Selection of thrombin-binding aptamers by using computational approach for aptasensor application

Alessandra Bini, Marcello Mascini, Marco Mascini and Anthony Turner

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Selection of Thrombin-binding aptamers by using computational approach for aptasensor application

A. Bini[1,2], K. Karim[2], M. Mascini[3*], M. Mascini[1] and A.P.F. Turner[2,4] &,

[1] Dipartimento di Chimica, Università degli Studi di Firenze, Via della Lastruccia 3 , 50019 Sesto Fiorentino, Italy

[2] Cranfield Health, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK

[3] Dipartimento di Scienze degli Alimenti, Università degli Studi di Teramo, 64023, Teramo, Italy

[4] Biosensors & Bioelectronics Centre, IFM-Linköping University S-58183, Linköping, Sweden

*Corresponding author: mmascini@unite.it

Abstract

The possibility of introducing a computationally-assisted method to study aptamer-protein interaction was evaluated with the aim of streamlining the screening and selection of new aptamers.

Starting from information on the 15-mer DNA aptamer (5'-GGTTGGTGTGGTTGG-3') that binds to thrombin (TBA), a library of mutated DNA sequences (994 elements) was generated and screened using a molecular dynamics method to generate computationally-derived binding scores. The TBA and three other mutated oligonucleotides, selected on the basis of their binding score (best, medium, worst), were incorporated into Surface Plasmon Resonance (SPR) biosensors . By reducing the ionic strength (binding buffer, 50 mM TrisHCl pH 7.4, 140 mM NaCl, 1 mM $MgCl_2$, diluted 1:50) in order to match the simulated condition, the analytical performances of the four oligonucleotide sequences were compared in terms of signal amplitude, sensitivity (slope), linearity (R^2) and reproducibility (CVav %). Thrombin was detected in the concentration range 0-50 nM with a minimum detectable amount of 5 nM. The experimental results indicated that slightly better performance was obtained using the mutated sequence with the best binding score, while a slightly worse performance was found using the mutated sequence with the worst binding score, in agreement with the simulation findings. The behaviours of the TBA and the mutated

sequence with medium binding score were very similar and intermediate between the other two sequences.

1.   Introduction

Aptamers, single stranded nucleic acids generated *in vitro*, have been extensively used in many analytical applications, such as chromatography, electrophoresis, mass spectrometry, molecular beacons, sensors and biosensors. In particular, aptamers have become increasingly important molecular tools for diagnostics and therapeutics, while also offering great flexibility for the design of novel biosensors with high detection sensitivity and selectivity (*1-2*).

The Systematic Evolution of Ligands by EXponential enrichment procedure (SELEX) is a powerful tool for screening large libraries of randomised nucleic acids (generally 10^{13}- 10^{15} different structures), generated by combinatorial chemistry (*3*). Experimentally, the selection of a proper ligand to attain a desired structure is largely a subject of trial and error, due to the large number of possible candidates to be synthesised and tested using a combinatorial approach. The use of molecular modelling software potentially offers a considerable assistance for the rational design of molecular systems with high selectivity and affinity (*4-7*). Moreover some of the drawbacks encountered in the generation of aptamers by the SELEX process could be avoided with the help of a computationally-assisted method. The introduction of a rationalised approach which makes use of a predictive model could overcome or minimise experimental problems currently encountered with the SELEX process, such as reagent stability and separation procedures (*8-9*), non-specific recognition of the support and the fact that pcr tends not to amplify secondary structures. Moreover, a method to partially pre-select aptamer sequences that are more likely to bind the target anylyte, would reduce the practical volumetric problems associated with screening that result in a typical SELEX procedure only searching around 10^{14} of a possible 10^{18} sequences.

Computational approaches are routinely applied to identify candidate compounds for new drugs using information obtained from the target structure, because of the

necessity to use cost- and time- effective ways to indentify lead compounds (*10*). The strength and type of interactions existing between ligands and target determine the recognition properties of the complex. In principle, the ligands with the highest binding score are the best candidates for use in molecular recognition systems.

In this work, the possibility of introducing a molecular modelling method to study aptamer-protein interaction and aptamer selection was evaluated with the aim of introducing a novel and streamlined technique to find new aptamers. Other comparative studies between simulated binding scores obtained from a database of candidates and experimental results, obtained by testing the ability of ligands to bind a target, have been reported, but only with oligopeptides (*4-5, 11-12*) or small organic compounds (*13-16*) as potential ligands.

Here, the thrombin-binding aptamer was chosen as a model and a retrospective docking study was performed by comparing the affinity of mutated sequences for thrombin with that of the thrombin-binding aptamer, based on computationally-derived binding score.

The DNA aptamer that binds to thrombin (TBA) was chosen as a model, because it is one of the most well characterised. Its structure has been determined (*17*), as well as its binding site (*18-19*) and its stability *in vivo* and *in vitro* (*20*). The 15-mer TBA sequence was identified by SELEX as the minimal consensus sequence (5'-GGTTGGTGTGGTTGG-3') from a pool of ~10^{13} 96-mer oligonucleotides containing a randomised 60-mer sequence (*21*). The crystal structure of the complex thrombin-TBA was solved by X-Ray crystallography (*22*). The Nuclear Magnetic Resonance (NMR) model of aptamer-thrombin complex (Protein Data Bank PDB code 1HAO) indicates that the dominant interactions in the fibrinogen exosite involve the TT loops, while the TGT loop interacts with the heparin-binding site of the neighbouring thrombin (*18*). Starting from information on the TBA sequence, we generated and screened a library of mutated DNA sequences using computational methods. Results extrapolated from the computational procedure were verified by correlating the binding scores with the analytical performance of biosensors developed by immobilising the experimental thrombin-binding aptamer and other mutated oligonucleotides, selected on the basis of their binding score, onto Surface Plasmon Resonance (SPR) chips.

Over the past 10 years, many assays, mainly biosensors, based on TBA for the detection of thrombin have been developed. The thrombin-binding aptamer has been coupled with different transduction principles to demonstrate the wide applicability of aptamers as bioreceptors in biosensors with relevance in clinical chemistry (*16, 23-31*).

