chain and side-chain components of the inhibitor. In contrast in the antibody–antigen complexes, the main-chain–main-chain interaction energy represents on average only 17% (in range

7–26%) of the total interaction energy and in no case is it ever the dominant contribution in the interaction. This role is replaced by the side-chain–side-chain interaction energy averaging 36% (in range 23–54%) of the total interaction energy, dominant in 9 of the 15 cases, or the main-chain (antigen)–side-chain (antibody) interaction averaging 27% (in range 11–43%) dominant in four of the six remaining cases. In contrast, in the protease-inhibitor com-

Fig. 1. The percentage contribution of main-chain and side-chain components to the protein-protein interaction energy. PDB codes for individual complexes are given for both the antibody-antigen and protease-inhibitor complexes (see Table 1). Black bars, main-chain-main-chain interaction energy; dark grey bars, main-chain (enzyme/antibody)-side-chain (inhibitor/antigen); unfilled bars, main-chain (inhibitor/antigen)-side-chain (enzyme/antibody); light grey, side-chain-side-chain interaction energy. Ranked according to contribution of main-chain-main-chain interaction energy.

plexes, the side-chain-side-chain interaction energy represents on average only 17% (in range 9–26%) of the total interaction energy, and in all but two cases is of less importance than the main-chain (of enzyme)-side-chain (of inhibitor) interaction.

In terms of the main-chain-side-chain interaction the main contributor is generally the main-chain (of enzyme)-side-chain (of inhibitor) in the protease-inhibitor complexes (10/12 cases) and the main-chain (of antigen)-side-chain (of antibody) in the antibody-antigen complexes (11/15 cases), i.e., the side-chains of the ligand (inhibitor or antibody) recognize the main-chain of the target (enzyme or antigen). However, this is not always the case, and the complexes of VH fragment CAB-LYS3; hen egg lysozyme and JE142; histidine containing protein are peculiar in that the main-chain (of antibody)-side-chain (of antigen) make the single largest contributions to the interaction energy and they have the two lowest side-chain-side-chain, and two of the three highest main-chain-main-chain contributions, making them more like the protease-inhibitor complexes than the other antibody-antigen complexes.

Contribution of nonrotameric residues to the interaction energy

Here a nonrotameric residue is defined as a residue that does not fit into one of the commonly occurring side-chain conformational states known as rotamers. Rotamers correspond to dense clusters of points in χ -angle space found by statistical analysis of protein crystal structures (Ponder & Richards, 1987). The percentage of the interaction energy comprising nonrotameric residues is calculated here using the rotamer library of Tuffery et al. (1991) to define rotameric residue conformations. A residue is considered as being rotameric if its side-chain dihedral angles (χ_1 , χ_2 , and χ_3) are within 40° of those of a side-chain rotamer in the library. The rotational symmetry of the residue dihedral angles of Phe, Tyr, Asp (χ_2), and Glu (χ_3) is taken into account. The percentage contri-

bution of nonrotameric residues to the total interaction energy is given in Figure 2. The percentage of polar/aromatic residues (see Methods for definitions) is also shown. Nonrotameric residues contribute on average 21% (in the range 6–34%) to the interface interaction energy. The average contribution of nonrotameric residues in the antibody-antigen complexes (23% of interaction energy) is similar to that of the protease inhibitors (19%). Figure 2 would suggest that in the antibody-antigen and to a certain extent the protease-inhibitors complexes, >90% of the nonrotameric residue energy contribution is due to polar/aromatic residues at the interface. These can be broken down further by residue type. Figure 3A shows the percentage energetic contribution of polar/aromatic residue types in all complexes given that they are nonrotameric. Particularly problematic in terms of the rotameric definitions used are Arg, Asp, Asn, Glu, and Gln. This trend appears to be largely maintained in both the antibody-antigen and protease-inhibitor complexes. All these residues have side-chain functional groups that have two or more hydrogen bond donor/acceptor groups and thus the ability to form two (or more) hydrogen bonds. The control, Figure 3B, shows the percentage energetic contribution of all polar/aromatic residues (rotameric and nonrotameric). It can be seen that the energetic contributions of nonrotameric residue types do not simply reflect the contributions of the residue types in general. Further investigation reveals that it is the result of the almost free rotation about the $-\text{CH}_2-\text{CO}(\text{O}/\text{NH}_2)$ dihedral angle Asp/Asn χ_2 and Glu/Gln χ_3 . Figure 3C shows the percentage energetic contribution of nonrotameric polar/aromatic residues if these dihedral angles are not included in the test of rotamericity. The nonrotameric, virtually uniform χ_2 distribution of Asn/Asp has been noted before (Schrauber et al., 1993) with high resolution data (resolution $\leq 2.0 \text{ \AA}$; R -values ≤ 0.202). Database analysis of the dihedral angle distributions for high resolution data show that Asn(χ_2) and Gln(χ_3) show virtually uniform distributions while Asp(χ_2) and Glu(χ_3) show a slight preference for

