PROTEINS: Structure, Function, and Genetics 35:364–373 (1999)

# Use of Pair Potentials Across Protein Interfaces in Screening Predicted Docked Complexes

**Gidon Moont, Henry A. Gabb, and Michael J.E. Sternberg***
*Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, London, United Kingdom*

***ABSTRACT*** **Empirical residue–residue pair potentials are used to screen possible complexes for protein–protein dockings. A correct docking is defined as a complex with not more than 2.5 Å root-mean-square distance from the known experimental structure. The complexes were generated by "ftdock" (Gabb et al. J Mol Biol 1997;272:106–120) that ranks using shape complementarity. The complexes studied were 5 enzyme-inhibitors and 2 antibody-antigens, starting from the unbound crystallographic coordinates, with a further 2 antibody-antigens where the antibody was from the bound crystallographic complex. The pair potential functions tested were derived both from observed intramolecular pairings in a database of nonhomologous protein domains, and from observed intermolecular pairings across the interfaces in sets of nonhomologous heterodimers and homodimers. Out of various alternate strategies, we found the optimal method used a mole-fraction calculated random model from the intramolecular pairings. For all the systems, a correct docking was placed within the top 12% of the pair potential score ranked complexes. A combined strategy was developed that incorporated "multidock," a side-chain refinement algorithm (Jackson et al. J Mol Biol 1998;276:265–285). This placed a correct docking within the top 5 complexes for enzyme-inhibitor systems, and within the top 40 complexes for antibody–antigen systems. Proteins 1999;35:364–373.** © 1999 Wiley-Liss, Inc.

**Key words: predictive docking algorithm; pair potentials; protein modeling; enzyme-inhibitor; antibody–antigen**

## INTRODUCTION

The diversity of interactions between residues provides the specificity of recognition in protein folding and ligand binding.[1–3] A simple model for these interactions is provided by residue–residue pair potentials. These have been widely used to evaluate the stability of protein folds in prediction.[4–10] In this article we evaluate the use of pair potentials to identify a near-native predicted model for a protein–protein complex from decoys of false positives obtained from our rigid-body docking program.[11]

The protein docking problem is to start with coordinates of two molecules in their uncomplexed state and hence predict the structure of the complex (reviewed in refs. 3, 12, and 13). A solution to this problem is becoming increasingly important as the number of experimentally determined protein structures (or protein domains) is increasing rapidly, without the corresponding characterization of their docked complexes (see the present entries in the protein data bank (PDB).[14,15]) Advances in computing have lead to the development of several algorithms that tackle the step of exhaustively searching all rigid-body dockings.[11,12,16–22] The approaches primarily match shape complementarity without too many steric clashes. Some then filter on burial of hydrophobic surfaces[16] and/or electrostatic complementarity.[11,17,18]

For many test systems, these approaches generate one or more complexes that are close to the native (typically root-mean-square [RMS] distance for Cα atoms of less than 2.5 Å at the interface) but also generate several false positives of comparable score to the true positive. The scoring function used during exhaustive searching must be fast to evaluate. However, more sophisticated and time-consuming treatments can be applied as a subsequent step of screening a limited set of alternative dockings.

Strategies that have been explored for a subsequent screening include atomic solvation potentials, empirical function for atom–atom surface contacts, and continuum models with Poisson-Boltzmann electrostatic calculations.[12,23–26] These approaches require a decomposition of the effects stabilizing the complex with consequential simplifications and omissions. In addition, the treatment of electrostatic effects is particularly sensitive to atomic positions, so these screening approaches tend to have a limited radius of convergence. Also, some of these approaches are time-consuming and so are less appropriate for screening hundreds of complexes. Therefore there is a requirement for an alternative strategy for screening that is both robust, with respect to the detailed atomic interaction, and fast enough to be applied to a large set of complexes.

These considerations have led us to evaluate the use of residue–residue pair potentials for screening docked complexes. Following earlier work,[27,28] several groups have derived these potentials from frequencies of residue–residue pairs in an appropriate database of experimentally determined protein structures.[4,5,7,9,10,29–31] The theory is that by applying Boltzmann's principle to the ratio of

Grant sponsor: EMBO.
*Correspondence to: Dr. Michael Sternberg, Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, Lincoln's Inn Fields, P.O. Box 123, London WC2A 3PX, UK. E-mail: m.sternberg@icrf.icnet.uk
Received 25 September 1998; Accepted 21 January 1999

observed to expected frequencies of pairings between two residue types one obtains an estimate of the mean force potential between those two residue types. This potential should then incorporate all the pertinent thermodynamic effects, including protein–solvent effects, interresidue van der Waals forces, and electrostatic interactions. The use of residue level (rather than atomic level) potentials provides a smoothness in the energy landscape that is likely to reduce the sensitivity of the function to precise atomic position. In addition, residue–residue potentials are fast to evaluate.

Residue pair potentials are often used in protein fold recognition (i.e., threading) to evaluate the fit of a sequence of unknown structure onto a known fold.[4–9] In addition, the potentials can be used to evaluate simplified folding simulations, including those on lattices.[32] However, several problems have been identified in simply applying Boltzmann's equation to observed frequencies to obtain a potential of mean force.[33–35] In particular, the difficulty in correctly identifying the random state and the validity of the quasi-chemical approximation that neglects the chain connectivity have been emphasized by some investigators.[33,35] With the quasi-chemical approximation there remains several possible reference states, including one that is purely compositional (mole-fraction) or one that incorporates the differing tendencies of residues to make pairs (contact-fraction).

In this article we acknowledge the problems of deriving potential of mean force from pairing frequencies and simply use the formalism to derive a statistical log odds ratio. These log odds are then used to screen docked complexes generated by our recently developed docking algorithm.[11] This employed a Fourier correlation approach[19] to identify protein–protein complexes with a high degree of shape complementarity, together with the absence of unfavorable electrostatic interactions. The algorithm was evaluated by starting with the unbound coordinates of six enzyme–inhibitor and of two antibody–antigen complexes. An additional two antibody–antigen systems were used, where the antigen was unbound, but the antibody was from the bound complex. After docking, a loose biochemical filter was applied that required either one of the enzyme active-site residues to contact any part of the inhibitor, or the antibody combining site to contact any part of the antigen. Gabb and colleagues[11] showed that for all but one system, after docking and loose filtering a set of between 25 and 750 complexes would have to be examined, within which was known to be a prediction with better than 2.5 Å (RMS) deviation for Cα atoms of the interface. The antibody results were generally poorer. This is possibly because of the differing recognition stereochemistry and lower affinity of antibody–antigen interactions, compared to those for enzyme-inhibitors.[36]