In this study, the reliability of computational results was tested by experimental measurements. For this purpose, the Thrombin-binding aptamer and other mutated sequences, selected on the basis of their binding score, were employed for the development of optical biosensors and the resulting analytical performances were compared in order to introduce a computational approach to the aptamer selection, thus streamlining the currently experimental selection technique used for aptamer development (*32*).

2.   Materials and methods

2.1.   Reagents and apparatus

15-mer DNA sequences were designed using the software package Hyperchem 7.5. All the other modelling steps were then performed using Openeye Scientific Software tolls (www.eyesopen.com) with free academic license. All molecular modelling experiments were performed on a Acer Veriton T661, RAM 3 GBytes, supported by platform Microsoft Windows XP.

The 15-mer thrombin aptamers (MWG Biotech; Milan, Italy) had the following sequences:

- Biotin-TEG 15-mer aptamer: 5'-Biotin-TEG-GGT TGG TGT GGT TGG-3' (Molecular Weight 5296 g/mol)

- Biotin-TEG 15-mer sequence mutated in positions 5 and 14 (named "Best"): 5'-Biotin-TEG-GGT TTG TGT GGT TAG-3' (Molecular Weight 5255 g/mol)

- Biotin-TEG 15-mer sequence mutated in positions 3 and 11 (named "Medium"): 5'-Biotin-TEG-GGC TGG TGT GAT TGG-3' (Molecular Weight 5265 g/mol)

- Biotin-TEG 15-mer sequence mutated in positions 4 and 15 (named "Wrost"): 5'-Biotin-TEG-GGT AGG TGT GGT TGC-3' (Molecular Weight 5265 g/mol)

With the aim of conferring some flexibility to the immobilised molecule, without interfering with the binding to protein, a TEG (triethylene glycol) tail ($C_2H_{14}O_4$) was

used as spacer arm. The aptamers were received lyophilised and then diluted in MilliQ water and stored at -20 °C.

For the optimisation of the binding procedure between the immobilised aptamer and thrombin by SPR experiments, the binding buffer and regeneration solution were, respectively, Tris-HCl pH 7.4, 140 mM NaCl, 1 mM MgCl2, 0.005% Tween 20 and NaOH 1 or 5 mM, both of which had been optimised in our previous work (*33-34*). The binding buffer without 0.005% Tween 20 was also used for thrombin dilution. The aptamer immobilisation buffer used was 150 mM NaCl, 20 mM Na_2HPO_4, 0.1 mM EDTA pH 7.4 (*35*). All buffer reagents were purchased from Sigma (Milan, Itlay) together with N-Hydroxysuccinimide (NHS), ethanolamine hydrochloride, acetic acid, sodium acetate (10 mM acetate buffer, pH 5.0) and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDAC).

For the SPR experiments, the Biacore™ X analytical device (Biacore; Uppsala, Sweden) was used. The monitoring and analysis of the data was performed using the manufacturer's software Bia-evaluation Version 3.1. Carboxylated dextran-coated chips CM5 (Biacore; Uppsala, Sweden) were employed for the immobilisation of biotinylated aptamers. A shift of 1000 RU represented an increase in surface protein concentration of approximately 1 ng/mm^2. The Biacore resolution was approximately 1.0 RU.

2.2. Computational assay

The 15-mer TBA aptamer (5'-GGT TGG TGT GGT TGG-3') was drawn as a single strand in helix conformation by means of HyperChem 7.5 and used as a reference structure where mutations have to be introduced. HyperChem allowed the substitution of nucleotide residues with others present in the database. The automatic creation of mutated sequences as single strands in helix conformation in HyperChem 7.5 was realised by using a script equipped with a modifiable input file. For limiting the number of sequences to be generated, the number of mutations to be performed with respect to the TBA sequence, was reduced to one or two, obtaining a library of 990 different sequences. The output file consisted of a set of oligonucleotide structures saved in format .pdb, which is compatible with both the HyperChem 7.5 and Openeye software packages. In order to briefly indicate the mutations performed, the following procedure was adopted: the

central number indicating the position in the sequence was positioned between the symbol indicating the nucleotide in the TBA sequence for that position (on the left) and the symbol indicating the mutated nucleotide (on the right). The library was added with three virtually generated sequences, which in literature have been experimentally tested for weakly binding to thrombin:

5'GGT GGT GGT TGT GGT3', having 6 mutations (Mutations T4G, G6T, T7G, G10T, T13G, G15T) with respect to the TBA sequence and referred to as sequence "a" (*21*).

5'GGT AGG GTC GGA TGG3' (Mutations T4A, T7G, G8T, T9C, T12A) referred to as sequence "b" (*36*).

5'GGT AGG GCA GGT TGG3' (Mutations T4A, T7G, G8C, T9A), referred to as sequence "c" (*36*). Therefore, the library contained 994 different sequences as single strands in helix conformation.

Once the library of 15-mer mutated sequences was generated starting from the TBA sequence, the binding score, the main result of computational screening, was obtained taking the conformers of the oligonucleotides library as the ligands database and the heparin-binding exosite or the fibrinogen-binding exosite as the target protein. It should be noted that thrombin shows two different binding sites for aptamers, but only the aptamer-thrombin complex, where the 15-mer TBA interacts with the fibrinogen-binding exosite, has been solved by NMR and it is present in PDB with the code 1HAO. With the aim of studying the aptamer-thrombin interaction also in the heparin-binding exosite, the heparin-thrombin structure (code 1TB6) was downloaded as well from PDB.

After, using Openeye scientific software applications the 994 oligonucleotides were compacted in a single file and fast minimized using Openeye application Szybki1.2.2, using Merck Molecular Force Field (MMFF) max_iter = 4000 (*37*). To take into account the flexibility of the oligonucleotides, conformers were generated for each library member by means of the application Omega2-2.2.1 a systematic high-throughput conformers generator by Openeye, fixing the maximum quantity of conformers generated to 100. Before calculating the binding score, also the two protein receptors, were prepared. By using the Openeye application Fred-Receptor, an active site box, defining the general space of the protein where aptamers should fit in and are expected to bind, was designed around both heparin-binding exosite and fibrinogen-binding exosite.