Fig. 2. The percentage contribution of nonrotameric residues to the total interaction energy. Light grey bars, all amino acids; dark grey bars, polar/aromatic amino acids only.

the eclipsed conformation (M.W. MacArthur & J.M. Thornton, pers. comm.). In terms of the energetic rotamer based model described here, the magnitude of the nonrotameric contribution of Asp/Asn and Glu/Gln in experimental structures (being by far the largest combined contribution) could be a major reason for the inability of the rotamer model to fully reproduce the interaction energy seen in the crystal complex. However, it should be noted that the higher resolution protease-inhibitor complexes (see Table 1) have on average a lower percentage contribution of nonrotameric residues to the energetic epitope than the antibody-antigen complexes. The exception is arginine, whose nonrotameric conformations contribute extensively to the interaction energy in both types of complex. Unlike the above residues the nonrotameric behavior of Arg takes place in the $-sp^3-sp^3$ -hybridized alkyl dihedral angles χ_1 , χ_2 , and χ_3 . Inspection of the data indicates this is the result of relatively few nonrotameric residues participating in particularly favorable interactions, raising the suggestion that for particular residue types particularly favorable nonbonded interactions may extend the normal boundaries defined by internal energy constraints. Residues that are commonly nonrotameric have two or more hydrogen bonding donor/acceptor groups in the side-chain moiety (arginine has three). This may be a prerequisite for storing the energy necessary to overcome internal energy constraints.

Electrostatic component

The percentage contribution of the electrostatic energy to the total interaction energy can vary between ~10–40% of the interaction energy. The average electrostatic component of the antibody-antigen interaction (29%, $\sigma = 10\%$) is larger than that of the protease-inhibitors (24%, $\sigma = 5\%$) and as reflected by the standard deviation, the antibody-antigens show a greater spread (9–43%) than the protease-inhibitors (16–32%). However, the size of the data sets is small enough to make the significance of this difference uncertain.

Implications for protein-protein interaction energy

It is not possible in this study to calculate directly the effect of (1) the percentage contribution of main-chain vs. side-chain interaction energy or the effect of (2) the nonrotameric contribution on binding, since only the bound conformations of the complex are available. The multiple copy representation of side chains includes the original side-chain rotamer (from the experimental coordinates), in addition to other rotamers generated according to the rotamer library. However, if only side-chain rotamers of the rotamer library are included in the atomic representation, then any memory of the original side-chain rotamer is lost. This is like starting from the unbound conformation in that there is no bias toward the original (bound) side-chain conformation. However, the orientation and conformation of the main chain is identical to the bound state, and will in turn bias side-chain conformations toward those of the experimental structure. The extent of “recovery” of the total interaction energy following refinement without the original side-chain rotamers (as opposed to with them) is shown in Figure 4. The average recovery for the protease-inhibitor (94%) is much higher than for the antibody-antigen complexes (76%), and the spread of individual values are largely clustered about the mean value. Also shown is the percentage recovery of the combined main-chain-side-chain/side-chain-side-chain interaction energy. It can be seen that this value is highly correlated with the total interaction energy (slope ≈ 1 , $r = 0.97$). The percentage recovery of the main-chain-main-chain interaction energy is also shown. It is generally more conserved following refinement than the main-chain-side-chain/side-chain-side-chain component with average recovery for the antibody antigen (92%) being very high and similar in value to that in the protease-inhibitor complexes (98%). There would appear to be no direct correlation of these values to the percentage contribution of nonrotameric residues to the interaction energy (Fig. 2). However, there would appear to be a relationship with the extent of main-chain vs. side-chain contributions for the two protein groups as a whole. The greater involvement of