## METHOD

### Generation of the Pair Potential Matrices

In this study, pair potentials are used to score interactions across an interface rather than to evaluate the stability of a protein fold. Accordingly, the form of the pair potential that is used considers only propensities of residue types to be with a given spatial cutoff distance, and does not need to include terms such as sequence separations that are often incorporated into potentials used for protein folding studies.[4,31]

Four types of pair potentials were used; a residue level potential based on Cβ atoms, a residue level potential based on all atoms, a residue level potential based on all side-chain atoms, and an atom level potential (see below for details). In addition, three different datasets were tested; a set of interdomain pairings from a nonredundant set of heterodimers, a set of interdomain pairings from a nonredundant set of homodimers, and a set of intramolecular residue pairings from a nonredundant set of protein domains. The set of heterodimers was taken from an extensive set used in a recent study involving protein interface surfaces,[37] and consisted of only 11 complexes with a resolution of 2.5 Å or better. The set of nonredundant dataset of homodimers was taken from a list from the same recent study,[37] and gave a set of 23 structures with a resolution of 2.5 Å or better.

In order to obtain a nonredundant dataset for the intramolecular potentials, that excluded duplicate homologous entries, a set of protein domains was taken from the Structural Classification of Proteins (SCOP)[38] database (Version 1.37). The best resolution structure of each superfamily was taken for each of the first four classes (α, β, α/β, α + β). Superfamilies where there was no structure with a resolution of 2.5 Å or better were ignored. This dataset totaled to 385 domains.

To generate the potentials, the number of pairings between each type of residue was counted. For the residue level potential based on Cβ atoms, a pair, $c_{i,j}$, was defined as occurring between residues $i$ and $j$ if the Cβ atoms in the two residues were within a given distance cutoff. In the case of the intramolecular pairings, the pairs were within the domain. In the case of the homodimers and heterodimers, the pairs spanned across the dimer interface.

For the residue level potential based on all atoms, a pair, $c_{i,j}$, was defined as occurring between residues $i$ and $j$ if any of the atoms in the two residues were within a given distance cut off. Similarly for the residue level potential based on all side-chain atoms, a pair, $c_{i,j}$, was defined as occurring between residues $i$ and $j$ if any of the side-chain atoms in the two residues were within a given distance cutoff.

For the atom level potential, each atom on every residue was assigned an atom type. We used 40 atom types, the same used as in a previous study of atomic level pair potentials.[39] To generate the potential we then did essentially the same as was done for the residue level potentials. A pair was defined as occurring between atom types $i$ and $j$ if they were within a given distance cutoff.

There are two methods of calculating the expected number of pairs between residues $i$ and $j$. One, the mole-fraction method, $e_{(\text{mole-fraction})i,j}$, is proportional to the product of the fractional abundances of the two residues in the pair. The other, the contact-fraction method,

$e_{(\text{contact-fraction})i,j}$, is proportional to the propensities of the two residues to be paired with any residue at all. that is,

$$e_{(\text{mole-fraction})i,j} = C \times \frac{n_i}{N} \times \frac{n_j}{N}$$

$$e_{(\text{contact-fraction})i,j} = C \times \frac{c_i}{C} \times \frac{c_j}{C}$$

$$c_i = \sum_{j=1}^{j=20} c_{i,j}$$

$$C = \sum_{i=1}^{i=20} c_i$$

$$N = \sum_{i=1}^{i=20} n_i$$

where $n_i$ and $n_j$ are the total occurrences of each residue.

For all types of pair potential we considered it necessary for the expected value of any given pair to have a value of at least 5. If the expected value was lower than this we considered it an indication that there were not enough data to generate a useful pair potential value. Where this occurred we made the pair potential value for that pair equal to zero. This was not common for most types of pair potential, but did sometimes occur with small distance cutoff values. However, in the case of the atom level calculations using the heterodimer and homodimer datasets, not enough data were available to make even a small number of the expected values large enough to be acceptable. Therefore we did not use those datasets for atomic level pair potentials.

The score for each pair was then taken as the log fraction of the actual count and the expected count.

$$\text{score}_{i,j} = \text{score}_{j,i} = \log\,(c_{i,j}/e_{i,j}).$$

The value of the score for each pair can be considered simply as a statistical measure of likelihood of that pair occurring. Since the quantity is a log fraction, the total likelihood for a structure is the sum of all the individual scores. A widely used approach is to equivalence this method to Boltzmann's law, and thereby relate the negative of the log fraction to an estimate of relative free energies for different residue pairings. This was not done for this work, although it would not alter the actual results.

Figure 1 shows the 205 different scores of a residue level potential based on all atoms, calculated using $e_{(\text{mole-fraction})}$ with a distance cutoff of 8 Å. There is a score value for each type of residue–residue pair, including pairings where the type is identical. The charge–charge interactions have score values of the expected sign (apart from the case of arg-arg), and pairings between hydrophobics have generally positive values. This is because hydrophobics tend to pair, yet the mole-fraction method does not take this information into account. The values calculated using $e_{(\text{contact-fraction})}$ exhibit this feature to a lesser extent.

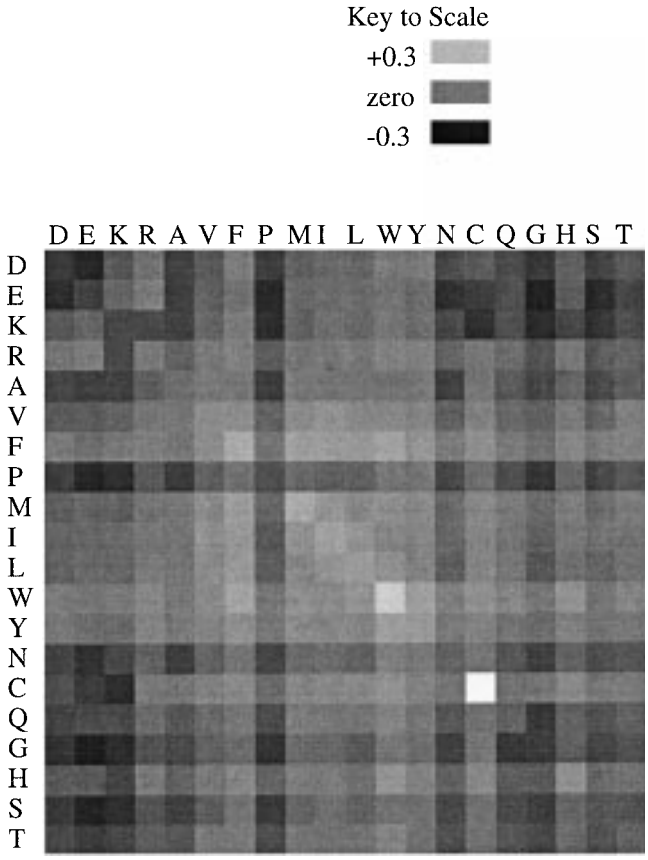


Fig. 1. Example matrix: generated from 385 SCOP domains, using all atoms and $e_{(\text{mole-fraction})}$ with an 8 Å cutoff.