After these preliminary steps for preparing conformers and protein receptors, FRED the protein-ligand docking program by Openeye was run with ShapeGauss (*38*), as scoring function, that attempts to quantitate how well a pair of atoms interact using a Gaussian representation of atomic volumes. The score is best (more negative) when atomic surfaces touch, but do not interpenetrate. Once the run completed the binding score together with a molecule file containing the docked structures of the top 100 scoring conformers was generated.

2.3. Experimental assay

For the SPR measurements, the biotinylated aptamer immobilisation was realised on dextran-modified CM5 Biacore chips, according to the final part of the protocol and using the following reaction times (*35*):

- Activation step with 50 mM NHS, 200 mM EDAC in water: 7 min.
- Covalent immobilisation of 200 ppm streptavidin in 10 mM acetate buffer, pH 5.0: 7 min.
- Saturation of free carboxylic groups with 1M ethanolamine hydrochloride, pH 8.6: 7 min.
- Aptamer immobilisation after a thermal treatment (95 °C 1 min, 0 °C 10 min): 20 min. The thermal treatment is expected to unfold the aptamer strand making the biotin label at 5' end available for the interaction with streptavidin on the chip surface. Therefore, before the immobilisation, the biotinylated aptamer was heated at 90 °C for 1 min to unfold the DNA strand and then cooled in ice for 10 min in order to block the DNA in its unfolded structure.

After the aptamer immobilisation, the surface was blocked with 500 ppm biotin solution in (20 min), to saturate the free streptavidin sites. The immobilisation procedure was performed at a constant flow rate of 5 µl/min at 25 °C, using as running buffer the same solution as binding buffer, degassed daily prior to use.

Then the binding between the immobilised receptor and the target protein was monitored. The measurement cycle consisted of recording the baseline, followed by target injection, washing with buffer and surface regeneration. The regeneration step released the sensor-bound analyte from the aptamer without affecting its capacity to further bind the protein and allowing a multiple use of the

sensor. The binding between the immobilised receptor and proteins was monitored with an interaction time of 15 min., followed by washing with running buffer and surface regeneration. Binding interactions were monitored at a constant flow rate of 2 µl/min at 25°C.

3. Results and discussion

3.1. Computational assay

The aptamer-thrombin complex NMR structure (PDB code, 1HAO), in which the 15-mer TBA in its typical chair-conformation interacts with the fibrinogen-binding exosite, was downloaded from PDB. By means of Fred-Receptor, a 25803 $Å^3$ binding box was defined and the 15-mer TBA in the chair-shaped conformation was docked to the protein as a reference ligand. Using ShapeGauss as scoring function, 100 possible binding poses were ranked for the chair-shaped aptamer, with binding scores comprised in the range -1762 ÷ -1549. Binding scores were expressed as arbitrary units describing the affinity between aptamer and protein. Using the same conditions above the heparin-binding exosite (PDB code, 1TB6) with a binding box volume of 28427 $Å^3$ was also studied. In this case, the binding scores of the 100 possible binding poses ranked for the 15-mer TBA in the chair-shaped conformation, were comprised in the range -1373 ÷ -1141.

The library of sequences contained the TBA sequence, 990 different sequences having one or two mutations respect to the TBA and the three low-binding sequences reported in literature (named "a", "b" and "c"). As a result, the library contained 994 sequences, which were minimised with a computing time of 1 day, 2 h, 29 min. To take into account the flexibility of the oligonucleotides, overcoming the serious limitation of considering nucleotides as rigid, conformers were generated for each sequence: the Openeye software OMEGA generated an average of 53 conformers for each sequence, in 5 days, 5 h, 1 min. Consequently, the database was represented by the conformers of the 994 oligonucleotide, around 53000 units. In order to give an idea of the minimisation energy and number of conformers obtained, the results for the sequences selected for the following experimental part and for the three low-binding sequences were quantitatively reported in Table 1.

Code	Sequence	Mutations	Minimisation Energy (Kcal/mol)	Conformers
TBA	5'-GGT TGG TGT GGT TGG-3'		-1030	44
Best	5'-GGT TTG TGT GGT TAG-3'	G5T, G14A	-993	91
Medium	5'-GGC TGG TGT GAT TGG-3'	T3C, G11A	-959	112
Worst	5'-GGT AGG TGT GGT TGC-3'	T4A, G15C	-996	61
"a"	5'-GGT GGT GGT TGT GGT-3'	T4G, G6T, T7G, G10T, T13G, G15T	-1050	45
"b"	5'-GGT AGG GTC GGA TGG-3'	T4A, T7G, G8T, T9C, T12A	-833	39
"c"	5'-GGT AGG GCA GGT TGG-3'	T4A, T7G, G8C, T9A	-834	80

Table 1. Computational summary results for the oligonucleotide sequences selected for experimental screening. The oligonucleotide sequences, the mutations with respect to the TBA sequence, the minimisation energy and the number of conformers generated are reported.

After these preliminary steps for the binding box and preparing conformers, FRED (the protein-ligand docking program by Openeye) was run with ShapeGauss as scoring function. The run completed in 1 day, 12 hours, 33 minutes when docking was performed in the fibrinogen-binding exosite and in 1 day, 10 hours, 31 minutes when docking was performed in the heparin-binding exosite. Therefore, the binding score together with a molecule file containing the docked structures of the top scored conformer for each sequence was generated. The binding score obtained was given only by one component, i.e. the shape score, while the hydrogen atoms were ignored. From the docking simulation in the fibrinogen-binding exosite, 174 sequences were not scored since no valid pose was found. The best conformer in terms of binding score was that coming from the mutated

