Using multiobjective optimization and energy minimization to design an isoform-selective ligand of the 14-3-3 Protein

Sanchez-Faddeev, H.; Emmerich, Michael T.M.; Verbeek, F.J.; Henry, A.H.; Grimshaw, S.; Spaink, H.P.; ... ; Steffen, B.

Using Multiobjective Optimization and Energy Minimization to Design an Isoform-Selective Ligand of the 14-3-3 Protein

Hernando Sanchez-Faddeev[1], Michael T.M. Emmerich[1], Fons J. Verbeek[1], Andrew H. Henry[2], Simon Grimshaw[2], Herman P. Spaink[3], Herman W. van Vlijmen[4], and Andreas Bender[4]

[1] Leiden Institute of Advanced Computer Science, Leiden University, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
[2] Chemical Computing Group, St John's Innovation Centre, Cowley Road, Cambridge, United Kingdom, Cambridge CB40WS, United Kingdom
[3] Institute of Biology, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands
[4] Medicinal Chemistry Division, Leiden / Amsterdam Center for Drug Research, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands

Abstract. Computer simulation techniques are being used extensively in the pharmaceutical field to model protein-ligand and protein-protein interactions; however, few procedures have been established yet for the design of ligands from scratch ('*de novo*'). To improve upon the current state, in this work the problem of finding a peptide ligand was formulated as a bi-objective optimization problem and a state-of-the-art algorithm for evolutionary multiobjective optimization, namely SMS-EMOA, has been employed for exploring the search space. This algorithm is tailored to this problem class and used to produce a Pareto front in high-dimensional space, here consisting of 23^{22} or about 10^{30} possible solutions. From the knee point of the Pareto front we were able to select a ligand with preferential binding to the gamma versus the epsilon isoform of the *Danio rerio* (zebrafish) 14-3-3 protein. Despite the high-dimensional space the optimization algorithm is able to identify a 22-mer peptide ligand with a predicted difference in binding energy of 291 kcal/mol between the isoforms, showing that multiobjective optimization can be successfully employed in selective ligand design.

Keywords: protein design, ligand design, de novo assembly, SMS-EMOA, multiobjective optimization, 14-3-3, Pareto front, multiobjective selection, hypervolume indicator.

1 Introduction

Over vast timescales, nature has optimized the genetic material to account for survival of organisms that are better adapted to the immediate, local environment [1]. The interplay of genetic variation and natural selection is the driving force of DNA evolution that, for instance, determines the function of all proteins encoded by the genome.

T. Margaria and B. Steffen (Eds.): ISoLA 2012, Part II, LNCS 7610, pp. 12–24, 2012.

Evolutionary algorithms (EA) seek to mimic this process on an algorithmic level [2]. Starting from an initial population of candidate solutions, the application of variation (mutation and recombination) operators and selection operators adapts the solutions to its environment (the fitness function).

In EAs the adaptation to the environment is given by a user-defined fitness function, which determines the likelihood of new genotypes for survival recombination and/or selection; equivalent to reproductive fitness in nature [2]. The defining difference is that fitness functions can be used to optimize the same type of molecules that have been created in the process of evolution with a user-defined objective in mind.

Taking into consideration the complexity of molecular interactions and the promiscuity of those interactions observed in nature [3,4], it is often not sufficient to have only one optimization goal; rather, when designing a ligand for a protein, it is important to make sure it is selective for the target(s) of interest, relative to other, possibly very similar targets [3]. As in the natural environment multiobjective EAs can deal with several conflicting objective functions at the same time and select trade-off solutions that are better suited for both objectives; however, it will favor those solutions which are superior to all others in at least one way (they are the 'non-dominated' solutions) [5].

The desired activity profile for a set of targets is relevant both for efficacy of a compound in a biological system, as well as to avoid adverse side effects such as in case of drugs that are applied to humans [3]. Given the current huge amount of bioactive data we are becoming aware of the suitability of a ligands with a bioactivity profile of interest, with areas such as 'chemogenomics' and 'proteochemometrics' gaining increasing importance [6,7,8].

