



[Artificial Receptors](#)

[Flexible Sidechains](#)

[GOLD SAMPL](#)

[GOLD Validation:
Virtual Screening](#)

[Purchasing Information](#)

Optimum Docking and Rescoring Protocols for Structure-based Virtual Screening

Introduction

It is desirable, when beginning a structure-based virtual screening campaign, to have available a docking and analysis protocol that appropriately trades speed with efficiency at finding hits, specific for the case in hand. However it may not be immediately clear what that best protocol is. In GOLD for instance the user has available three scoring functions for docking (GoldScore, ChemScore and the Astex Statistical Potential (ASP)) and a variety of settings controlling the length of docking (No. of GA attempts, No of docking operations etc). In addition GOLD has four scoring functions which can be used for rescoring an existing set of poses (GoldScore, Chemscore, ChemScore with Receptor Density Scaling, and ASP). A docking and rescoring strategy may turn out to be significantly more effective than docking on its own, yet it is then important to know which is the right combination of docking and scoring function to use. The most effective protocol will almost certainly be different for different protein types used and possibly even depend on the make up of the library to be screened.

We hope here to provide some guidelines as to the type of protocols that might be of use. [Previous work](#) had suggested the length of docking protocol that was appropriate for High Throughput Virtual Screening (HTVS). Here we present a more thorough investigation of protocol length and suggest new guidelines for fast yet efficient protocols. In addition we look at using rescoring protocols to see if they can significantly enhance enrichment.

Two protein models of a single target, factor Xa are investigated in this study, (PDB codes *1ezq* and *2bok*), and only one Decoy/Active set (the factor Xa DUD set of [Irwin](#)). Because only one protein target is considered there is no guarantee that the optimum protocols described here will work on other systems, though they are quite likely to be applicable to other serine protease targets. However the active site of factor Xa is large and the active ligands tend to be large and flexible i.e. harder to dock successfully, so the optimum docking protocols found for factor Xa, are likely to be, if anything, more lengthy than required for many other systems. In this sense factor Xa might be considered a 'hard' test case. Nevertheless similar work on other protein systems is necessary. This is ongoing and will be reported on at a later date.

GOLD Docking and rescoring Protocols

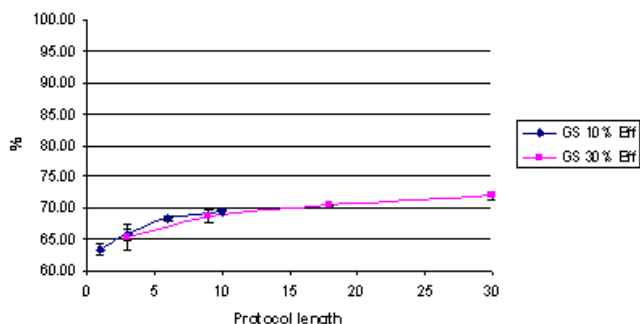
The factor Xa model used for the first part of this study was the structure with PDB code *1ezq*, which was protonated according to physiological pH using an in-house methodology (currently available with GOLD 4.x) but otherwise unmodified. The active site was defined as a 14 Å sphere around the NH atom Gly 216. Default settings were used for all docking parameters other than number of GA attempts and number of operations. These two parameters were systematically varied. The previous work had suggested that using Auto settings at 10% efficiency was the shortest docking protocol that should be used for a serine protease target. Therefore we used this number of operations (10% efficiency; minimum number of operations, 1000; maximum number of operations; 12500). We also used a longer protocol (30% efficiency; minimum number of operations, 3000; maximum, 37500). This we feel is the longest protocol that could be accommodated for very large virtual screens (>100,000 structures) without massive computational resource. The other variable changed was the number of GA attempts carried out per ligand (i.e. number of poses generated for each ligand). This varied as 1, 3, 6 and 10. All attempts were saved, not just the best. This was necessary for the rescoring part of the study. Rescoring protocols are likely to be most effective when a variety of reasonable poses is saved. The top pose out of the docking may be over-scored if there is a deficiency in the binding interaction not picked up by the docking scoring function. A rescoring function which picks up this deficiency may instead correctly place top a pose that is not ranked highest by the docking scoring function.