## Scoring a Docked Structure

A score was calculated for each complex by summing the appropriate scores of pairs that spanned the interface of the complex. The pairs were considered to exist in exactly the same way as when generating the matrices. That meant that the exact method of scoring was different for each different type of pair potential.

A minimum relative surface accessibility (MRSA) was used as a further constraint on whether two residues are paired.[40] The program used to calculate this value was NACCESS, written by Simon Hubbard when at University College London (present URL is http://sjh.bi.umist.ac.uk/naccess.html). This was used to assign a relative percentage accessibility value to a residue while in the unbound state, by dividing the accessible area of the residue by its accessibility in a standard conformation. By making the constraint that both residues in the bound pair have to have had at least a given MRSA while unbound, residues that are not accessible (either buried or on the surface but largely unexposed) when unbound can be ignored.

### Test Systems

All the methods above were tested on the nondegenerate results of the systems used by Gabb and coworkers[11] to test ftdock. The 10 systems used in that study where the exhaustive dataset at the time. Since then there are several new systems we could have used as well, but by using these results there was a readily available set of possible complexes with which to evaluate our scoring methods. (These complexes had also been evaluated by multidock, discussed below, thereby making comparisons readily available.) We used the list of possible complexes generated by a global search of binding space followed by filtering using biological information, followed by local refinement with a 1.5-Å surface thickness. It was known that there was at least one correct docking (defined as being a complex with RMS of 2.5 Å or less from the experimentally determined complex) in these lists of possible complexes for 9 of the 10 systems. We did not use the one system where there was no correct docking to be found (SIC).[11]

The enzyme-inhibitor systems consisted of the following experimentally determined complexes and components of those complexes (with their PDB codes):

1. CHI, human pancreatic trypsin inhibitor (1apt)[41] bound to α-chymotrypsinogen (1chg).[42] The PDB code of the complex is 1cgi.[43]
2. CHO, ovomucoid (2ovo)[44] bound to α-chymotrypsin (5cha).[45] The PDB code of the complex is 1cho.[46]
3. KAI, bovine pancreatic trypsin inhibitor (1bpi) bound to kallikrein A (2pka).[47] The PDB code of the complex is 2kai.[48]
4. PTC, bovine pancreatic trypsin inhibitor (4pti)[49] bound to trypsin (2ptn).[50] The PDB code of the complex is 2ptc.[51]
5. SNI, chymotrypsin inhibitor 2 (2ci2)[52] bound to subtilisin (1sup). The PDB code of the complex is 2sni.[53]

The antibody–antigen systems used consisted of the following Fab's and Fv's bound to lysozyme (1lza)[54]:

1. FDL, D44.1 Fab (1mlb), complex (1fdl).[55]
2. MLC, D1.3 $F_v$ (1vfa), complex (1mlc).[56]
3. HFL, HyHel-5 Fab (2hfl).[57]
4. HFM, HyHel-10 Fab (3hfm).[58]

Unfortunately, native crystal structures for the antibody in 2hfl and 3hfm are not yet solved. So the bound forms of the Fab's are used in HFL and HFM docking. Only the Fv regions of 1mlb, 2hfl, and 3hfm were used during docking.

The test systems were tested using varying parameters. For the residue level potential based on Cβ atoms, the distance cutoff was varied from 5 to 15 Å at 1 Å steps. For the other three types of potential, the distance cut off was varied from 4 to 10 Å at 1 Å steps. This difference is due to the first type being measuring essentially an interresidue distance, whereas the other three methods are measuring an interatom distance. The MRSA was varied at values of 0, 5, and 20%.

The complexes were then ordered according to their scores, and the position in the sorted list of the correct dockings was determined.

Where it was found that the true complex with a given parameter set would have less than 20 pairs, that set was not used. This was because the results were too erratic when less than this number of pairs were involved, especially if a good result was due to say only 3 pairs. This often occurred when the distance cutoff was low and the MRSA high.

The value of interest was at what rank at least one of the correct dockings could be found in all the systems. To facilitate comparison between the different systems, the rank was converted to a percentage rank. The size of the lists of possible complexes ranged over an order of magnitude, from 26 to 762, and an absolute rank would not take this into consideration.

## RESULTS

### Screening Unbound Complexes by Pair Potentials

Table I shows all the results using the optimal parameters for each method. The majority of the pair potential methods improve in ranking correct dockings high up the list of complexes compared to the ranking by ftdock. The ftdock rank is from the shape complementarity value given by the ftdock program.[11] Although the ftdock algorithm is good at being able to generate a correct docking in a small list of complexes, it is not so capable of selecting the correct from the incorrect within that list.

The key observations are that the best dataset for any given type of pair potential is intramolecular, and that the mole-fraction method of calculating the expected is better for all types and datasets. The only exception to this is for the all side-chain atoms pair potential with the homodimer data set. It is not clear why this breaks the trend.

Two types of pair potentials performed particularly well; the residue level potential based on all atoms and the residue level potential based on all side-chain atoms. The best result is for the residue level all side-chain atoms type using the intramolecular data set and a mole-fraction calculated expected. The results with the other two data sets and a mole-fraction calculated expected show no clear difference between these two types.

Table I also shows that the MRSA value of the optimal parameters is either 0% or 5%, but never 20%. This may show that using 20% excludes some still useful pairings. However, since there were results with a 20% MRSA parameter (not shown) not significantly worse than the best parameters shown in Table I, it is hard to be sure of making any firm conclusion from this.