We have applied structure-based multiobjective ligand design to the family of 14-3-3 proteins, which are present in multiple isoforms in all eukaryotic organisms. Given that this protein is also of large biomedical interest such as in cancer research, and the requirement for a ligand to prefer some 14-3-3 isoforms over others, we chose this case study for the in silico design of isoform-selective peptide ligand design.

14-3-3 proteins participate in many biological processes including protein kinase signaling pathways within all eukaryotic cells, being involved in progression through the cell cycle, initiation and maintenance of DNA damage checkpoints, regulation of mitosis, prevention of apoptosis, and coordination of integrin signaling and cytoskeleton related dynamics [9]. Current studies have demonstrated the important role that 14-3-3 plays in cancer [10], particularly leukaemia as described by Dong et al.[11], and Alzheimer's disease as reported by Jayaratnam et al.[12]. The 14-3-3 proteins are intensively studied in many animal species such as zebrafish, mouse and human.

We recently reported on the 14-3-3 protein isoforms in zebrafish (*Danio rerio*) that are encoded by eleven genes named after their Homo sapiens homologues [13], while the human isoforms γ, β, ε and θ each possess two homologue isoforms in zebrafish. Zebrafish isoforms are the subject of this study due to their similarity to the human isoforms, as well as to be able to validate the predictions later directly by means of in vitro experiments.

The differential expression of various isoforms in different tissues and diseases suggests that different isoforms possess different functionality, which implies different binding preferences for particular ligands [14]. However, only little is known on the differences in the binding specificities to 14-3-3 proteins. Given the large number of isoforms of 14-3-3 and their different roles, the design of specific ligands is important to achieve; yet it is a task that is not trivial in practice. Given recent advances in molecular modeling as well as computational optimization techniques, this study now aims at merging the best of both worlds in order to establish advanced computational methods for 14-3-3 γ1 isoform specific ligand design.

A recent review [15] outlines the opportunities and challenges in the application of computer tools to design peptide based drugs, an area of which we present an application of particular interest. Several previous studies have focused on in silico peptide screening for potential new therapeutic entities [16,17,18,19,20,21,22,23,24]; however, in those cases only existing peptides were screened virtually and evaluated with respect to their ability to bind to a protein of interest. The approach taken in this work is rather different, however: instead of screening a library of known peptides and scoring the best solutions, this study focuses on mutating the peptide in a step-wise optimization process, in order to achieve better affinity and to access novel chemical space, in a 'de novo' peptide design approach.

Relating this to previous work, Li et al. [25] analyzed peptide binding to the p53–MDM2/MDMX interface by randomly mutating and evaluating affinities using computational methods. Our random mutation process is similar, except for the fact that Li et al. [25] used a single objective function, so no other protein interactions other than the one with the intended target were taken into account.

However, in this work we emphasize both the de novo, as well as the multiobjective nature of peptide ligand design. We do so by trying to identify a peptide with high binding affinity for the γ1 isoform, as compared to the ε1 isoform. ε1 and γ1 isoforms have been selected as they have been suggested to have different biological functions, namely γ1 has a specialized function in adult physiology, and ε1 is highly expressed during the embryonic stage [13,26,27].

Multiobjective optimization is meant to find good compromises (or "trade-offs"), rather than a solution that is optimal in a single objective function only. If the number of conflicting objectives is low, a well-established approach is to approximate the Pareto front of the problem, i.e., the set of non-dominated optimal solutions, or rephrasing the above a set of optimal "trade off" solutions [28,31]. With two objectives minimization the solution is "Pareto-optimal" if there exists no other solution which improves the one of the objective function values without causing a simultaneous deteriorating of the other objective function value. This is visualized in Figure 1A, which contains a Pareto-optimal set of solutions that were generated in this study.

Multi-objective optimization is not easy to perform in high-dimensional spaces due to the sheer size of the hypothesis space. EA, due to their population-based search concept and high number of generated solutions, lend themselves very well for the task of generating and maintaining Pareto-optimal solutions in higher-dimensional spaces [29]. Pareto optimization hence recently received increasing attention in drug design problems [19,20,30] and bioinformatics [31], besides other application fields.