All three scoring functions were initially used for the docking part of the study. However it was found that the ASP scoring function performed poorly at generating good enrichments with this test set. These results are shown later. Consequently a full analysis of docking protocol was only carried out for GoldScore and ChemScore. Each docking run was repeated three times and results averaged. In addition each run was rescored with three different scoring functions, ASP, ChemScore and ChemScore with Receptor Density scaling. A significant difference for ChemScore RDS, was that Simplex optimisation of the docked pose was **not** allowed during rescore. It was thought that Simplex minimisation might lead to poses with unrealistically buried portions. Rescoring ChemScore dockings with GoldScore was also tried in one or two cases but poor results were obtained and so a full analysis was not carried out. Typical results are however tabulated later.

Analysis

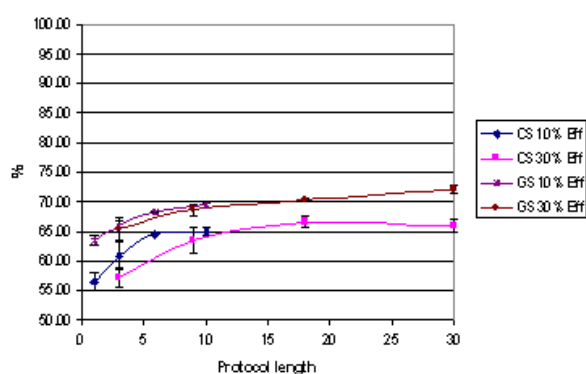
Three different metrics were used for measuring enrichment. The primary measure was the area under the Receiver Operator Curve (ROC). In addition enrichment factors (= No. of actives retrieved/no. actives expected assuming random pick) were calculated at both the 1%, and the 10% cut-offs of the ranked dataset. These results are graphed (Graphs 1-10). The Y axis in each graph is one of the three enrichment metrics. The X axis in each graph represents a measure of protocol length. This is calculated for a given protocol by multiplying the % search efficiency by the number of GA attempts and dividing by 10. This allows all protocols to be placed on the same graph. Strictly speaking this is only valid if a 30% protocol using one GA attempt is as time consuming computationally as a 10% protocol using 3 GA attempts. This however will be approximately true.

Enrichment when docking with GoldScore or ChemScore



Graph 1. AUC under ROC values against protocol length for GoldScore dockings.

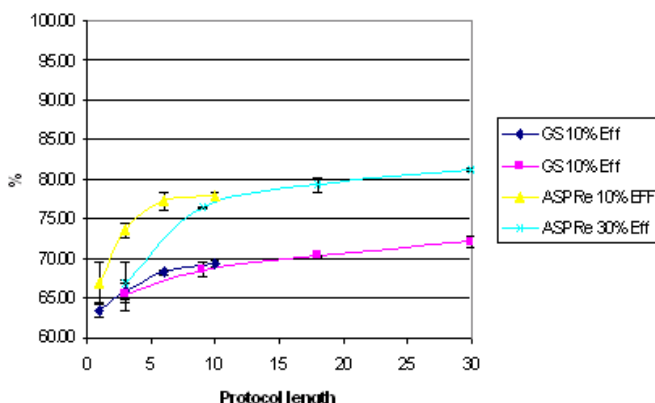
Two conclusions are apparent from looking at the way the enrichment metric varies with length of protocol. Firstly, although longer protocols always appear to lead to high enrichment, performance is seen to plateau off once the Protocol length rises above 9. A docking protocol which balances speed and accuracy will lie just beyond the apex of the curve at the start of the plateau. This corresponds to a 10% Eff. protocol with 10 GAs or a 30% Eff. protocol with 3GAs. Secondly, the two curves on each graph are almost coincident within experimental error. So, for this case, a 3X increase in search efficiency does appear to be equivalent, in terms of performance, as a 3X increase in the number of GAs run.



Graph 2. AUC under ROC values against protocol length for ChemScore Dockings (GoldScore included for comparison).