Fig. 3. The percentage contribution of polar/aromatic residues, by amino acid type, to the total interaction energy. Light grey bars, antibody-antigen complexes; dark grey bars, protease-inhibitor complexes. **A:** For all non-rotameric side-chains. **B:** For all (rotameric and nonrotameric) side-chains. **C:** For all nonrotameric side-chains excluding the dihedral angle χ_2 of Asp and Asn, and dihedral angle χ_3 of Glu and Gln.

the main-chain-main-chain interaction in the protease-inhibitor complexes provides the energetic and/or structural constraints to recover a greater extent of the overall interaction energy.

Contribution of different amino acid types to the interaction energy

Novotny et al. (1989), Padlan (1990), and Mian et al. (1991) observed that the aromatic and polar residues contribute prominently at the interface of antibody-antigen complexes. It has also been suggested that nonentropic residues contribute prominently at the

interface of protease-inhibitors (Krystek et al., 1993). The generality of these observations have been analyzed using a larger data set of nonhomologous complexes but in terms of the contribution to the interaction energy. In addition, the relative importance of different residue types in the receptor (enzyme/antibody) and ligand (inhibitor/antigen) interfaces have been analyzed. The interaction energy has been partitioned into different residue type contributions (see Methods). These are polar/aromatic, nonentropic, and hydrophobic. The relevant contributions are shown in Figure 5. In terms of the energetic epitope, polar/aromatic residues dominate the interaction in the antibody-antigen complexes averaging 81% of the energy, as opposed to 56% in the protease inhibitors. Nonentropic residues contribute much more to the interaction energy in the protease inhibitors (average, 23%) than the antibody antigens (average, 10%). This is in general agreement with the findings of Krystek et al. (1993) who analyzed a smaller data set. In addition, the hydrophobic residues contribute more to the protease-inhibitor interaction energy (average, 21%) than the antibody-antigen complexes (9%). The protease-inhibitor complexes have a higher contribution of nonentropic and hydrophobic interactions at the expense of the polar/aromatic contribution. The protein desolvation and loss in side-chain entropy are both predominantly entropic effects that would favor protease-inhibitor association over that of antibody-antigen association (not measured in this study).

It is useful to analyze residue contributions at the amino acid level. The individual residue type contributions are calculated as a percentage of the interaction energy of all complexes from a particular protein type, antibody, antigen, inhibitor, and protease.

Antibody binding

There is a predominance of tyrosine in the interaction of antibodies with their antigens. It represents over a quarter (26%) of the total interaction energy donated by the antibody. The contribution of Trp (7%) is also considerably more important in antibody binding than it is in the antigen (1%). However, equally notable is the low contribution of phenylalanine to the antibody interaction energy (2%). Negatively charged Asp (12%) and Glu (8%), and small dipolar residues Asn (9%), Ser (8%), Thr (5%), and Gly (5%) would also appear to play an important role in the antibody binding surface. If arginine (5%) is added, then these nine residues make up 87% of the overall interaction energy of the paratope. All of these residues have the capacity to hydrogen bond through their side-chain (except glycine, which can do so through its main-chain). Notably, five of the seven lowest energy residue type contributions (Val, Met, Pro, Leu, and Ala with $\leq 1\%$) have no side-chain hydrogen bonding capacity, reflecting the generally polar nature of the antigen's paratope. These results are generally consistent with previous studies of antibody CDR residue propensity (Padlan, 1990; Mian et al., 1991).

Protein antigen binding

While arginine makes a significant contribution in the antibody interaction energy, lysine (unlike the other charged amino acids) makes a negligible contribution (1%). However, the antigen Arg (15%) and Lys (15%) make the two largest contributions to the epitope, with Asp (9%), Tyr (6%), and Glu (4%) making large but proportionately less of a contribution than in the antibody. As in the antibody interface the residues Asn (11%), Ser (5%), Thr (5%), and Gly (5%) all appear to play an important role in the antigen binding surface.