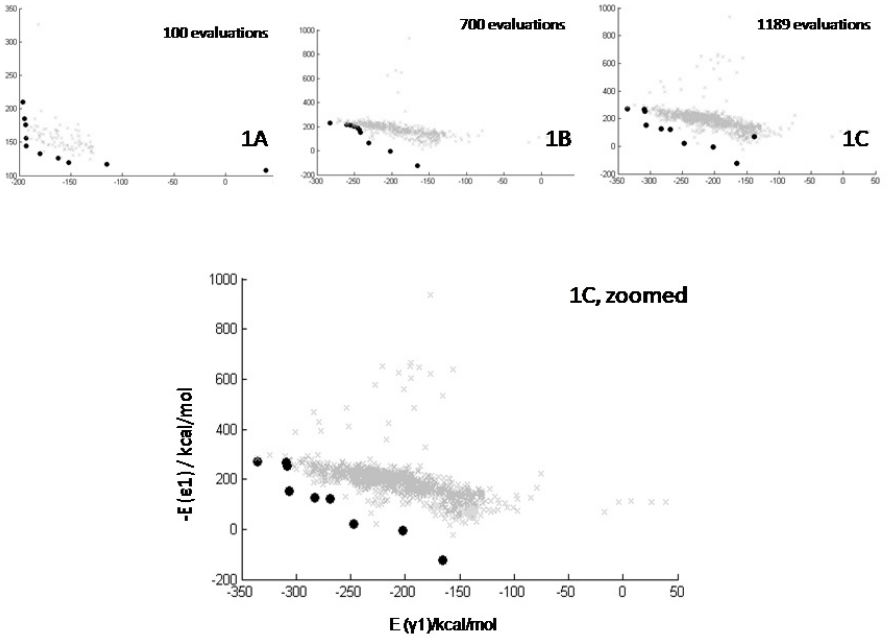

Fig. 1. Binding energy of peptides to the γ1 and ε1 isoforms of the 14-3-3 protein as a function of the number of optimization iterations using the SMS-EMOA algorithm. The X-axis shows the potential energy of interaction with γ1 (which was desired) while the Y-axis shows the inverse of the potential energy of interaction with ε1 (which we attempted to 'design out' of the peptide). The solutions that belong to the Pareto front are represented by circles. The intermediate solutions rejected during the SMS-EMOA run are represented by crosses. It can be seen that already after 100 iterations partially selective peptides are obtained, while after the full number of 1,189 iterations even peptides with no affinity to the ε1 isoform, but 335 kcal/mol binding energy to the γ1 isoform could be identified. Hence, our optimization can be considered successful even in this 22-dimensional search space.

This study intends to use EA with multiobjective optimization to find a binding peptide with relative high binding affinity for the γ1 isoform, as compared to the ε1 isoform [32]. We use multiobjective EA as implemented in SMS-EMOA [28,33], a state-of-the-art Pareto optimization algorithm for this purpose as described in the following.

2 Methods

2.1 Sequence Data

The nucleotide sequences of 14-3-3 isoforms in zebrafish were described by Besser et al. [13]. They performed a phylogenetic analysis of the 14-3-3 family together with microarray expression analysis; the results provided the basis for the choice of 14-3-3 isoforms analyzed in this study.

2.2 Homology Modeling

Zebrafish 14-3-3 homology models were generated via the Molecular Operating Environment [34]. Table 1 displays the PDB templates used for homology modeling and the sequence similarity with the corresponding zebrafish isoforms. The high sequence identity of 96% allowed a construction of highly reliable models. An RMS gradient of 0.1 was employed to build intermediate homology models and an RMS gradient of 0.01 was used for generating the final models. AMBER99 (default) distance-dependent force field parameters were applied in energy minimization after homology modeling.