Table II shows the absolute rankings for the best results, namely, using the residue level all side-chain atoms potential with a cutoff of 6 Å and an MRSA of 5%. Pair potentials substantially improve on the ranking for all the systems, apart from for PTC, where the rank is only just below. Table II also shows that the pair potentials are better at

**TABLE I. All Results**

| | Percentage Ranks | | | | | |
|---|---|---|---|---|---|---|
| | Heterodimers | | Homodimers | | Intramolecular | |
| Type | Molar | Contact | Molar | Contact | Molar | Contact |
| $C_\beta$–$C_\beta$ | 32.3 | 72.9 | 33.6 | 48.4 | 29.9 | 35.8 |
| Residue all atoms | 22.0 | 43.4 | 27.3 | 46.2 | 16.3 | 30.1 |
| Residue side-chain atoms | 21.3 | 60.7 | 29.8 | 18.9 | 11.8 | 41.9 |
| Atom–atom | — | — | — | — | 38.7 | 53.8 |
| | Parameters | | | | | |
| $C_\beta$–$C_\beta$ | 14 Å, 0% | 11 Å, 0% | 14 Å, 5% | 14 Å, 0% | 6 Å, 5% | 14 Å, 0% |
| Residue all atoms | 10 Å, 0% | 7 Å, 0% | 8 Å, 5% | 4 Å, 5% | 6 Å, 5% | 8 Å, 0% |
| Residue side-chain atoms | 7 Å, 0% | 5 Å, 0% | 5 Å, 0% | 4 Å, 0% | 6 Å, 5% | 10 Å, 0% |
| Atom–atom | — | — | — | — | 7 Å, 0% | 3 Å, 5% |

Compared to above percentages, ftdock puts a correct docking within 43.8% of a ranked list. The parameter values are the distance cut off (in angstroms), and the MRSA (as a percentage).

**TABLE II. Best Results**[†]

| System[a] | | | Rank | | | |
|---|---|---|---|---|---|---|
| | | | At least one correct solution found | | All correct solutions found | |
| | $N$ | $N_{good}$ | ftdock | Pair potentials | ftdock | Pair potentials |
| CGI | 93 | 1 | 3 | 2 | 3 | 2 |
| CHO | 85 | 5 | 11 | 6 | 39 | 23 |
| KAI | 349 | 16 | 128 | 13 | 336 | 131 |
| PTC | 205 | 7 | 12 | 14 | 145 | 49 |
| SNI | 26 | 2 | 8 | 1 | 23 | 4 |
| FDL | 636 | 2 | 149 | 75 | 401 | 89 |
| MLC | 539 | 4 | 34 | 24 | 493 | 182 |
| HFL | 498 | 2 | 218 | 36 | 416 | 153 |
| HFM | 700 | 4 | 48 | 6 | 342 | 220 |

[†]Residue side-chain atoms potential with a 6-Å cutoff and MRSA of 5% using $e_{(\text{mole-fraction})}$.
[a]See the section Test Systems for abbreviations. $N$ is the number of possible dockings generated by ftdock for that system, after biological filtering. $N_{good}$ is the number of those possible dockings that are within 2.5 Å of the correct structure. The parameter values are the distance cutoff (in angstroms), and the MRSA (as a percentage).

ranking all the correct dockings toward the top. The worst rank they produce is still in the top 40% of the list. Compared to this, ftdock puts correct dockings right at the end of the list.

Table III shows where the actual crystal structure is ranked when it is included into the list of possible dockings produced by ftdock. The ranking is that given when using the same method as produced the best results (i.e., all side-chain atoms potential with a cutoff of 6 Å and an MRSA of 5%). Clearly, the rank is not the ideal top rank we would like, and although the ranks for the enzyme–inhibitor systems still seem to be acceptable, the ranks given for the antibody–antigen systems are very bad. This shows that however good is the method to produce a list of possible dockings for the pair potentials to evaluate, is the pair potentials on their own are unlikely to improve much on what they are currently capable of when given results from the present version of ftdock.

Figure 2 shows how the values for the best and all-inclusive ranks vary with the distance cut off parameter (MRSA = 5%). It shows that the method is stable around the optimal parameter for the best method. Therefore, even if the optimal parameters chosen using our tests are

**TABLE III. Ranking of Correct Structure**[†]

| System[a] | $N$[b] | Rank |
|---|---|---|
| CGI | 93 | 5 |
| CHO | 85 | 12 |
| KAI | 349 | 9 |
| PTC | 205 | 5 |
| SNI | 26 | 1 |
| FDL | 636 | 283 |
| MLC | 539 | 259 |
| HFL | 498 | 246 |
| HFM | 700 | 120 |

[†]Residue side-chain atoms potential with a 6-Å cutoff and MRSA of 5% using $e_{(\text{mole-fraction})}$.
[a]See the section Test Systems for abbreviations.
[b]$N$ is the number of possible dockings generated by ftdock for that system, after biological filtering.

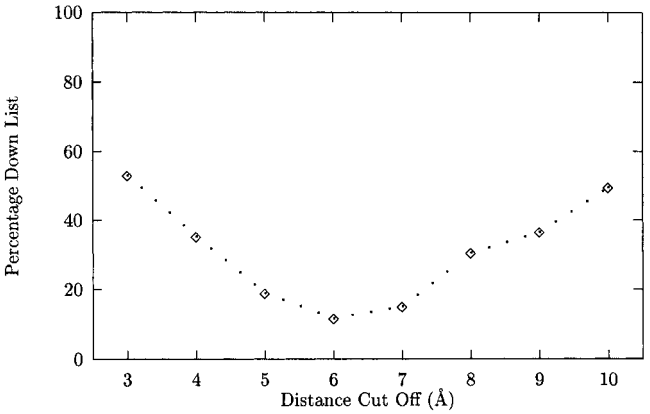


Fig. 2. Stability of results around the minima.

not ideal for another system, they should still produce results which are useful.

### Control-Bound Complexes

To investigate how sensitive the pair potential algorithm is to precise atomic positions, the experiment was repeated using the best method and parameters using the bound forms of the two parts of the system. Table IV shows the results with the bound forms are no better than those of the unbound. Hence the algorithm is clearly able to cope well with the side-chain flexibility that occurs when two unbound proteins dock.

### Combining Algorithms

Pair potentials, although producing false positives, do not rank a correct docking as completely wrong. This is shown by the all-inclusive rank, which is the lowest rank of any correct docking (Table II). By contrast, the scoring functions of ftdock and multidock (see below) produce large all-inclusive ranks that are of little use. The multidock algorithm, also developed in our group, is a simulation that allows for movement of side chains into lower energy states, once a complex is formed. When run on the same dataset as used in this study,[59] the best ranks were comparable to those given by pair potentials, but the all-inclusive rank showed that correct dockings can be placed as completely wrong.

Accordingly, we explored if the two algorithms could be run sequentially to produce a useful combined rank. By cutting the pair potential score-ranked list of complexes at a given percentage down the list, we could still be sure that at least one correct docking was in the list, and so allow for fewer complexes to be evaluated by multidock. We decided to cut the list at 25% down the list, over double the length at which a correct docking was always found. This whole combined process is illustrated in Figure 3.

From the previous work we had the ranks that multidock gave for the datasets used in our tests. Once the lists were cut to the smaller sizes, the multidock ranks were reordered, giving a combined rank. Table V shows all the results together. The ftdock ranks were not used in the combination, and are put there for comparison only.