sequence G5T, G14A, with a total score of −1748. Considering the full ligand library, the total score ranged from −1748 to −667, but only 98 elements reached a score between −1748 and −1500, while the majority of oligonucleotides (657 elements) had total score between −1500 and −1000 and 65 sequences had a total score lower than -1000 (Figure 1).
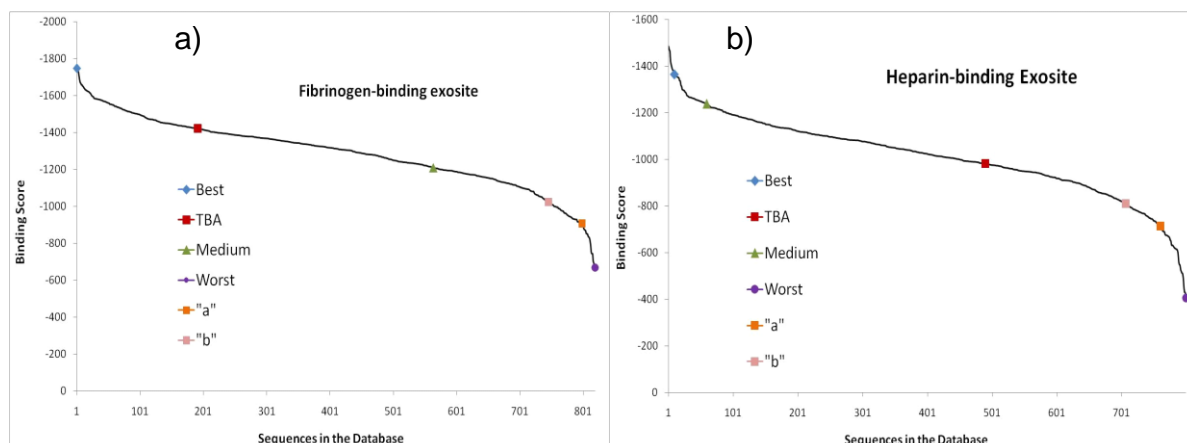

Figure 1. Binding score for each sequence in the database docked to: a) the fibrinogen-binding exosite; b) the heparin-binding exosite.

The docking step in the heparin-binding exosite failed for 194 sequences, which was not scored since no valid pose was found. The best conformer in term of binding score was that coming from the mutated sequence G6T, with a total score of -1486. Considering the full ligand library, the total score was comprised between -1486 and -404, therefore lower binding scores were in general found in the heparin-binding exosite with respect to the fibrinogen-binding exosite. Only 90 elements reached a score between −1486 and −1200, while the majority of oligonucleotides (672 elements) had total score between −1200 and −700 and 38 sequences had a total score higher than -700.

Taking into consideration the two binding exosites, the mutated sequence G5T,G14A (named "Best") was the top scored in the fibrinogen-binding exosite and the tenth scored in the heparin-binding exosite; the TBA sequence was ranked 192th and 490th in the two exosites respectively, the mutated sequence T3C,G11A (named "Medium") was ranked 564th and 60th respectively and the mutated sequence T4A,G15C (named "Worst") was bottom ranked in both the binding sites. These binding scores were taken into consideration in the choice of the ligands for the experimental test, in order to have a variety of candidates: the best and worst ranked, for which a huge difference in term of binding scores was

highlighted, plus a sequence for which an intermediate behaviour had to be expected. A binding score % was calculated by assigning value 100 to the mutated sequence G5T,G14 ("Best") for the fibrinogen-binding site and to the mutated sequence G6T for the heparin-binding site.

Regarding the three sequences experimentally tested in literature for low-binding, docking of sequence "c" failed in both the exosites, while sequences "a" and "b" were ranked bottom in both the exosites (binding Score sequence "a": -907 and -713 respectively; binding Score sequence "b": -1023 and -811 respectively).

The simulation results for the most interesting sequences are summarized in Table 2.

Code	Fibrinogen-binding exosite 820 docked sequences			Heparin-binding exosite 800 docked sequences		
	Binding Score	Hit-list position	Binding Score %	Binding Score	Hit-list position	Binding Score %
Best	-1748	1	100	-1364	10	91.8
TBA	-1422	192	81.3	-982	490	66.1
Medium	-1209	564	69.2	-1256	60	83.3
Worst	-667	820	38.2	-404	800	27.2
"a"	-907	799	51.9	-713	761	48.0
"b"	-1023	746	58.5	-811	707	54.6

Table 2. Binding scores and relative hit-list position for the oligonucleotide sequences chosen for the experimental test and for the docked sequences reported in literature for low-binding to thrombin.

Binding scores obtained by docking the database in the fibrinogen-binding exosite were graphically correlated with the binding scores obtained by docking the library in the heparin-binding exosite (Figure 2). This aimed to immediately represent the behaviour of each docked sequence in both the binding sites. As can be observed, most of the sequences tended to bind the fibrinogen-binding site rather than the heparin-binding site, and the mutated sequence named "Medium" obtained the same binding scores for both the exosites.
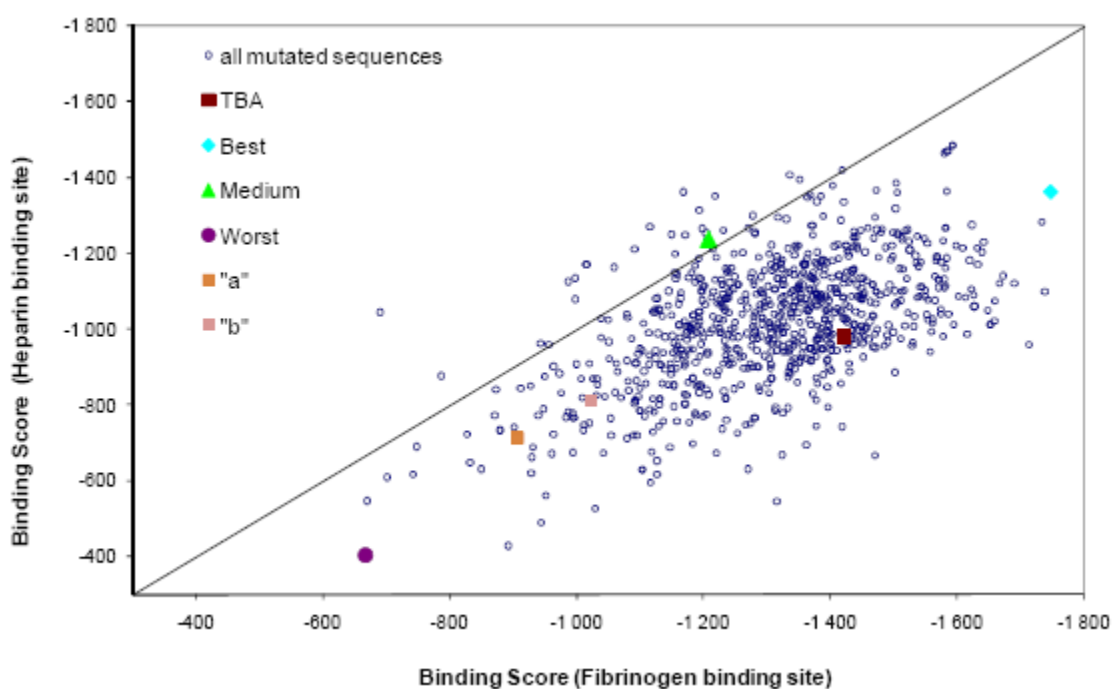
Figure 2. Scatter plot of the binding scores obtained by docking the database in the fibrinogen-binding site versus the binding scores obtained by docking the database in the heparin-binding site.