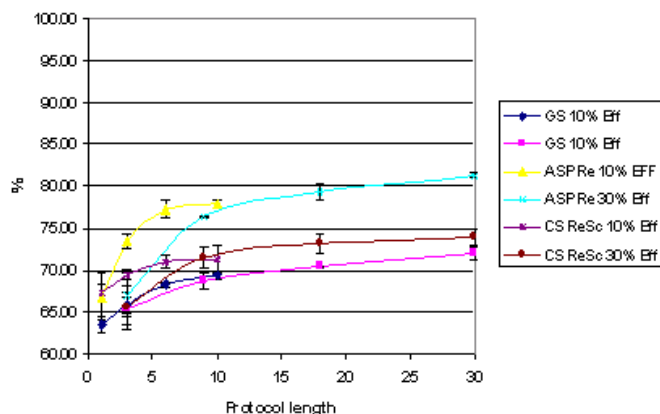
ChemScore enrichments do not appear to be as good as those for GoldScore. A near plateau in the AUC is again reached for both curves. Interestingly however the 10% efficiency protocols plateau more quickly than the 30% protocol. So a 6 GA 10% protocol is now equivalent to a 3GA 30% protocol, despite the fact the former is probably significantly faster.

Rescoring GoldScore Poses with ASP, ChemScore and ChemScore RDS



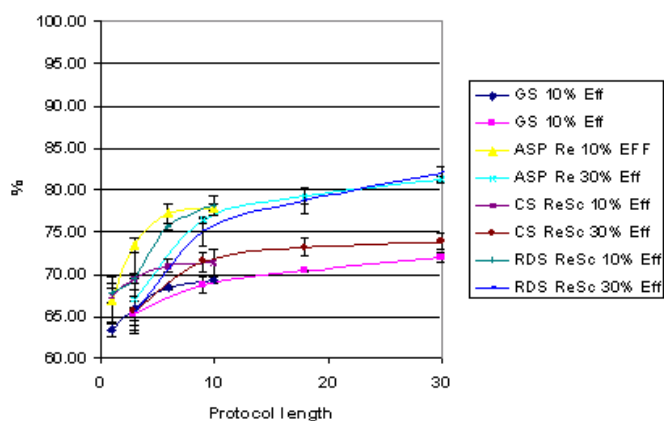
Graph 3. AUC under ROC values against protocol length for GoldScore dockings rescored with ASP.

The ASP rescore results are graphed alongside the original GoldScore data. A very clear improvement in enrichment is generally seen at all protocol lengths, if the ASP score for each rescored pose is used to re-rank the dataset. AUC under ROC of greater than 80% are now achievable. The graph for the longer protocol continues to rise even at the right hand of the graph. Nevertheless the rise is only modest after a protocol length of 12 has been reached and so a protocol of 3 or 4 GA attempts at 30% efficiency can be considered the best trade-off protocol. It is very interesting that the shorter protocol levels out much earlier and a protocol of 6GAs at 10% efficiency may be sufficient to get good results.



Graph 4. AUC under ROC values against protocol length for GoldScore dockings rescored with ChemScore.

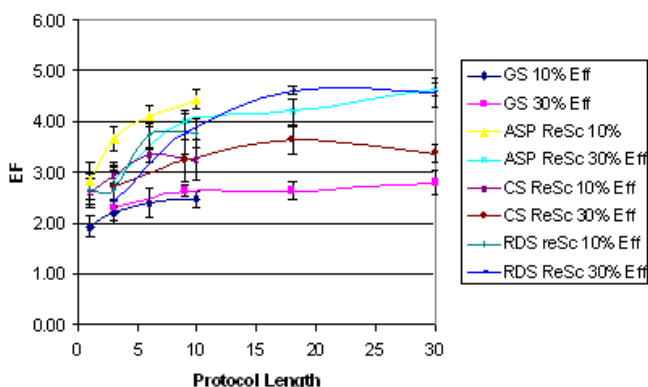
The ChemScore rescore results are added to the previous graph. Better enrichment is again observed than with the original GoldScore scores. It is worth pointing out that this is despite the fact that use of ChemScore for docking leads to poorer enrichments than GoldScore (Graph 2). Rescoring with ChemScore is not however as effective as rescoring with ASP, for this system.