The combined rank clearly improves beyond either pair potentials or multidock alone. Due to the order in which the algorithms were applied, it is impossible for the combined rank to be worse than that for multidock alone, but it is possible for it to be worse than the rank given by pair potentials alone. This shows in SNI where there is a deterioration in its rank from the pair potential rank. However, overall, the distance down the list of complexes at which all systems have a correct docking is reduced. This is particularly found in the antibody–antigen systems, where there was more scope for improvement. All the systems now have a correct docking within the top 8%, as opposed to the top 12% for pair potentials alone, or top 34% for multidock alone.

This therefore shows that a combined approach to filtering a list of complexes can yield better results than single algorithms alone, and that pair potentials are particularly useful in reducing a list of possible complexes.

### Control-Significance of Results Above Random

In some of the systems studied, ftdock generates several dockings that are correct. It is possible that in a list of a limited size, with a large number of correct solutions to find, that our algorithms for screening a correct docking from the complexes generated by ftdock, though seemingly impressive, were not performing any better than chance. We therefore ran simulations of placing $N_{good}$ solutions in a list of length $N$, to see what the probability was of obtaining our ranks or better. For each system, the computer simulation was run 10,000 times, and used a random number generator.

Table VI shows that for most systems, the probability of obtaining by chance the observed rank or better is small. As the probabilities are independent of each other, the total probability of getting all these ranks or better together is the multiple of the probabilities for each individual system. Thus the success of each algorithm can be expressed as a single number. These values show that all the algorithms are well above random, and there is a clear progression to better values in the order ftdock $\Rightarrow$ multidock $\Rightarrow$ pair potentials $\Rightarrow$ combined method. If for a given algorithm, the probability of getting the rank given for any system was 50/50 (i.e., a coin toss), then the combined probability of success in 9 systems would be $(0.5)^9$, which is 0.002. All of our algorithms do better than this.

### False Positives

Table VII shows the RMS and percentage correct pairs for the top-ranked structure for each method. It can be seen that ftdock does not have any top rank with both less than 6 Å RMS and more than 20% correct pairs. Both multidock and pair potentials have one top rank with an RMS of less than 6 Å and more than 20% correct pairs. The best method by these criteria is the combined method with three such top ranks.

The reasons why these structures are ranked at the top are not consistent. For the four trypsin complexes, at least one residue of the catalytic triad is in the interface. However, compared to any lower ranked incorrect complex, there is no increase in hydrophobic pairs across the

**TABLE IV. Best Bound Results†**

| System[a] | | | Rank | | | |
|---|---|---|---|---|---|---|
| | | | At least one correct solution found | | All correct solutions found | |
| | $N$ | $N_{good}$ | ftdock | Pair potentials | ftdock | Pair potentials |
| CGI | 123 | 11 | 2 | 1 | 59 | 12 |
| CHO | 170 | 9 | 23 | 25 | 138 | 65 |
| KAI | 370 | 18 | 24 | 14 | 287 | 188 |
| PTC | 410 | 7 | 59 | 47 | 309 | 151 |
| SNI | 44 | 11 | 5 | 1 | 42 | 20 |
| FDL | 574 | 1 | 210 | 62 | 210 | 62 |
| MLC | 464 | 5 | 2 | 48 | 215 | 183 |
| HFL | 708 | 5 | 68 | 43 | 299 | 210 |
| HFM | 578 | 1 | 94 | 13 | 94 | 13 |

†Residue side-chain atoms potential with a 6-Å cutoff and MRSA of 5% using $e_{(\text{mole-fraction})}$.
[a]$N$ is the number of possible dockings generated by ftdock for that system, after biological filtering. $N_{good}$ is the number of those possible dockings that are within 2.5 Å of the correct structure. The parameter values are the distance cutoff (in angstroms), and the MRSA (as a percentage).

interface, or salt bridges. Overall, there is no discernible single reason why these false positives are ranked as they are.

### Correlation of Score to Correct Pairs

Looking at all the systems together, we initially examined if there was a correlation between pair-potential rankings and RMS for the complexes being screened. None was observed. However, when we considered the percentage of correct pairs formed by the known complex, which are then found in the complex, correlations were observed (Figures 4, 5, and 6). A pair was considered to exist between two residues if any of the side chain atoms in the two residues were within 6 Å of each other. There is still no discernible correlation with shape complementarity ranks from ftdock. However, pair potentials now show a clear correlation, even when the percentage of correct pairs is as low as 50%. Multidock also shows a correlation, which although not as good overall, is in fact better when the percentage of correct pairs is above 70%. This shows that though pair potentials have a larger radius of convergence than multidock, once a complex is very near to the correct solution, multidock will select it to a higher rank.

These results are consistent with the level of representation used in the modeling. Only surfaces are considered by ftdock, pair potentials are at the residue level, while multidock is at the atomic level. The order in which the two rankings are integrated to produce the combined rank is a consequence of their different correlations and radii of convergence. For protein complexes with limited conformational change on association, this stepwise refinement of molecular representation via residue pair potentials provides a useful strategy to predict docked protein complexes.

## DISCUSSION AND CONCLUSION

The key conclusion of this study is that pair potentials have considerable power in correctly selecting correct

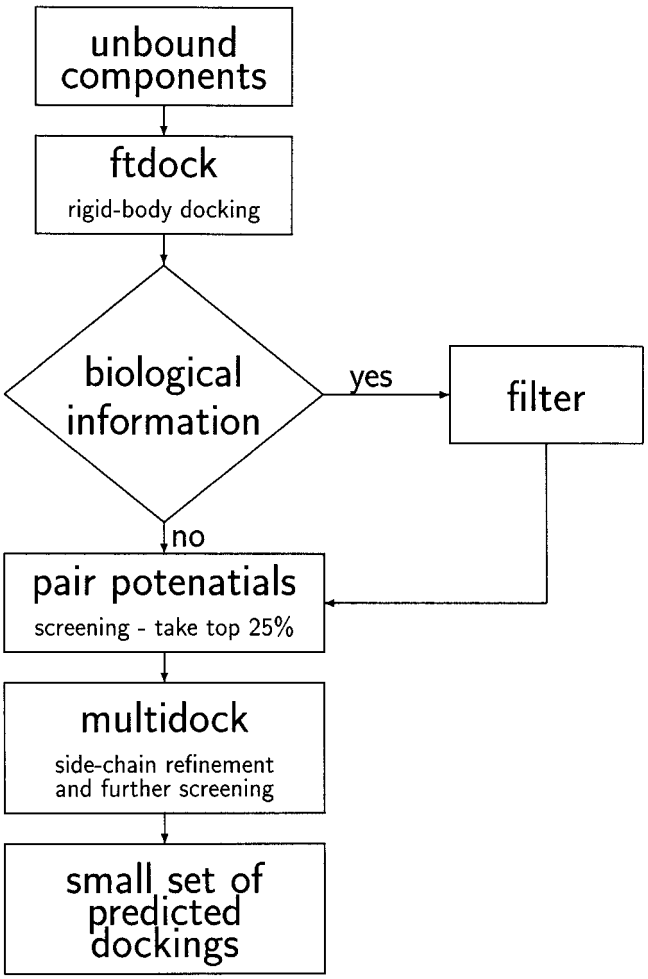


Fig. 3. Flow chart summarizing the combined methods.