3.2. Experimental assay

Experimental tests were conducted in order to verify if the binding scores agreed with the analytical behaviour of the sequences. For this purpose, the 15-mer thrombin-binding aptamer as well as the best sequence, the worst sequence and the sequence with intermediate behaviour resulting from the simulation experiments were immobilised on CM5 Biacore chips, in concentration 0.5 µM, following the procedure described in previous paragraph. To afford each immobilised sequence a certain flexibility, a non oligonucleotide spacer (biotin-TEG tail) was chosen in order to avoid base-pairing, which affects the aptamer's secondary structure. Two chips were employed for each sequence and, from the resulting immobilisation shifts, a similar density of immobilised molecules was estimated considering 1000 RU = 0.8 ng/mm^2 (Biacore User Manual). This density similarity allowed a direct comparison among the different sequences when testing their sensitivity for the detection of thrombin. The results for the immobilisation step are reported in Table 3.

Sequence Code	Chip 1		Chip 2	
	Immobilisation Shift (RU)	Density (molecules/cm^2)	Immobilisation Shift (RU)	Density (molecules/cm^2)
TBA	1116	1.0×10^{13}	1194	1.1×10^{13}
Best	931	8.5×10^{12}	918	8.4×10^{12}
Medium	1126	1.0×10^{13}	1080	9.9×10^{12}
Worst	957	8.8×10^{12}	911	8.3×10^{12}

`

Table 3. Comparison between different immobilised Biotin-TEG oligonucleotide sequences onto Biacore CM5 chips. Concentration of each sequence equal to 0.5 µM.

3.2.1. Optimisation of experimental conditions with Biacore

As a first optimisation step, SPR measurements were performed in order to determine the optimal aptamer-analyte binding protocol. For this purpose, the 5'-biotinylated 15-mer thrombin aptamer with TEG tail was immobilised onto the SPR chip at a concentration of 0.5 µM. The binding buffer optimised in a previous work (*34*) was tested for the dilution of thrombin to 20 nM and the binding with the immobilised aptamer was then monitored. The regeneration solution, studied in a previous work as well (*33*), was NaOH at two concentrations 1 or 5 mM depending of the drift to be regenerated, with the aim of releasing the bound analyte from the sensor without affecting the aptamer capacity to bind thrombin. These optimised binding conditions (binding buffer and regeneration solution) were then transferred to the system.

Binding experiments were performed with the target thrombin solution prepared in the optimised buffer, as reported above. The same buffer was used also as running solution. The thrombin-aptamer interaction time was 15 minutes and the surface was regenerated after each binding cycle by a treatment with 1 or 5 mM of NaOH.

When testing the TBA sequence, a binding shift was recorded in the thrombin concentration range 0-10 nM, with 0.5 nM as minimum detectable amount of the analyte. On the contrary, when testing the mutated sequences, a change in refractive index was observed during the protein injection, but no binding shift was recorded, even at very high thrombin concentration.

It should be noticed that in the computational experiments, phenomena occurring in solution such as ionic strength, pH and in trace compounds interferences were not taken into account. For this reason, in the experimental test the ionic strength of the binding buffer was drastically reduced, in order to reproduce the binding conditions of the simulation. For this purpose, a 400 nM thrombin solution was prepared by diluting the 4000 nM stock solution (in 50 mM TrisHCl, pH 7.4, 140 mM NaCl, 1 mM $MgCl_2$) in MilliQ water and the binding curve was registered, using the binding buffer as running solution. Under these conditions, a binding curve was registered, but a complete dissociation was observed at the end of injection. On the contrary, when MilliQ water was used as running buffer as well as for the preparation of 400 nM thrombin, the binding curve did not dissociate at the end of the injection and the measurement of a binding shift was possible.