Fig. 4. The extent of recovery of the interaction energy following refinement without the original side-chain rotamers as a percentage of the interaction energy following refinement with the original side-chain rotamers. Light grey bars, total interaction energy; dark grey bars, main-chain-side-chain and side-chain-side-chain interaction energy; unfilled bars, main-chain-main-chain interaction energy.

Protease inhibitor binding

In terms of protease-inhibitor binding, the overall predominance of arginine (18%) and lysine (9%) in the interaction of the inhibitor with the protease possibly reflects the fact that the inhibitor P₁ specificity residue of the inhibitor is Arg/Lys in 5 of the 12 complexes. Similarly, leucine (9%) and methionine (5%) are the P₁ specificity residue in six of the remaining seven structures. However, if these contributions are ignored, the nonentropic cysteine (9%) and proline (7%), and hydrophobic valine (6%) and isoleucine (6%) make prominent contributions.

Serine protease binding

In terms of the protease targets themselves, the major contributing residues serine (14%), glycine (13%), aspartate (9%), and histidine (8%) largely reflect the recurrence of certain homologous “key” residues at the proteases active site, which are recognized by the inhibitor. Thus, the protease-inhibitor interactions have conserved residue patterns, which are utilized in the recognition process throughout this group of complexes.

The residue-type interactions can be further analyzed at the level of residue-residue pair interactions. For the antibody-antigen com-

Fig. 5. The interaction energy of the receptor-ligand complex is partitioned according to residue type (see Methods). Light grey bars, polar/aromatic; dark grey bars, nonentropic; unfilled bars, hydrophobic.

plexes, there is a striking dominance of charge–charge interactions and charge–aromatic interactions. The ion pair interactions Asp-Arg (5.8%), Asp-Lys (6.1%), Glu-Arg (5.1%), and Glu-Lys (3.2%) contribute 20% of the favorable interaction energy in the antibody-antigen complexes (i.e., 2% of the pair interactions contribute 20% of the energy). The charge–aromatic (tyrosine) interactions Tyr-Arg (2.5%), Tyr-Lys (3.3%), Tyr-Asp (4.5%), and Tyr-Glu (1.1%), and the polar/aromatic (tyrosine) interactions Tyr-Asn (4.0%), Tyr-Gln (1.4%), Tyr-Ser (2.2%), and Tyr-Thr (1.4%), together with the charge–charge interactions, contribute 41% of the favorable interaction energy (i.e., 6% of the pair interactions contribute 41% of the energy). This results in a highly skewed distribution of relative importance of individual pair interactions. Given the homologous nature of the protease-inhibitor interaction motif, the residue–residue pair interactions landscape appears to be surprisingly evenly populated (figure not shown), when compared to that of the antibody-antigen complexes (which are sparsely populated with several high intensity peaks). The outstanding pair interactions include expected pairing resulting from the homologous nature of the binding sites, notably Arg-Asn (2.0%), Arg-Asp (3.5%), and Lys-Asp (4.2%). However, there are a large number of minor contributing pair interactions in the ~1% level that represent residue pair interactions from the nonentropic and hydrophobic types (nonentropic-nonentropic: Ala-Gly, Cys-Gly, and Gly-Pro; nonentropic-hydrophobic: Ala-Phe, Cys-Leu, Cys-Met, Gly-Leu, Gly-Met, Gly-Val, and Pro-Leu; hydrophobic-hydrophobic: Ile-Phe, Leu-Leu, Leu-Met, and Leu-Val). Thus, the pair interactions used by the antibody antigens appear much more limited than those of the protease-inhibitor complexes.

Discussion

The differences in the contribution of the main-chain/side-chain components to the overall interaction energy in the two different classes of protein complex is pronounced. In the protease-inhibitor complexes, the predominant contribution to the interaction energy