**TABLE V. Best Combined Results†**

| System[a] | | | Rank at which at least onecorrect solution found | | | Combined | |
|---|---|---|---|---|---|---|---|
| | $N$ | $N_{good}$ | ftdock | Pair potentials | Multidock | $N_{good}$ | Rank |
| CGI | 93 | 1 | 3 | 2 | 2 | 1 | 2 |
| CHO | 85 | 5 | 11 | 6 | 1 | 4 | 1 |
| KAI | 349 | 16 | 128 | 13 | 2 | 12 | 2 |
| PTC | 205 | 7 | 12 | 14 | 23 | 7 | 3 |
| SNI | 26 | 2 | 8 | 1 | 12 | 2 | 2 |
| FDL | 636 | 2 | 149 | 75 | 211 | 2 | 38 |
| MLC | 539 | 4 | 34 | 24 | 101 | 2 | 21 |
| HFL | 498 | 2 | 218 | 36 | 29 | 1 | 29 |
| HFM | 700 | 4 | 48 | 6 | 9 | 2 | 2 |

†Residue side-chain atoms potential with a 6-Å cutoff and MRSA of 5% using $e_{(\text{mole-fraction})}$.
[a]$N$ is the number of possible dockings generated by ftdock for that system, after biological filtering. $N_{good}$ is the number of those possible dockings that are within 2.5 Å of the correct structure. The parameter values are the distance cutoff (in angstroms), and the MRSA (as a percentage).

**TABLE VI. Probabilities Showing Significance of Results**

| System* | | | Probability of getting rank | | | |
|---|---|---|---|---|---|---|
| | $N$ | $N_{good}$ | ftdock | Pair potentials | Multidock | Combined |
| CGI | 93 | 1 | 0.031 | 0.023 | 0.023 | 0.023 |
| CHO | 85 | 5 | 0.504 | 0.311 | 0.058 | 0.058 |
| KAI | 349 | 16 | 0.999 | 0.466 | 0.092 | 0.092 |
| PTC | 205 | 7 | 0.352 | 0.401 | 0.579 | 0.099 |
| SNI | 26 | 2 | 0.530 | 0.082 | 0.724 | 0.159 |
| FDL | 636 | 2 | 0.417 | 0.218 | 0.554 | 0.114 |
| MLC | 539 | 4 | 0.230 | 0.166 | 0.567 | 0.144 |
| HFL | 498 | 2 | 0.690 | 0.144 | 0.116 | 0.116 |
| HFM | 700 | 4 | 0.256 | 0.034 | 0.051 | 0.010 |
| Total probabilities | | | $5 \cdot 10^{-5}$ | $2 \cdot 10^{-8}$ | $1 \cdot 10^{-7}$ | $4 \cdot 10^{-11}$ |

*$N$ is the number of possible dockings generated by ftdock for that system, after biological filtering. $N_{good}$ is the number of those possible dockings that are within 2.5 Å of the correct structure. The parameter values are the distance cutoff (in angstroms), and the MRSA (as a percentage).

**TABLE VII. False Positives: RMS and PCP for Top Ranks†**

| | ftdock | | Multidock | | Pair potentials | | Combined | |
|---|---|---|---|---|---|---|---|---|
| System* | RMS | PCP | RMS | PCP | RMS | PCP | RMS | PCP |
| CGI | 11.61 | 0.0 | 6.05 | 20.0 | 7.96 | 0.0 | 6.05 | 20.0 |
| CHO | 8.29 | 17.4 | 1.52 | 73.9 | 6.20 | 21.7 | 1.52 | 73.9 |
| KAI | 7.13 | 11.1 | 4.85 | 5.6 | 6.28 | 5.6 | 4.85 | 5.6 |
| PTC | 5.98 | 10.5 | 6.57 | 5.3 | 7.79 | 0.0 | 5.00 | 21.1 |
| SNI | 5.98 | 4.8 | 7.52 | 9.5 | 1.56 | 28.6 | 8.49 | 4.8 |
| FDL | 8.59 | 0.0 | 12.96 | 0.0 | 9.46 | 11.1 | 4.68 | 22.2 |
| MLC | 12.20 | 0.0 | 9.28 | 21.1 | 9.86 | 10.5 | 10.04 | 0.0 |
| HFL | 10.68 | 15.4 | 10.27 | 0.0 | 12.92 | 0.0 | 10.27 | 0.0 |
| HFM | 17.94 | 0.0 | 17.70 | 3.2 | 10.09 | 0.0 | 13.53 | 3.2 |

†Pairs considered up to distance of 6 Å pair potential is residue side-chain atoms potential with a 6-Å cutoff and MRSA of 5% using $e_{(\text{mole-fraction})}$.
*RMS, root-mean-square distance from crystal structure; PCP, percentage correct pairs compared to crystal structure.

dockings from a list of complexes. Although the scoring function also produces false-positives, pair potentials can position a correctly docked complex at or near the top of a score-ranked list.

Our study also shows that of the various models proposed for the random state required in generating pair potentials, it is the mole-fraction method which should be used when using pair potentials across an interface. However, the best strategy for docking need not necessarily apply to the use of pair potentials within a single domain, as is done for threading.

Of the various datasets from which we generated pair potentials, it is clear that the best results came from the intramolecular pairings in a database of nonhomologous protein domains. Why this is so is not clear. It may be that pair potentials generated from intermolecular pairings will ultimately be a better method, and the reason they were not so here was due to the small size of the datasets available. This is particularly true of the heterodimer dataset, which should most closely represent the propensities of intermolecular pairings in complexes, yet at the moment was limited to 11 structures.

The results for the enzyme–inhibitor systems are clearly better than those for antibody–antigen systems. This problem is not unique to pair potentials, and as was discussed above is possibly due to the differing recognition stereochemistry and lower affinity of the antibody–antigen interactions, compared to those for enzyme-inhibitors.[36] However, the results even for the antibody–antigen systems are still significantly more useful than relying only on the shape complementarity as calculated by ftdock.