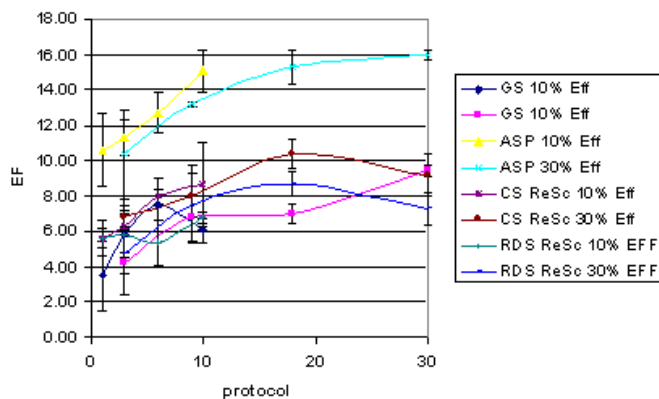
Graph 5. AUC under ROC values against protocol length for GoldScore dockings rescored with ChemScore with Receptor Density Scaling (RDS).

ChemScore with Receptor Density Scaling is a modified version of ChemScore which better rewards those interactions made deep within subpockets of the enzyme active site. The ChemScore RDS results have been added to the previous graph. This time enrichments comparable to those achieved with ASP rescores are achieved. Indeed there is some evidence that if the protocols were extended beyond the longest used here, this would become the preferred protocol for this system. A protocol of 10GAs at 10% or 4GAs at 30% are the best trade-off protocols. Factor Xa is characteristic of serine proteases in that it has two major subpockets (S1 and S4) which are invariably filled by active inhibitors. Therefore rewarding these interactions more strongly, as ChemScore RDS does, might well lead to better enrichment performance against this class of proteins.

GoldScore Docking and Rescoring: Enrichment factors



Graph 6. Enrichment Factors calculated for the top 10% of the dataset for GoldScore docking and rescoring runs.

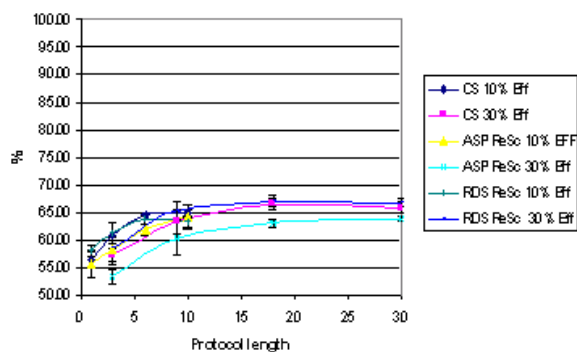


Graph 7. Enrichment Factors calculated for the top 1% of the dataset for GoldScore docking and rescoring runs.

Graphs 6 and 7 show Enrichment Factors calculated for the top 10% and 1% of the ranked dataset, for all GoldScore docked datasets, graphed against protocol length. The first thing to observe is that the error on individual EF values is much greater than that shown for AUC under ROC. This is one reason why these metrics are not considered the best metrics for reporting enrichment. However they do closely represent the situation that pertains to a practical virtual screen. It would be quite usual to select the top 1% of a ranked dataset for biological screening, so the 1% EF in particular is a useful indicator of early enrichment performance in this regard.

The trends observed for the EF graphs follow reasonably those of AUC under ROC. Rescoring usually improves the enrichment over that seen for the original dockings. In addition the rescoring protocols appear to rank in performance in a similar way. Thus ASP rescoring is the best rescoring protocol, followed by ChemScore RDS and ChemScore. One significant difference that should be noted is that the ASP rescoring protocol very significantly outperforms the ChemScore RDS protocol for 1% EF. The two protocols were equivalent using the AUC under ROC metric. We will return to this observation later.