comes from the main-chain–main-chain interaction energy where as in the antibody-antigen complexes it is the side-chain–side-chain or the side-chain–main-chain interaction that predominates with the main-chain–main-chain contribution being of much less importance. It has been suggested that this may be a result of the greater probability of main-chain atoms being at the interfaces of the protease-inhibitor complexes. This would be reflected by the greater percentage of main-chain surface area buried on complex formation. Figure 6 shows the percentage contribution of main-chain atoms to the change in surface area on complex formation. This is compared with the percentage contribution of the main-chain–main-chain interaction enthalpy (Fig. 1). There is some correlation between the percentage of surface area buried and interaction enthalpy within both of the two classes of complex. However, the average main-chain atom burial in the enzyme inhibitors (19%) is only slightly greater than that of the antibody antigens (13%) and does not explain the large difference in the contribution to the main-chain–main-chain interaction enthalpy (on average 41% and 17%, respectively). The difference between the two systems is explained by the differing extent to which one main-chain interacts directly with the other molecule's main-chain. In a biological sense these differences are possibly a reflection of the different roles played by the two classes of protein, while at a chemical level the differences may partly explain the large difference in binding affinity of the two classes of complex. The serine protease inhibitors need to bind tightly to their target proteases. This may be best achieved by a highly constrained main-chain conformation, which recognizes the main-chain motif of its target protein. In so doing, the inhibitor is highly committed to the enzyme in an evolutionary sense since a change in the target main-chain motif would severely compromise the binding affinity of the inhibitor; however, the arrangement is entropically highly favorable since the major determinant of the interaction energy (the main-chain) is in general more conformationally restrained than side-chain degrees of freedom. The very specific nature of the protease-inhibitor interaction is consistent with a large body of thermodynamic and crystallo-

Fig. 6. Comparison of the percentage contribution of main-chain atoms to the change in solvent accessible surface area on complex formation (light grey bars) to the percentage contribution of the main-chain–main-chain interaction energy to binding (dark grey bars).

graphic data available on inhibitor mutants (Lu et al., 1997; Qasim et al., 1997). Although mutations at the P₁ position of inhibitors can severely compromise the binding energy, (1) the P₁ residue of the mutant protein always binds in the same way in the specificity pocket irrespective of the deleterious effect on the binding constant, (2) any changes in conformation on complex formation are very small suggesting that the other components of the interaction are sufficient to retain the integrity of the binding conformation. In the antibody-antigen systems, a very different set of criteria prevail. To be a general utility binding molecule, the antibody main-chain conformation cannot be too committed to any particular main-chain motif since it must bind to all protein motifs— α -helix or β -sheet protein, polyanionic nucleic acids, and even lipids. For example, a binding motif that exploited the potential of an N-terminal helix dipole would clearly have problems binding a large proportion of regions on helical proteins. In some cases, however, a particular antibody may partially overcome these constraints via the specific conformations of its CDRs. However, a more extensive repertoire of interactions is most easily achieved through the large number of potential side-chain arrangements, which are sterically accessible at the antibody surface. In fact, this suggests the differences in the main-chain/side-chain contribution explain the difference in binding promiscuity of the two systems. The number of accessible degrees of freedom at the protein-protein interface is related to the number of false positive solutions that score higher than the near native solution in molecular docking (Jackson et al., 1998). The implications of these differences are important for computer simulation of molecular docking. Clearly a protein-protein interface that is largely determined by a static main-chain-main-chain interaction is more easily predicted than a somewhat plastic interface largely determined by side-chain-side-chain interactions. Indeed the difference in the level of success achieved in molecular docking with unbound subunits of the protease-inhibitor complexes as opposed to the antibody-antigen complexes largely confirms this view (Jackson et al., 1998), as do the results of the “recovery” of total interaction energy with and without the original side-chain conformation being present. In this context a particularly relevant question is: how representative are these different types of binding for protein-protein complexes in general? Both classes of binding protein display characteristics that may not be typical of protein-protein interactions in general. The overall dominance of the main-chain-main-chain interaction in the protease-inhibitor complexes might be expected to be somewhat class specific given the nature of serine protease function and the substrate like nature of inhibitor binding. Similarly the antibody-antigen interfaces are generally recognized as being less complementary than homodimers, protease-inhibitor complexes, and permanent hetero-complexes when analyzed by a shape correlation statistic (Lawrence & Colman, 1993), or interface gap volume (Jones & Thornton, 1997). Furthermore, given the extent of binding diversity and corresponding limit on structural diversity, recognition through the side-chains of the antibody is likely to be of major importance. While a full analysis of all other protein-protein interactions is beyond the scope of this paper, a small number of nonhomologous heterodimers and homodimers have been sampled. These were taken from the analysis of Jones and Thornton (1997). The results are shown in Figure 7 for 4 heterodimers (metallo proteinases were not included) and 10 homodimers ordered by main-chain-main-chain contribution, and placed in the context of the maximum and minimum main-chain-main-chain contributors of the antibody-antigen and protease-inhibitor complexes. The four heterodimers