Finally, we have shown that as part of a combined approach, the large radius of convergence makes pair potentials useful in screening large numbers of structures before more detailed all-atom refinement procedures are used.

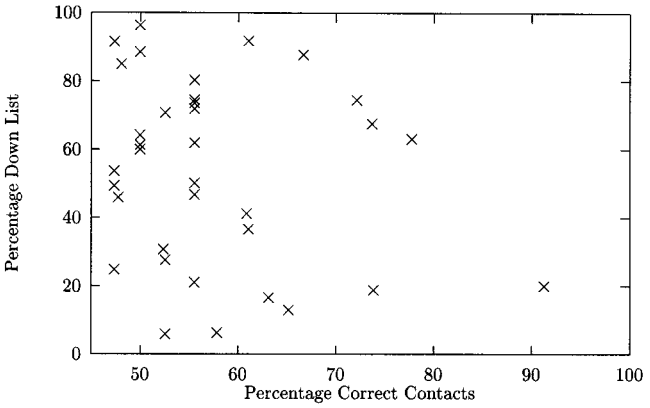


Fig. 4. Correlation of ftdock ranks to percentage correct contacts.

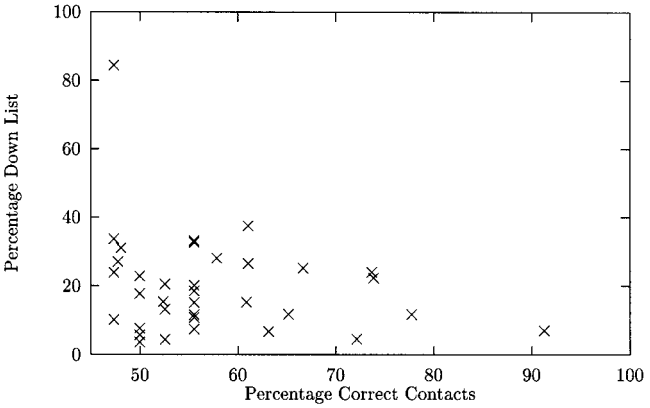


Fig. 5. Correlation of pair potential ranks to percentage correct contacts.

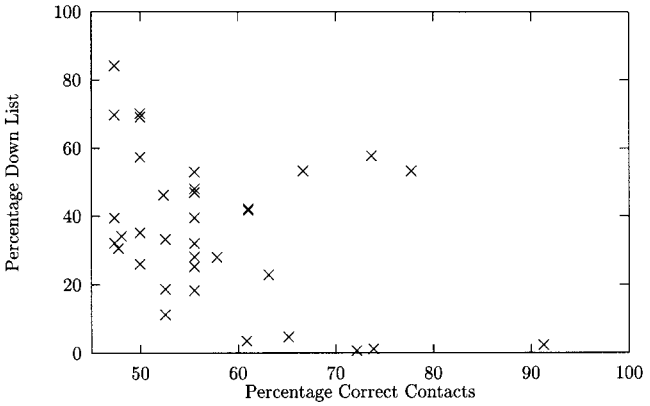


Fig. 6. Correlation of multidock ranks to percentage correct contacts.

## ACKNOWLEDGMENTS

H.A. Gabb was supported by a long-term EMBO fellowship. Thanks to R.M. Jackson (ICRF) for discussion and help with multidock.

## REFERENCES

1. Chothia C, Janin J. Principles of protein–protein recognition. Nature 1975;256:705–708.
2. Dill KA. Dominant forces in protein folding. Biochemistry 1990;29: 7133–7155.
3. Janin J. Protein–protein recognition. Prog Biophys Mol Biol 1995;64:145–166.
4. Sippl MJ. Calculation of conformational ensembles from potentials of mean force: an approach to the knowledge-based prediction of local structures in globular proteins. J Mol Biol 1990;213: 859–883.
5. Godzik A, Kolinski A, Skolnick J. Toplogy fingerprint approach to the inverse protein folding problem. J Mol Biol 1992;227:227–238.
6. Jones DT, Taylor WR, Thornton JM. A new approach to protein fold recognition. Nature 1992;358:86–89.
7. Bryant SH, Lawrence CE. An empirical energy function for threading a protein sequence through the folding motif. Proteins 1993;16:92–112.
8. Jones DT. A new approach to protein fold recognition. Curr Opin Struct Biol 1997;7:377–387.
9. Torda AE. Perspectives in protein-fold recognition. Curr Opin Struct Biol 1997;7:200–205.
10. Vajda S, Sippl M, Novotny J. Empirical potentials and functions for protein folding and binding. Curr Opin Struct Biol 1997;7:222–228.
11. Gabb HA, Jackson RM, Sternberg MJE. Modelling protein docking using shape complementarity, electrostatics, and biochemical information. J Mol Biol 1997;272:106–120.
12. Shoichet BK, Kuntz ID. Protein docking and complementarity. J Mol Biol 1991;221:327–346.
13. Sternberg MJE, Gabb HA, Jackson MJ. Predictive docking of protein–protein and protein–dna complexes. Curr Opin Struct Biol 1998;8:250–256.
14. Bernstein FC, Koetzle TF, Williams GJB, Meyer Jr EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The protein data bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
15. Abola EE, Bernstein FC, Bryant SH, Koetzle TF, Weng J. Protein data bank. In: Allen FH, Bergerhoff G, Sievers R, editors. Crystallographic Databases-Information Content, Software Systems, Scientific Applications. Bonn: Data Commission of the International Union of Crystallography, 1987;107–132.
16. Cherfils J, Duquerroy S, Janin J. Protein–protein recognition analyzed by docking simulation. Proteins 1991;11:271–280.
17. Jiang F, Kim S. Soft docking: matching of molecular surface cubes. J Mol Biol 1991;219:79–102.
18. Walls PH, Sternberg MJE. New algorithm to model protein–protein recognition based on surface complementarity. J Mol Biol 1992;228:277–297.
19. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Wodak SJ. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc Natl Acad Sci USA 1992;89:2195–2199.
20. Totrov M, Abagyan R. Detailed ab initio prediction of lysozyme–antibody complex with 1.6 Å accuracy. Nature Struct Biol 1994;1: 259–263.
21. Norel R, Lin SL, Wolfson HL, Nussinov R. Molecular surface complementarity at protein–protein interfaces: the critical role played by surface normals at well placed sparse points in docking. J Mol Biol 1995;252:263–273.
22. Peters KP, Fauck J, Frommel C. The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. J Mol Biol 1996;256:201–213.
23. Cummings M, Hart T, Read R. Atomic solvation parameters in the analysis of protein–protein docking results. Protein Sci 1995;4: 2087–2099.
24. Jackson RM, Sternberg MJE. A continuum model for protein–protein interactions: application to the docking problem. J Mol Biol 1995;250:258–275.
25. Wallqvist A, Covell DG. Docking enzyme–inhibitor complexes using a preference-based free-energy surface. Proteins 1996;25: 403–419.
26. Weng ZP, Vajda S, Delisi C. Prediction of protein complexes using empirical free-energy functions. Protein Sci 1996;5:614–626.
27. Tanaka S, Scheraga HA. Medium and long range interaction