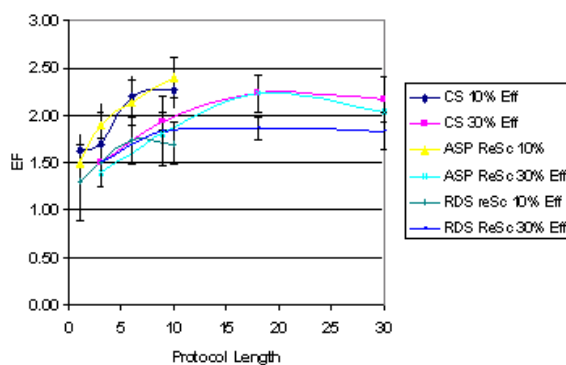
Rescoring ChemScore Poses with ASP and ChemScore RDS



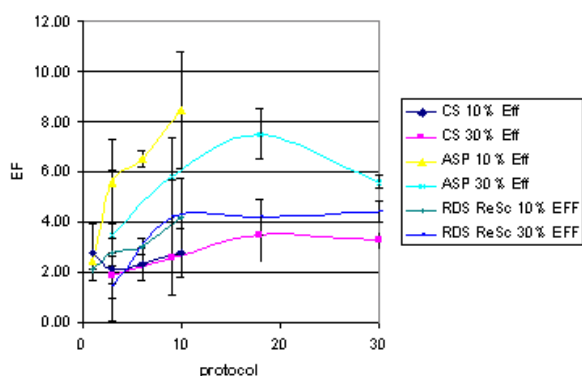
Graph 8. AUC under ROC values against protocol length for ChemScore dockings rescored with ASP and ChemScore RDS.

Rescoring with either scoring function leads to no significant benefit, compared to the original ChemScore dockings. Rescoring with GoldScore was also tried but with an even poorer outcome (representative data is presented later).

ChemScore Docking and Rescoring: Enrichment factors



Graph 9. Enrichment factors calculated for the top 10% of the dataset for ChemScore docking and rescoring runs.



Graph 10. Enrichment Factors calculated for the top 1% of the dataset for ChemScore docking and rescoring runs.

Errors are again high for the individual enrichment factors. No rescoring protocol is clearly better than any other although the ASP rescoring protocol provides the highest enrichment factors at the 1% level.

Rescoring: The Effect of Simplex Optimisation

	Enrich. Factor 10%	Enrich. Factor 1%	AUC under ROC (%)
ASP Rescore without Simplex Opt.	4.4	12.0	77.9
ASP Rescore with Simplex Opt.	4.8	17.6	81.5
CS RDS Rescore without Simplex Opt.	4.6	6.3	81.9
CS RDS Rescore with Simplex Opt.	6.1	16.2	86.7

Table 1. Rescoring a single set of GoldScore docked poses (10 GA, 30% Efficiency) with and without Simplex Optimisation.

Rescoring of docked poses can be carried out either with or without Simplex optimisation of the pose geometry to maximise the value of the rescore function. Currently (GOLD 4.0.1) Simplex optimisation is set as the default for rescoring with ASP and ChemScore. Table one gives the results of equivalent rescoring experiments on a single set of GoldScore generated poses using the ASP and ChemScore RDS scoring functions. The result is clear cut. Use of Simplex optimisation significantly enhances the AUC under ROC and, notably, the 1% enrichment factor for both scoring functions. It was previously noticed that ASP rescore returned better 1% EF than CS RDS (Graph 7). However now we see this is due to the fact that ASP was utilising Simplex optimisation (SO) and ChemScore RDS was not. Rescoring with ChemScore RDS including SO appears, in the light of the results in Table 1, to be an optimum protocol for this protein system. It is perhaps not surprising that allowing the ligand pose to relax whilst rescoring, should lead to better enrichment results and it is therefore a strength of the rescore methodology in GOLD that this can be done. Further experimentation will be carried out on other systems to prove the utility of SO with RDS.

Why is an Effective Rescoring Function not Effective for Pose Generation?

	Enrich. Factor 10%	Enrich. Factor 1%	AUC under ROC (%)
GoldScore Dock (10GA, 30% Eff.)	3.1	9.2	72.7
ASP Rescore	4.8	17.6	81.5
CS Rescore	4.2	9.2	75.7
ASP Dock (10GA 30% Eff.)	1.8	4.2	41.7
GS Rescore	1.5	3.5	57.1
CS Docking (10% GA 30% Eff.)	2.0	2.8	65.3
GS Rescore	1.3	1.4	58.0

Table 2. Comparison of docking vs. rescoring for GoldScore, ChemScore and ASP scoring functions