Table 1. Templates used for homology modeling of the 14-3-3 isoforms

Zebrafish Isoforms	modeled residue range	PDB template (resolution)	Sequence Identity [%]	E-value
Epsilon-1	3 to 232	2br9A (1.75Å)	96.5	1.79E-114
Gamma-1	2 to 234	2b05A (2.55Å)	96.1	7.16E-111

2.3 Starting Complex of Protein and Ligand

The starting point for generating peptide ligands was 22 amino acids long in order to allow interaction both with the binding groove as well as the regions immediately outside to achieve selectivity. This length has also been chosen based on the location of variable regions of 14-3-3 as well as low energy desolvation sites identified previously [14]. 23 possible amino acids could be selected in each position, namely the 20 natural amino acids as well as phosphorylated tyrosine, serine and threonine. The reason for also including phosphorylated amino acids in the study was that in particular phosphorylated serines and threonines are known to be of relevance for peptides interacting with the 14-3-3 protein from previous work [14].

Homology models of both isoforms were aligned sequentially using the Blosum62 matrix and subsequently structurally aligned using the MOE protein alignment tool. The ligand template formed by 22 alanines in an extended conformation was positioned inside the binding groove, with sufficient space to prevent clashing at posterior mutation steps and optimization (the resulting structure can be found in the supporting material). In this orientation the peptide extends from the binding groove to the regions that have been identified as possible interaction sites due to their low desolvation energy to allow peptide selectivity to be achieved in the optimization step [14].

2.4 Estimation of Peptide Binding Energy

The MOE Protonate 3D function was used to assign ionization states and position hydrogen atoms in the macromolecular structure. Subsequently the MM function of MOE was employed to perform potential energy minimization by use of the AMBER99 force field. Finally, the Potential function was used to evaluate the resulting potential energy of the complex.

2.5 Molecular Search Space and Landscape Analysis

In combinatorial search correlated landscapes neighborhoods are typically induced by a set of small mutations. However, not all neighborhoods can be explored with an EA that uses consecutive mutations to find better fitness value. An important requirement for EA to work is the correlation between parent solutions and offspring solutions in fitness space, called the 'causality requirement'. Using landscape analysis it is possible to get indications on the causality of the search space and the difficulty for optimization [35]. This requirement was assessed empirically in a preliminary study.

In our case the set of configurations or solutions is the set of all sequences of a 22-mer peptide sequence that can be built from 23 possible amino acids, and neighbors are given by solutions that differ in only one amino acid. The correlation and other properties such as ruggedness of the molecular landscape was assessed with the MOE forcefield fitness function in combination with random walks on this surface, based on previous work [29,35]. This study indicated a positive correlation of the fitness function and the proximity of solutions in search space, measured as Hamming distance. A positive correlation was observed up to thirty random steps, which indicated a causal relation between parents and offspring fitness for the given mutation type. Hence it could be seen as promising to perform evolutionary optimization in this search space.

2.6 Multiobjective Optimization

The SMS-EMOA algorithm was used as a multi-objective evolutionary optimization algorithm [28,33]. The instantiation of this algorithm consisted of ten parents and one offspring ((10+1)-SMS-EMOA). A population of ten peptides was maintained throughout the run. In each of the iterations a new sequence was generated by mutating the least recently changed peptide from this population. The new peptide was generated by randomly replacing a residue of the peptide at a random position with a random new amino acid. The potential energy of the complex of the ligand with the γ1 isoform was considered as a first objective function, and the inverse of the potential energy of the complex of the ligand with the ε1 isoform was considered as the second objective function. The inverse was taken since binding against this isoform was not desired and the standard implementation of SMS-EMOA aims at minimization of objectives.

The acronym SMS-EMOA stands for **S**-**M**etric **S**election **E**volutionary **M**ultiobjective **O**ptimization **A**lgorithm. As indicated in the name, its selection is based on the S-Metric, a metric for measuring the quality of a Pareto front approximation which does not require a-priori knowledge of the true Pareto front. The S-Metric is nowadays more commonly referred to as the hypervolume indicator. It measures the size (area in two dimensions, hypervolume in higher dimensions) of subspace that is dominated by a Pareto front approximation and cut from above by a reference point. A high value of the hypervolume indicator corresponds to a good approximation to the Pareto front. The hypervolume contributions are also positively correlated with the distance between neighbors. Hence its maximization promotes diversity of solutions on the