It seemed that the binding curve shape as well as the binding shift depended upon the ionic strength of the binding and running buffer, which had to be minimised to reproduce the conditions used in the computational experiments. Curve shape suggested that a successful rapid binding to a stable plateau was not achieved when experiments were performed in water; on the contrary, a slow binding was observed, without reaching a plateau during the ligand pulse. This could cause an irreproducible binding, probably due to the binding buffer conditions (Biacore User Manual). Moreover, the protein structure could be affected in a highly diluted environment and then a baseline drift was observed when working in MilliQ water. To overcome these problems, further experiments were performed in order to find the lowest ionic strength that supported a binding curve without departing from the computational conditions. For this reason, the optimised binding buffer was diluted 1:10 or 1:50 with MilliQ water and the resulting solutions were used as running and binding buffer. Using the buffer diluted 1:10, a slight dissociation was observed when 100 nM thrombin was bound to the mutated sequence G5T,G14A ("Best") (Figure 3). On the contrary, using the buffer diluted 1:50 the binding curve did not dissociate at the end of the injection. Therefore, this solution (1 mM TrisHCl, pH 7.4, 2.8 mM NaCl, 0.02 mM $MgCl_2$) was chose for the following experiments, aimed to verify if the experimental and computational analysis agreed.
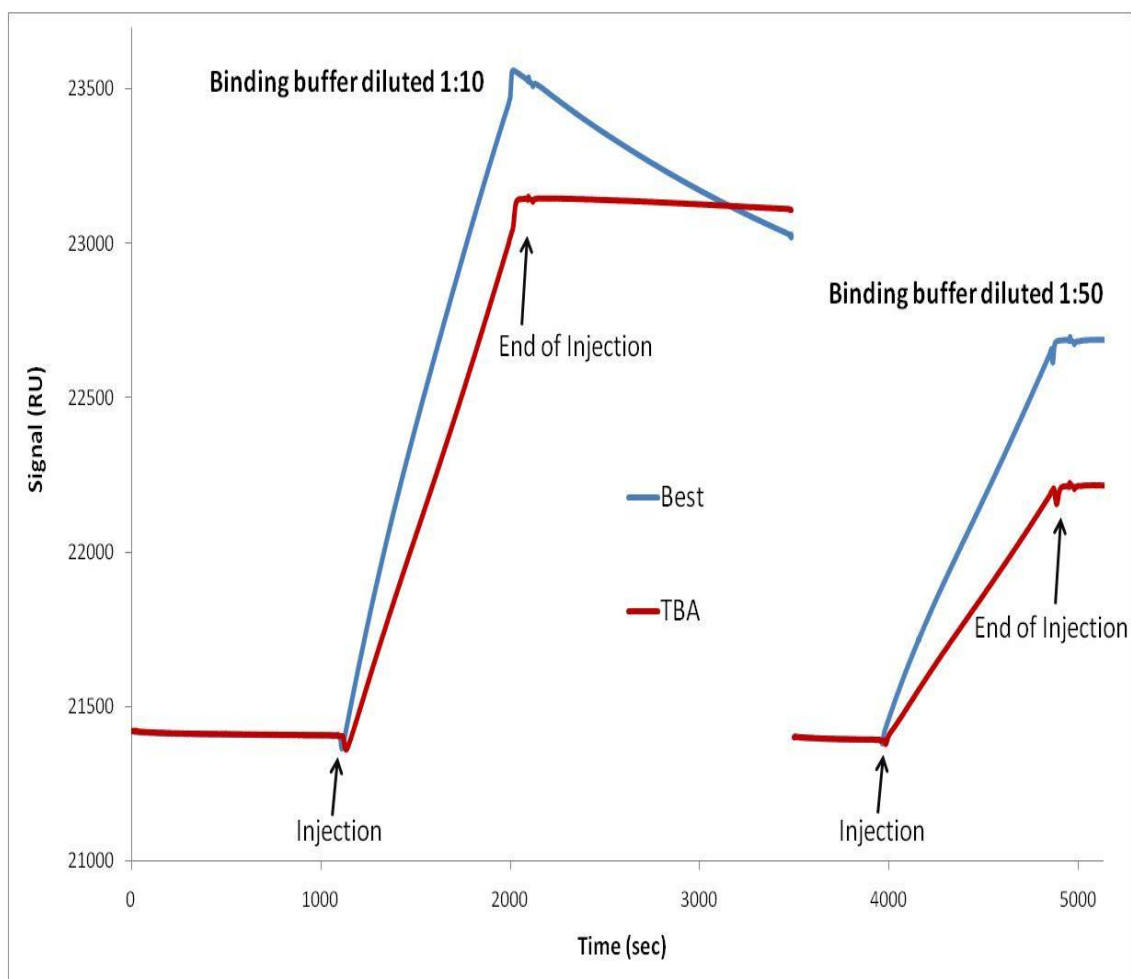
Figure 3. Binding curves for 100 nM thrombin interacting with the immobilised mutated sequence G5T,G14A ("Best") and the TBA sequence, when binding buffer diluted 1:10 or 1:50 was used as running solution as well as for the dilution of the stock thrombin solution. In the case of binding buffer diluted 1:50, binding shifts of 1296 RU and 824 RU were recorded when the immobilised sequence was G5T,G14A or TBA, respectively.

3.2.2. Comparison between the experimentally tested sequences

Using the binding buffer diluted 1:50, the analytical performances of the four oligonucleotide sequences were compared in terms of signal amplitude, sensitivity (slope), linearity (R^2) and reproducibility (CVav %). The results indicated that a slightly better performance was obtained with the mutated sequence G5T,G14A ("Best"), while a slightly worse performance was obtained with the mutated sequence T4A,G15C ("Worst"), in agreement with the simulation findings (Table 4). The behaviours of the TBA and the mutated sequence T3C,G11A ("Medium") were very similar and intermediate between the other two sequences.

Sequence Code	Linear Regression	R^2	CVav %

TBA	y = 5.35x – 10.7	0.992	19
Best	y = 7.86x – 18.8	0.994	15
Medium	y = 5.05x – 8.2	0.997	18
Worst	y = 3.33x – 7.5	0.985	21