Fig. 7. The percentage contribution of main-chain and side-chain components to the interaction energy, as in Figure 1. PDB codes are given for the heterodimer and homodimer complexes. The complexes with the maximum and minimum main-chain-main-chain interaction energy in the antibody-antigen and protease-inhibitor complexes are shown for purpose of comparison.

are significantly antibody-antigen “-like” in their interface interactions, with a fairly small main-chain-main-chain contribution, and relatively large side-chain-side-chain contribution. The 10 homodimers span the range of values, with 5 homodimers showing a greater main-chain-main-chain contribution than the maximum antibody-antigen contributor (1me1). Two of these (1cdt and 2sod) have main-chain-main-chain contributions within the maximum/minimum values of the protease-inhibitor complexes. Furthermore, both these complexes have a correspondingly small side-chain-side-chain contribution reminiscent of the contribution seen in the protease-inhibitor complexes. This analysis of dimer interactions can only be considered a “straw poll.” There are representatives of both types of binding from among the small number of homodimeric complexes studied here. Hence, the antibody-antigen and protease-inhibitor complexes may represent two ends of the spectrum.

The contribution of nonrotameric residues to the interaction energy is approximately the same in the antibody-antigen and protease-inhibitor complexes, averaging ~20% of the interaction energy. In addition to the nonrotameric behavior of the $-\text{CH}_2\text{-CO}(\text{O}/\text{NH}_2)$ dihedral of Asn, Gln, Asp, and Glu, the nonrotameric behavior of arginine can have the most severe consequences for reproducing the observed interaction energy. A rotameric description of side-chain conformation is inadequate for these particular residues. However, the magnitude of the nonrotameric contribution does not appear to be linked directly to the ability of the rotameric side-chain model to recreate the energetics of the protein-protein interface. The protease-inhibitor interaction energies are far less affected than those of the antibody-antigen by the inadequacies of the rotamer model when the original side-chain rotamer is absent from the simulation. Hence, it is the relative contribution of main-chain-main-chain vs. side-chain-side-chain/main-chain interaction energy that is of greatest importance. This is expected to most affect the applicability of the current model to protein-protein docking problems. Indeed, the level of success using the model is expected to be correlated to the degree of main-chain-main-chain involvement, making the docking methodology useful for a limited number of systems. Prior knowledge of the likely extent of main-chain-main-chain involvement would assist in deciding whether the current

method is appropriate. The rotameric model tells us much about the extent of diversity between different protein-protein interactions and suggests problems that must be addressed if a more universal docking method is to be developed.

Methods

Structure selection

The set of crystallographically determined serine endopeptidase protein-inhibitor and antibody-antigen complexes were extracted from the Brookhaven Data Bank (PDB) (Bernstein et al., 1977) January 1998 release plus pre-releases (daily mirror of PDB up to February 1998). From this list of complexes, those with a resolution greater than 3.0 Å were excluded. The structures F9.13.7 (1fbi) and the collagenase-ecotin complex (1azz) had a number of missing atoms that require side-chain rebuilding. Since none of these residues are in the relevant protein-protein interface these residues were rebuilt as alanine residues. The effect on the calculation of interaction energy is expected to be small. Where more than one structure was available for a given complex, the highest resolution structure was used. In the serine endopeptidase protein-inhibitor complexes, the enzyme components or the inhibitor components could be homologous, as long as both components were not homologous. Enzyme structures were treated as homologous if they were classified as being in the same protein superfamily in the structural classification of proteins (SCOP) (Murzin et al., 1995). Protease inhibitors were classified as homologous if they belong to the same family of serine protease inhibitors (Laskowski & Kato, 1980; Bode & Huber, 1992) classified by sequence similarity, topological similarity, and mechanism of binding. In the case of the antibody-antigen complexes, this constraint must be relaxed since antibodies are homologous, as are certain antigens such as lysozyme and neuraminidase. Hence, complexes are included if the antigen recognition site is different (implying a different antibody CDR-antigen interaction). All hetero-atoms including all water molecules were ignored in subsequent calculations. The table of structures selected using the above criteria are given in Table 1, including PDB code, protein name, and resolution.