Pareto front. In its selection, either a dominated solution or otherwise one with lowest hypervolume contribution is removed from the population [33]. While the SMS-EMOA specifies the algorithmic details for the selection step, it is generic in terms of search space representation and variation operators. SMS-EMOA is well suited for Pareto front approximation in large search spaces and small population sizes, where the goal is to find well spread Pareto front approximations with relatively few evaluations, such as in the present problem. The reason for this is that SMS-EMOA concentrates the distribution in the so-called 'knee-point' regions of the Pareto front, where good compromise solutions are found, while representing regions with an unbalanced trade-off with a decreased density of points.

The structural energy minimization and evaluation of potential energy of the isoform complexes with the mutated template took on average 4 minutes per complex (8 minutes per iteration for both isoforms). The computational overhead of the internal operations performed in SMS-EMOA is negligible in case of two and three objective functions. More precisely, all hypervolume computations for a single iteration require only subquadratic time. Hence the computational effort is essentially determined by the number of objective function evaluations.

3 Results

The SMS-EMOA implementation in the Molecular Operating Environment (MOE) (available from the authors on request) evaluated 1,089 random mutations (and evaluations of the objective function) after one week of processing time on four Xeon 2.5 GHz processors machine with Scientific Linux. 1,089 mutations correspond to 1,089 / 23^22 × 100 ≈ 1.19954208 × 10^-25 per cent of all possible solutions; however even with this small number of evaluations the Pareto front (X/Y axes: potential energy of interaction with γ1/ ε1) took its characteristic J-shape (a line bending towards the optimizing direction) after about 100 iterations (Figure 1A).

To evaluate the behavior of the algorithm properly it is important to present the development of peptide fitness as a function of time. Figure 1A hence visualizes the Pareto front obtained after 100 iterations of the algorithm. For the 10 peptides that formed part of the Pareto front the number of substitutions of the initial alanine amino acids varied from one to seven while the energy of the complex varied from the original -142 to -196 kcal/mol for the γ1 isoform. The solution of the first Pareto front already shows considerable improvement over the starting point clearly a big advancement since already after eight replacements of alanine amino acids a difference in binding between the isoforms of 49 kcal/mol was obtained (detailed numbers regarding the evolution of the Pareto front are given in the supplementary material).

Figure 1B contains the corresponding Pareto front after 700 evaluations. At this point most of the alanine residues from the initial template were replaced. The potential energy difference of the complex with the 14-3-3 γ1 isoform, compared to the ε1 isoform, varied from 7 kcal/mol to 291 kcal/mol in this case while he best potential energy upon binding the γ1 isoform has reached -282 kcal/mol.

Finally, at the end of our SMS-EMOA run, Figure 1C displays the potential energy along the Pareto front after a total of 1,189 iterations. We observe a set of 9 Pareto-optimal solution and 1 dominated solution with energy of the complex ranging from -138 to -335 kcal/mol for γ1 and 125 to -284 kcal/mol for ε1. Hence, the algorithm was successful in navigating a very high-dimensional (22-dimensional) search space to arrive to peptides of interest for both optimization criteria.

Figure 2 and Table 2 show the energy differences for ε1 and γ1 isoform binding for the final Pareto front. Throughout all solutions, at least an energy difference of about 40 kcal/mol is maintained, which grows to 291 kcal/mol in case of the most selective peptide listed at position 9. The solution at position 7 might be the one most relevant in practice, since it exhibits only minimal binding to the ε1 isoform (free energy of 23 kcal/mol), while binding relatively tightly to the γ1 isoform (free energy of 247 kcal/mol).