Table 2 compares the scoring functions GoldScore, ChemScore and ASP, for their ability to identify actives, either via docking or via rescoring. In each case the longest docking protocol was applied (10GA, 30% Efficiency). The rescore runs are carried out on the poses arising from the parent docking run. As previously demonstrated, ASP and ChemScore rescore protocols applied to the GoldScore derived docking poses, give better enrichment than the GoldScore ranked poses alone. However ChemScore and, more strikingly, ASP, both perform less well than GoldScore when used as the docking scoring function. Moreover, in neither case does rescoring the ChemScore or ASP poses with GoldScore lead to good enrichment metrics. The most likely conjecture to arise from this is that, for this system, the best scoring function for finding good quality i.e. realistic, binding poses, is not the best scoring function for distinguishing the binding of inactives and actives. GoldScore/ASP and GoldScore/ChemScore or ChemScore RDS work well as docking/rescoring protocols because GoldScore is able to identify reasonable binding poses and ASP/ChemScore (RDS) are able to rank these poses highly against poses of inactive molecules. The reverse is not true however. Neither ASP or ChemScore are able to identify realistic binding poses as well and GoldScore is not well suited to accurately rank as high those active poses identified by ASP or ChemScore

It must be stressed that this result should not be considered general. A scoring function other than GoldScore might generate good binding poses for other targets and ASP may be less appropriate as a rescoring function in some cases. Further investigation is ongoing to identify best docking/rescoring protocols for other target types.

Docking and Rescoring against another Factor Xa Model

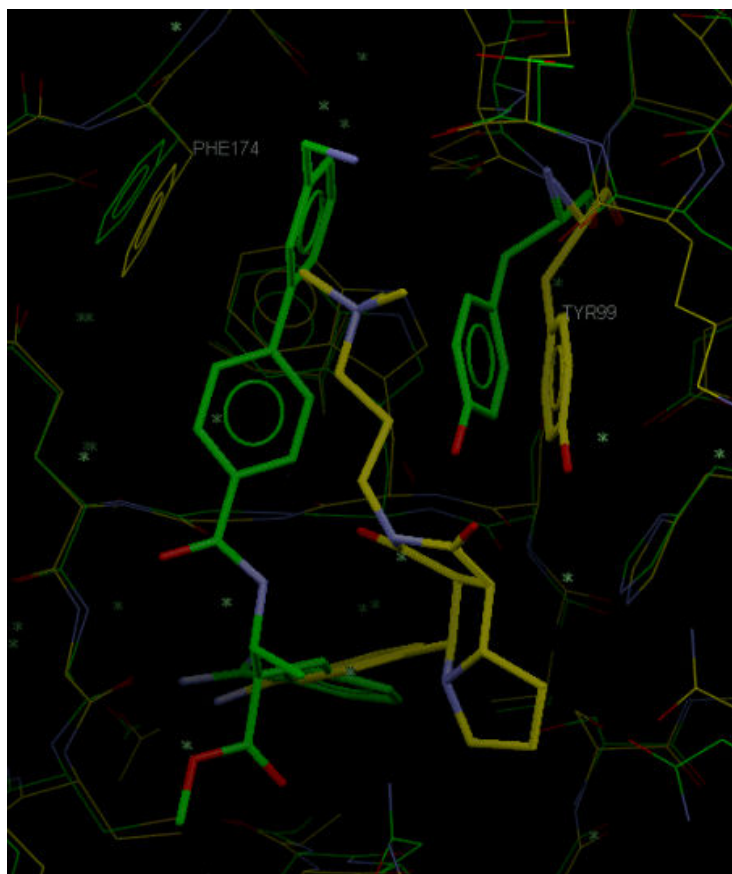


Figure 1. Comparison of *1ezq* and *2bok* active sites. *2bok* is in yellow, *1ezq* in green. The Tyr99 residue at the side of the S4 pocket is highlighted in both structures.

The second model examined here, *2bok*, differs from *1ezq* in only one important aspect. The S4 pocket, a large electron rich box which normally accepts both lipophilic groups and cations, is more constrained in *1ezq* than *2bok*. Therefore actives with large groups in S4 may be penalised by the scoring function in *1ezq*. Conversely ligands which have small groups in S4 may not be so well rewarded in *2bok*. A superposition of the relevant portions of both active sites is shown above.