Table 4. Analytical behaviour of the four experimentally tested immobilised sequences in the optimised binding conditions (binding buffer diluted 1:50). The CVav % was calculated using three replicates (n=3).
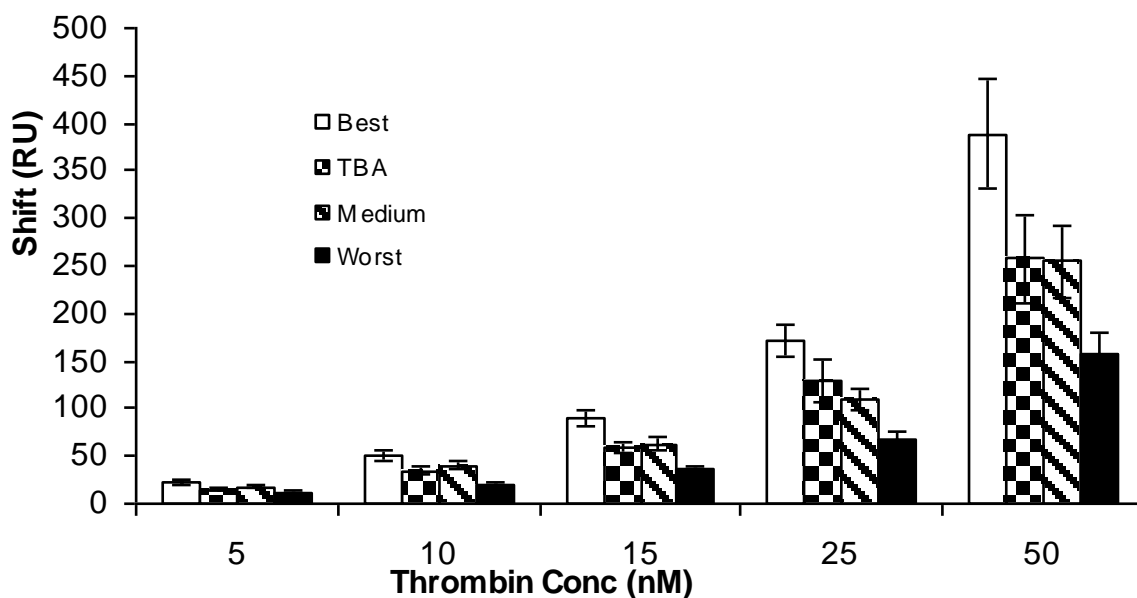

Figure 4. Histogram of the thrombin binding shifts recorded with the four experimentally tested immobilised sequences in the optimised binding conditions (binding buffer diluted 1:50). Each error bar represents the standard deviation of three replicates (n=3).

As a consequence of the reduction in ionic strength, thrombin was detected in the concentration range 0-50 nM with a minimum detectable amount of the analyte of 5 nM, with all the four sequences tested. In contrast, a thrombin concentration range of 0-10 nM was achieved with the immobilised TBA sequence, using the optimised buffer as binding and running buffer, with a minimum detectable amount of analyte of 0.5 nM.

4. Conclusion

The possibility of introducing a computational method for aptamer selection was evaluated. The introduction of an *in silico* approach to aptamer selection could reduce the drawbacks and limitations currently faced by SELEX and other experimental procedures. The affinity of each mutated sequence towards thrombin was described by a binding score and compared with the one obtained with the TBA. The sequences showing the best, medium and worst binding score among the docked oligonucleotides were employed for the development of optical biosensors, the analytical characteristics of which were compared with those obtained for TBA. A good correlation between *in silico* selection and experimental results was found when reproducing an experimental environment close to that computationally simulated. Despite the fact that the TBA binding score corresponded to only 81.3% of the best candidate, the computational approach partially confirmed results observed in SELEX, identifying high affinity binding sequences (TBA) from low affinity binding sequences reported in literature. This work could have a significant impact onthe effieiciency and effectiveness of aptamer selection for sensors and diagnostics, By facilitating the partial pre-selection of apatmer sequences to be further tested in an experimental screening process, this computational approach could allow the full range of relevant apatmer sequences to be tested in a manageable experimental format. In addition, this new research approach may encourage interdisciplinary research in this field, with the objective of facilitating the development of new and improved algorithms.

5. References

1.   Tombelli, S., Minunni, M., and Mascini, M. (2005) Analytical applications of aptamers, *Biosensensors & Bioelectronics 20*, 2424–2434.
2.   Tombelli, S., Minunni, M., and Mascini, M. (2007) Aptamer-based assays for diagnostics, environmental and food analysis, *Biomolecular Engineering 24*, 191-200.
3.   Ellington, A. D., and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands, *Nature 346*, 818-822.
4.   Mascini, M., Guilbault, G. G., Monk, I. R., Hill, C., Del Carlo, M., and Compagnone, D. (2008) Screening of rationally designed oligopeptides for Listeria monocytogenes detection by means of a high density colorimetric microarray, *Microchimica Acta 163*, 227–235

5. Mascini, M., Sergi, M., Monti, D., Del Carlo, M., and Compagnone, D. (2008) Oligopeptides as mimic of acethylcholinesterase: from the rational design to the application in solid-phase exctraction for pesticedes, *Analytical Chemistry 80* 9150–9156.

6. Rajamani, R., and Good, A. C. (2007) Ranking poses in structure-based lead discovery and optimisation: current trends in scoring function development, *Current Opinion in Drug Discovery & Development 10*, 308-315.

7. Sanchez-Barragan, I., Karim, K., Costa-Fernandez, J. M., Piletsky, S. A., and Sanz-Medel, A. (2007) A molecularly imprinted polymer for carbaryl determination in water, *Sensors and Actuators B 123*, 798–804.

8. Kim, N., Gan, H. H., and Schlick, T. (2007) A computational proposal for designing structured RNA pools for in vitro selection of RNAs, *Bioinformatics 13*, 478-492.

9. Hamula, C. L. A., Guthrie, J. W., Zhang, H., Li, X., and Le, X. C. (2006) Selection and analytical applications of aptamers, *Trends in Analytical Chemistry 25*, 681-691.

10. Goldenhuys, W. J., Gaasch, K. E., Watson, M., Allen, D. D., and Van der Schyf, C. J. (2006) Optimizing the use of open-source software applications in drug discovery, *Drug Discovery Today 11*, 127-132.

11. Mascini, M., Del Carlo, M., Compagnone, D., Cozzani, I., Tiscar, P. G., Mpmhanga, C. P., and Chen, B. (2006) Piezoelectric sensors based on biomimetic peptides for the detection of heat shock proteins (HSPs) in mussels, *Analytical Letters 39*, 1627-1642.

12. Mascini, M., Macagnano, A., Monti, D., Del Carlo, M., Paolesse, R., Chen, B., Warner, P., D'Amico, A., Di Natale, C., and Compagnone, D. (2004) Piezoelectric sensors for dioxins: a biomimetic approach, *Biosensors and Bioelectronics 20*, 1203-1210.