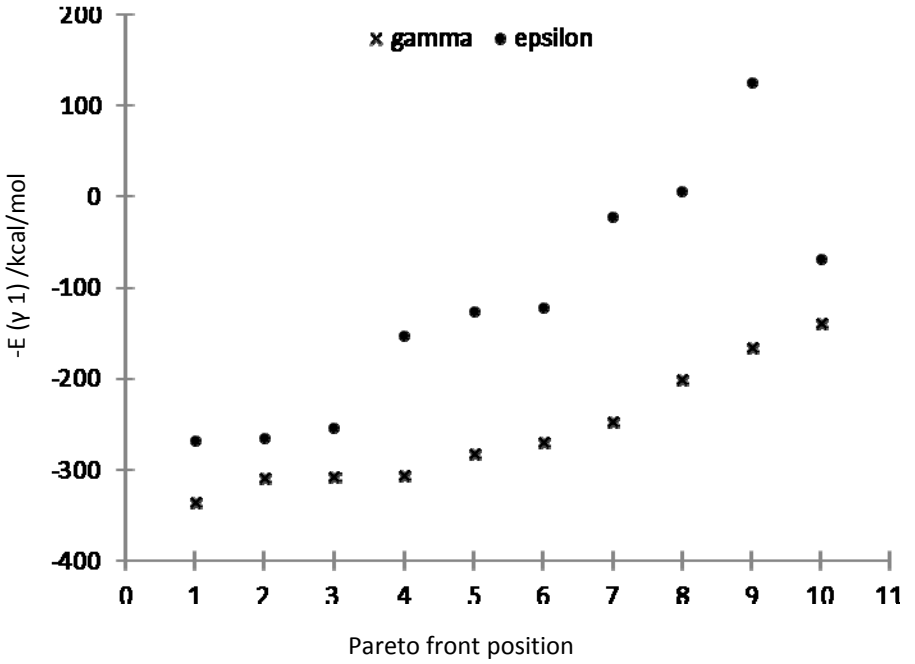

Fig. 2. Potential energy of peptide binding to the ε1 and γ1 isoforms of the 14-3-3 protein after 1,189 iterations of the SMS-EMOA algorithm. The x axis corresponds to the ligand position on the final Pareto front approximation from the leftmost to the rightmost solution in Figure 1C while the y axis represents the potential energy of the complex. It can be seen that multiple trade-offs between affinity and selectivity can be chosen, with solution 7 representing probably a solution of relevance in practice; high selectivity while at the same time high affinity to the γ1 isoform is maintained.

Table 2. Sequence and potential energy of the final solutions obtained after 1,189 iterations of the SMS-EMOA algorithm, optimizing the difference in the potential energy between the ε1 and γ1 isoforms. The ligand position corresponds to the order from the leftmost to the rightmost solution in Figure 1C on the final Pareto front, with all binding energies also visualized in Figure 2.

Aminoacid Sequence																				ligand #	gamma1	epsilon1	difference
PHE	ILE	TPO	ARG	SEP	GLY	TYR	SER	TPO	TRP	ASP	ASN	ARG	ARG	TYR	ARG	TYR	SEP	ASN	ASN	ALA	ALA	1	-335	-269	67
PHE	ILE	TPO	ARG	SEP	GLY	GLY	SER	TPO	ALA	ASP	ASN	ARG	ARG	TYR	LEU	TYR	MET	ASN	ASN	ALA	ALA	2	-309	-266	43
PHE	ILE	TPO	ARG	SEP	GLY	TYR	SER	TPO	ALA	ASP	ASN	ARG	ARG	TYR	ARG	TYR	VAL	ASN	ASN	ALA	ILE	3	-308	-254	54
PHE	ILE	TPO	ARG	SEP	GLY	TYR	SER	TPO	ALA	ASP	ASN	ARG	ARG	TYR	LEU	TYR	MET	ASN	ASN	ALA	ALA	4	-306	-154	152
LEU	TRP	TPO	ARG	SEP	GLY	TRP	ASN	TPO	ALA	ASP	ASN	PRO	ARG	GLU	ARG	TYR	MET	ASN	ASN	ALA	ALA	5	-283	-126	157
LEU	PHE	TPO	ARG	SEP	GLY	TRP	ASN	TPO	ALA	ASP	ASN	PRO	ARG	GLU	ARG	TYR	MET	ASN	ASN	ALA	ALA	6	-269	-123	147
LEU	PHE	TPO	CYS	SEP	GLY	TYR	ASN	TPO	ALA	SER	ASN	LEU	GLN	GLU	ARG	ALA	MET	ASN	ASN	ALA	ALA	7	-247	-23	224
ALA	VAL	ALA	CYS	SEP	GLY	TRP	ASN	TPO	ALA	SER	ASN	ALA	ALA	GLU	ALA	ALA	MET	TPO	ASN	ALA	ALA	8	-202	5	207
LYS	TRP	ALA	ALA	ALA	TRP	ALA	ALA	ALA	ALA	ALA	PRO	ALA	ALA	LEU	ALA	ALA	ALA	ALA	ALA	ALA	ALA	9	-166	125	291
LYS	TRP	ALA	ALA	ALA	TRP	ALA	ALA	ALA	ALA	ALA	PRO	ALA	ALA	LEU	PHE	ALA	ALA	ALA	ALA	ALA	ALA	10	-139	-69	69