Interaction energy

The interaction energy in the protein-protein interface was calculated starting from the structurally determined crystal complex. The energetic refinement of the interface was restricted to rigid-body energy minimization of the interface. Side-chain degrees of freedom were not allowed to move during refinement of the complex. The method of rigid-body minimization including the molecular mechanics representation has already been described in detail (Jackson et al., 1998). In brief, only interface residues whose C β (or C α in the case of glycine) are within 15 Å of any C β of the other molecule are included in the minimization. A 10 Å atom-atom nonbonded cutoff distance was used. Also used was a distant dependent dielectric, $\epsilon = 4r$ (McCammon et al., 1979; Pickersgill, 1988), with scaling of van der Waals and electrostatic interactions for unrealistically close atomic contacts (Jackson et al., 1998). The enzyme or antibody (generally the larger molecule) was kept stationary while the inhibitor or antigen (generally the smaller molecule) was moved in a rigid-body manner according to the path of steepest descents determined by the numerical derivatives calcu-

lated using the atomic model and potential energy function. Note that no solvation energy contribution is included in the calculation. Minimization was continued until the energy of the system decreased by less than 1×10^{-6} kcal/mol for any given step. The above was implemented in a modified version of MULTIDOCK v1.0 (Jackson et al., 1998) (available at <http://www.icnet.uk/bmm>), which allows specification of rigid-body minimization alone and in which individual residue interaction energies and their side-chain/main-chain contributions are output for analysis. In "recovering" the interaction energy, while excluding the original crystallographic rotamer, a different interface refinement protocol was followed from that described above. This involved both mean field refinement and energy minimization. Two separate simulations were carried out (1) with the experimental (original) side-chain rotamer present as well as those of the rotamer library and (2) without the experimental rotamer present and only rotamers from the library.

The interaction energy has been partitioned into different residue type contributions. These are polar/aromatic (Arg, Lys, Glu, Asp, Asn, Gln, Tyr, Trp, His, Cys[h], Ser, and Thr), nonentropic (Pro, Gly, Ala, and Cys[-Cys]), and hydrophobic (Ile, Leu, Met, Val, and Phe). The nonentropic residues do not possess any side-chain degrees of freedom; hence, there is no loss in side-chain conformational entropy on binding. Certain residues could be placed in more than one group, however, given that redundancy does occur the most appropriate grouping has been chosen depending on side-chain length and physicochemical properties.

Surface area

The change in surface area of main-chain and side-chain atoms in the interface was calculated using the complexed and solvent separated subunits. The algorithm of Lee and Richards (1971) was used to calculate the solvent accessible surface area as implemented in the program NACCESS (Hubbard & Thornton, 1993), using a solvent probe radius of 1.4 Å and the van der Waal radii of Chothia (1976).

Acknowledgments

R.M.J. is grateful to Prof. Janet Thornton for helpful discussions and critical reading of the manuscript, and to Dr. Malcolm MacArthur for sharing results of unpublished work. This work was supported by the Wellcome Trust in the form of a Career Development Fellowship to R.M.J.

References

- Ajay, Murcko MA. 1995. Computational methods to predict binding free energy in ligand-receptor complexes. *J Med Chem* 38:4953-4967.
- Bernstein FC, Koetzle TF, Williams GJ, Meyer EE Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer based archival file for macromolecular structures. *J Mol Biol* 112:535-542.
- Bode W, Huber R. 1992. Natural protein protease inhibitors and their interactions with proteinases. *Eur J Biochem* 204:433-451.
- Cherfils J, Janin J. 1993. Protein docking algorithms: Simulating molecular recognition. *Curr Opin Struct Biol* 3:265-269.
- Chothia C. 1976. The nature of the accessible and buried surfaces in proteins. *J Mol Biol* 105:1-12.
- Gabb HA, Jackson RM, Sternberg MJE. 1997. Modeling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* 272:106-120.
- Hubbard SJ, Thornton JM. 1993. NACCESS, computer program. London: Department of Biochemistry and Molecular Biology, University College London.