13. Chianella, I., Lotierzo, M., Piletsky, S. A., Tothill, I. E., Chen, B., Karim, K., and Turner, A. P. F. (2002) Rational design of a polymer specific for microcystin-LR using a computational approach, *Analytical Chemistry 74*, 1288-1293.

14. Piletska, E. V., Tuner, N. W., Turner, A. P. F., and Piletsky, S. A. (2005) Controlled release of the herbicide simazine from computationally designed molecular imprinted polymers, *Journal of controlled release 108*, 132-139.

15. Yakaiah, T., Lingaiah, B. P. V., Narsaiah, B., Shireesha, B., Ashok Kumar, B., Gururaj, S., Parthasarathy, T., and Sridhar, B. (2007) Synthesis and structure activity relationship of novel pyrimido[1,2-b]indazoles as potential anticancer agents against A-549 cell lines, *Bioorganic & Medicinal Chemistry Letters 17*, 3445-3453.

16. Li, M., Huang, Y.-J., Tai, P. C., and Wang. B. (2008) . 368. (2008) Discovery of the first SecA inhibitors using structure-based virtual screening, *Biochemical and Biophysical Research Communications 368*, 839-845.

17. Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A., and Feigon, J. (1993) Thrombin-binding DNA aptamer forms an unimolecular quadruplex structure in solution, *Proceedings of the National Academy of Sciences 90*, 3745-3749.

18. Kelly, J. A., Feigon, J., and Yeates, T. O. (1996) Reconciliation of the X-ray and NMR structures of the thrombin-binding aptamer d(GGTTGGTGTGGTTGG), *Journal of Molecular Biology 256*, 417-422.

19. Padmanhabhan, K., and Tulinsky, A. (1996) An ambiguous structure of a DNA 15-mer thrombin complex, *Acta Crystallographica D52*, 272-282.

20. Keefe, A. D., and Schaub, R. G. (2008) Aptamers as candidate therapeutics for cardiovascular interactions, *Current Opinion in Pharmacology 8*, 147-152.

21. Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H., and Toole, J. J. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin, *Nature 355*, 564-566.

22. Padmanhabhan, K., and Tulinsky, A. (1993) The structure of α-Thrombin inhibited by a 15-mer single stranded DNA aptamer., *The Journal of Biological Chemistry 268*, 17651-17654.

23. Zheng, J., Feng, W., Lin, L., Zhang, F., Cheng, G., He, P., and Fang, Y. (2007) A new amplification strategy for ultrasensitive electrochemical aptasensor with network-like thiocyanuric acid/gold nanoparticles, *Biosensors and Bioelectronics 23* 341-347.

24. Schlensog, M. D., Gronewold, T. M. A., Tewes, M., Famulok, M., and Quandt, E. (2004) A Love-wave biosensor using nucleic acids as ligands, *Sensors and Actuators B 101*, 308-315.

25. Potyrailo, R. A., Conrad, R. C., Ellington, A. D., and Hieftje, G. M. (1998) Adapting selected nucleic acid ligands (aptamers) to biosensors, *Analytical Chemistry 70*, 3419-3425.

26. McCauley, T. G., Nobuko, H., and Stanto, M. (2003) Aptamer-based biosensor arrays for detection and quantification of biological macromolecules, *Analytical Biochemistry 319*, 244-250.

27. Lin, C., Katilius, E., Liu, Y., Zhang, J., and Yan, H. (2006) Self-assembled signaling aptamer DNA arrays for protein detection, *Angewandte Chemie International Edition 45*, 5296-5301.

28. Li, Y., Lee, H. J., and Corn, R. M. (2007) Detection of protein biomarkers using RNA aptamer microarrays and enzimatically amplified Surface Plasmon Resonance Imaging, *Analytical Chemistry 79*, 1082-1088.

29. Lee, M., and Walt, D. R. (2000) A fiber-optic microarray biosensor using aptamers as receptors, *Analytical Biochemistry 282*, 142-146.

30. Hianik, T., Ostatna, V., Zajacova, Z., Stoikova, E., and Evtugyn, G. (2005) Detection of aptamer-protein interactions using QCM and electrochemical indicator methods, *Bioorganic and Medicinal Chemistry Letters 15*, 291-295.

31. He, P., Shen, L., Cao, Y., and Li, D. (2007) Ultrasensitive electrochemical detection of proteins by amplification of aptamer-nanoparticle bio bar codes, *Analytical Chemistry 79*, 8024-8029.

32. Hall, B., Hesselberth, J. R., and Ellington, A. D. (2007) Computational selection of nucleic acid biosensors via a slip structure model, *Biosensors and Bioelectronics 22*, 1939-1947.

33. Bini, A., Minunni, M., Tombelli, S., Centi, S., Mascini, M., and (2008) Development of an optical RNA-based aptasensor for C Reactive Protein, *Analytical and Bioanalytical Chemistry 390*, 1077-1086.

34. Centi, S., Tombelli, S., Minunni, M., and Mascini, M. (2007) Aptamer-based detection of plasma proteins by an electrochemical assay coupled to magnetic beads, *Analytical Chemistry 79*, 1466-1473.

35. Storri, S., Santoni, T., Minunni, M., and Mascini, M. (1998) Surface modification for the development of piezoimmunosensors, *Biosensors and Bioelectronics 13*, 347-357.

36. Tasset, D., Kubik, M. F., and Steiner, W. (1997) Oligonucleotide inhibitors of human thrombin that bind distinct epitopes, *Journal of Molecular Biology 272*, 688-698.

37. Halgren, T. A. (1996) Merck Molecular Force Field I. Basis, form, scope, parameterization and performances of MMFF94, *Journal of Computational Chemistry 17*, 490-519.

38. McGann, M. R., Almond, H. R., Nicholls, A., Grant, J. A., and Brown, F. K. (2003) Gaussian Docking Functions, *Biopolymers 68*, 76-90.