parameters between amino acids for predicting three dimensional structures of proteins. Macromolecules 1976;9:945–950.
28. Miyazawa S, Jernigan RL. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules 1985;18:534–552.
29. Huang ES, Subbiah S, Levitt M. Recognizing native folds by the arrangement of hydrophobic and polar residues. J Mol Biol 1995;252:709–720.
30. Miyazawa S, Jernigan RL. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 1996;256: 623–644.
31. Park BH, Levitt M. The complexity and accuracy of discrete state models of protein structure. J Mol Biol 1995;249:493–507.
32. Hinds DA, Levitt M. Exploring conformational space with a simple lattice model for protein structure. J Mol Biol 1994;243:668–682.
33. Godzik A, Kolinski A, Skolnick J. Lattice representations of globular prtoeins: how good are they? J Comp Chem 1993;14:1194–1202.
34. Thomas PD, Dill KA. Statistical potentials extracted from protein structures: how accurate are they? J Mol Biol 1996;257:457–469.
35. Skolnick J, Jaroszewski L, Kolinski A, Godzik A. Derivation and testing of pair potentials for protein folding: when is the quasichemical approximation correct? Protein Sci 1997;6:676–688.
36. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. J Mol Biol 1993;234:946–950.
37. Jones S, Thornton JM. Prediction of protein–protein interaction sites using patch analysis. J Mol Biol 1997;272:133–143.
38. Murzin A, Brenner SE, Hubbard T, Chothia C. scop: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
39. Melo F, Feytmans E. Novel knowledge-based mean force potential at atomic level. J Mol Biol 1997;267:207–222.
40. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.
41. Hecht HJ, Szardenings M, Collins J, Schomburg D. Three-dimensional structure of a recombinant variant of human pancreatic secretory trypsin inhibitor (Kazal type). J Mol Biol 1992;225: 1095–1103.
42. Freer ST, Kraut J, Robertus JD, Wright HT, Xuong NH. Chymotrypsinogen, 2.5 angstroms crystal structure, comparison with alpha-chymotrypsin, and implications for zymogen activation. Biochemistry 1970;9:1997–2009.
43. Hecht HJ, Szardenings M, Collins J, Schomburg D. Three-dimensional structure of the complexes between bovine chymotrypsinogen a and two recombinant variants of human pancreatic secretory trypsin inhibitor (Kazal type). J Mol Biol 1991;220:711–722.
44. Bode W, Epp O, Huber R, Laskowski Jr M, Ardelt W. The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3). Eur J Biochem 1985;147:387–395.
45. Blevins RA, Tulinsky A. The refinement and the structure of the dimer of alpha-chymotrypsin at 1.67 Å resolution. J Biol Chem 1985;260:4264–4275.
46. Fujinaga M, Sielecki AR, Read RJ, Ardelt W, Laskowski Jr M, James MNG. Crystal and molecular structures of the complex of alpha-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution. J Mol Biol 1987;195:397–418.
47. Bode W, Chen Z, Bartels K, Kutzbach C, Schmidt-Kastner G, Bartunik H. Refined 2 angstroms x-ray crystal structure of porcine pancreatic kallikrein a and a specific trypsin-like serine proteinase. crystallization and structure determination and crystallographic refinement and structure and its comparison with bovine trypsin. J Mol Biol 1983;164:237–282.
48. Chen Z, Bode W. Refined 2.5 Å x-ray crystal structure of the complex formed by porcine kallikrein a and the crystallization, patterson search, structure comparison with its components and with the bovine trypsin-pancreatic trypsin inhibitor complex. J Mol Biol 1983;164:283–311.
49. Wlodawer A, Deisenhofer J, Huber R. Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. J Mol Biol 1987;193:145–156.
50. Fehlhemmer H, Bode W, Huber R. Crystal structure of bovine trypsinogen at 1.8å resolution. II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin. J Mol Biol 1977;111:415–438.
51. Marquart M, Walter J, Deisenhofer J, Bode W, Huber R. The geometry of the reactive site and of the peptide groups in trypsin and trypsinogen and its complexes with inhibitors. Acta Crystallogr 1983;39:480.
52. McPhalen CA, James MNG. Crystal and molecular structure of the serine proteinase inhibitor ci-2 from barley seeds. Biochemistry 1987;26:261–269.
53. McPhalen CA, James MNG. Structural comparison of two serine proteinase–protein inhibitor complexes. eglin-c-subtilisin Carlsberg and CI-2-subtilisin novo. Biochemistry 1988;27:6582–6589.
54. Maenaka K, Matsushima M, Song H, Sunada F, Watanabe K, Kumagai I. Dissection of protein–carbohydrate interactions in mutant hen egg-white lysozyme complexes and their hydrolytic activity. J Mol Biol 1995;247:281–293.
55. Fischmann TO, Bentley GA, Bhat TN, Boulot G, Mariuzza RA, Phillips SEV, Tello D, Poljak RJ. Crystallographic refinement of the three-dimensional structure of the D1.3-lysozyme complex at 2.5Å resolution. J Biol Chem 1991;266:12915–12920.
56. Braden BC, Souchon H, Eisele JL, Bentley GA, Bhat TN, Navaza J, Poljak RJ. Three-dimensional structures of the free and the antigen-complexed Fab form monoclonal anti-lysozyme antibody D44.1. J Mol Biol 1994;243:767–781.
57. Sheriff S, Silverton EW, Padlan EA, Cohen GH, Smith-Gill SJ, Davies DR. Three-dimensional structure of an antibody-antigen complex. Proc Natl Acad Sci USA 1987;84:8075–8079.
58. Padlan EA, Silverton EW, Sheriff S, Cohen GH, Smith-Gill SJ, Davies DR. Structure of an antibody–antigen complex. crystal structure of the HyHEL-10 Fab-lysozyme complex. Proc Natl Acad Sci USA 1989;86:5938–5942.
59. Jackson RM, Gabb HA, Sternberg MJE. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. J Mol Biol 1998;276:265–285.