- Jackson RM, Gabb HA, Sternberg MJE. 1998. Rapid refinement of protein interfaces incorporating solvation: Application to the docking problem. *J Mol Biol* 276:265–285.
- Jackson RM, Sternberg MJE. 1995. A continuum model for protein–protein interactions: Application to the docking problem. *J Mol Biol* 250:258–275.
- Janin J. 1995. Protein–protein recognition. *Prog Biophys Molec Biol* 64:145–166.
- Jones S, Thornton JM. 1997. Analysis of protein–protein interaction sites using surface patches. *J Mol Biol* 272:121–132.
- Krystek S, Stouch T, Novotny J. 1993. Affinity and specificity of serine endopeptidase–protein inhibitor interactions. Empirical free energy calculations based on X-ray crystallographic structures. *J Mol Biol* 234:661–679.
- Laskowski M, Kato I. 1980. Protein inhibitors of proteinases. *Annu Rev Biochem* 49:593–626.
- Lawrence MC, Colman PM. 1993. Shape complementarity at protein/protein interfaces. *J Mol Biol* 234:946–950.
- Lee B, Richards FM. 1971. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol* 55:379–400.
- Lu WY, Apostol I, Qasim MA, Warne N, Wynn R, Zhang WL, Anderson S, Chiang YW, Ogin E, Rothberg I. 1997. Binding of amino acid side-chains to S-1 cavities of serine proteinases. *J Mol Biol* 266:441–461.
- MacCallum RM, Martin AC, Thornton JM. 1996. Antibody–antigen interactions: Contact analysis and binding site topography. *J Mol Biol* 262:732–745.
- McCammon JA, Wolynes PG, Karplus M. 1979. Picosecond dynamics of tyrosine side-chains in proteins. *Biochemistry* 18:927–942.
- Mian IS, Bradwell AR, Olson AJ. 1991. Structure, function and properties of antibody binding sites. *J Mol Biol* 217:133–151.
- Murzin A, Brenner SE, Hubbard T, Chothia C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structure. *J Mol Biol* 247:536–540.
- Novotny J, Brucoleri RE, Davis M, Sharp KA. 1997. Empirical free energy calculations: A blind test and further improvements to the method. *J Mol Biol* 268:401–411.
- Novotny J, Brucoleri RE, Saul FA. 1989. On the attribution of binding energy in antigen–antibody complexes McPC 603, D1.3, and HyHEL-5. *Biochemistry* 28:4735–4749.
- Padlan EA. 1990. On the nature of antibody combining sites: Unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins* 7:112–124.
- Pickersgill RW. 1988. A rapid method of calculating charge–charge interaction energies in proteins. *Protein Eng* 2:247–248.
- Ponder JW, Richards FM. 1987. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* 193:775–791.
- Qasim MA, Ganz PJ, Saunders CW, Bateman KS, James MNG, Laskowski M. 1997. Interscaffolding additivity. Association of P-1 variants of eglin C and of turkey ovomucoid third domain with serine proteinases. *Biochemistry* 36:1598–1607.
- Schrauber H, Eisenhaber F, Argos P. 1993. Rotamers: To be or not to be? An analysis of amino acid side-chain conformations in globular proteins. *J Mol Biol* 230:592–612.
- Sela M, Pecht I. 1996. The nature of the antigen. *Adv Protein Chem* 49:289–323.
- Sternberg MJE, Gabb HA, Jackson RM. 1998. Predictive docking of protein–protein and protein–DNA complexes. *Curr Opin Struct Bio* 8:250–256.
- Tuffery P, Etchebest C, Hazout S, Lavery R. 1991. A new approach to the rapid determination of protein side-chain conformations. *J Biomol Struct Dyn* 8:1267–1289.
- Webster DM, Henry AH, Rees AR. 1994. Antibody–antigen interactions. *Curr Opin Struct Biol* 4:123–129.
- Weng Z, Vajda S, Delisi C. 1996. Prediction of protein complexes using empirical free energy functions. *Protein Sci* 5:614–626.
- Wilson IA, Stanfield RL. 1993. Antibody–antigen interactions. *Curr Opin Struct Biol* 3:113–118.