	Enrich. Factor 10% (Max = 10)	Enrich. Factor 1% (Max = 36)	AUC under ROC (%)
GoldScore Dock (10GA, 30% Eff.)	5.4	13.4	82.0
ASP Rescore	7.9	21.8	90.3
CS Rescore	6.7	19.0	87.3
CS RDS Rescore (Simp.)	8.2	28.2	93.4
ASP Dock (10 GA, 30% Eff.)	3.6	6.3	72.1
GS Rescore	2.8	6.3	65.1
CS Rescore	4.0	12.0	75.8
CS RDS Rescore (Simp.)	4.8	14.1	79.7
CS Dock (10 GA, 30% Eff.)	2.8	7.0	70.8
GS Rescore	3.7	9.9	71.1
ASP Rescore	3.7	14.1	71.4
CS RDS Rescore (Simp.)	2.5	7.7	71.0

Table 3. Results from Docking and Rescoring protocols against *2bok*.

Dockings were all carried out at 30% efficiency, saving 10 GAs each time. Only one run was carried out in each case. The first observation is that all scoring functions generate better enrichment against *2bok* than they do for *1ezq*. This is especially true for the ASP scoring function which now gives an acceptable ROC enrichment metric (for *1ezq* the result was worse than random). Clearly, in this set there are many active ligands which cannot be well accommodated unless the S4 pocket is spacious. The second observation is that rescoring with a second scoring function again gives benefits, at least if the original docks are carried out with GoldScore or ASP (and even with ChemScore, the early enrichment metric is improved on rescore). In addition, the docking/rescoring protocols that appear best (Dock with GoldScore, rescore with either ASP or ChemScore RDS) are the same as those that worked for *1ezq*. Therefore we have

reasonable evidence that these protocols are likely to be useful too, if other factor Xa models are used, and it is possible, may also work if other serine protease enzymes are being screened against. It is worth noting that the best protocol (dock with GoldScore, rescore with ChemScore RDS) gives close to maximal performance.

Conclusions

1. Optimum GOLD docking protocols for virtual screening have been established which adequately balance speed and accuracy. The shortest protocol that should be considered is one running under autoseettings at 10% search efficiency, saving the results of 3GAs per ligand. In most cases the longest protocol that need be considered is a 30% efficiency protocol that saves six GA attempts per ligand. Fast protocols should be used in cases where the binding site is small, slower ones where the binding site is extended and active ligands may contain a number of rotatable bonds. These protocols assume that no protein side chain movement or variable water is allowed during docking.
2. Two docking and rescoring protocols have been found to be the most effective virtual screening protocols for the system considered here, factor Xa. The optimum combination is to dock with GoldScore and rescore with either the Astex Statistical Potential (ASP) or ChemScore RDS. ChemScore RDS is a new variant of ChemScore which rewards burial of hydrogen bonding functionality within subpockets.
3. The best combination approach seems to be to use a scoring function that can reliably recreate reasonable binding poses, and then rescore with a scoring function that is good at identifying structures that make poor contacts with the protein. It is found that, if an optimum pair of scoring functions is used in reverse (i.e. dock with ASP, rescore with GoldScore) poor results are obtained.
4. When using a docking/rescoring protocol, there may be merit in increasing the number of GA poses saved rather than increasing the protocol length. Thus a 10% efficiency protocol saving 10 poses may be as effective as a 30% protocol saving 6 poses.
5. The efficacy of rescoring appears to be magnified if simplex minimisation of the binding pose is allowed during the rescore. This observation is not surprising, however few docking packages allow this option on rescoring. Further work is necessary to confirm that minimisation is always effective for ChemScore RDS rescoring.
6. Comparison of results against two different factor Xa models showed a very significant difference in enrichment performance, particularly in respect to the ASP scoring function. However the most effective docking/rescoring protocols were the same for both models.
7. This work has been carried out using factor Xa as a test case. The optimum protocols established here are quite likely to work well for similar binding cavities i.e for the active sites of other serine proteases. However there is no guarantee that they will work will on systems unrelated to factor Xa. The next target of this work is to establish good virtual screening protocols for a wide range of target types.

Jan 2009