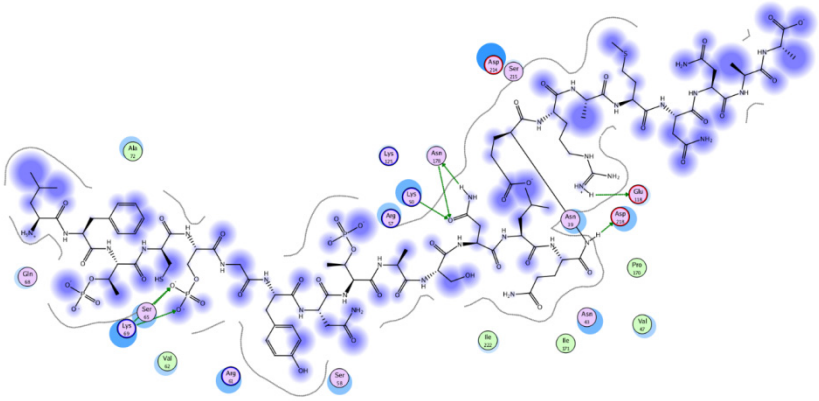

Fig. 3. Ligand interaction plot of solution 7 from the Pareto front with the γ1 isoform of 14-3-3. Salt bridges between the phosphoserine residue in the peptide and Lys69 are formed, which is in agreement with interactions seen in crystal structures for 14-3-3 ligands. A charge interaction between Glu118 and an arginine residue in the peptide ligands results in strong interactions. These are supplemented by hydrogen bonds between an asparagine residue in the ligand and Lys50 and Asn178 in the protein. The resulting free energy of binding is -247 kcal/mol.

Among the most frequent interactions present in the solutions are those with Arg61, Lys69, Asn178 and Asp218. This is in agreement with literature since Arg61 and Lys69 are located above the commonly accepted binding pocket, and Asp218 is located below the binding pocket at the sites predicted by other studies of human 14-3-3 [14]. Asn178 on the other hand is located very close to the binding pocket and may also be involved in recognition of natural ligands as well.

In order to understand ligand selectivity better, the interactions for solution 7 from the final Pareto front shall be discussed here in more detail.

As can be seen, in the ligand complexed with the γ1 isoform of 14-3-3 (displayed in Figure 3) salt bridges between the a phosphoserine residue in the peptide and Lys69 are formed, which is in agreement with interactions seen in crystal structures for 14-3-3 ligands, as well as a charge interaction between Glu118 and an arginine

residue in the peptide ligands, resulting in strong interactions. These are supplemented by hydrogen bonds between an asparagine residue in the ligand and Lys50 and Asn178 in the protein, resulting in a free energy of binding of -247 kcal/mol. On the other hand, binding interactions with the ε1 isoform (Figure 4) are much weaker, leading only to a free energy of binding of -23 kcal/mol. While the salt bridge of Lys69 to the phosposerine is retained, Glu118 is not able to form an electrostatic interaction with the arginine residue of the ligand anymore. Additional hydrogen bonds such as the one to Asp216 are formed; however they are on average weaker than those in the γ1 complex, resulting in a decrease in binding affinity.

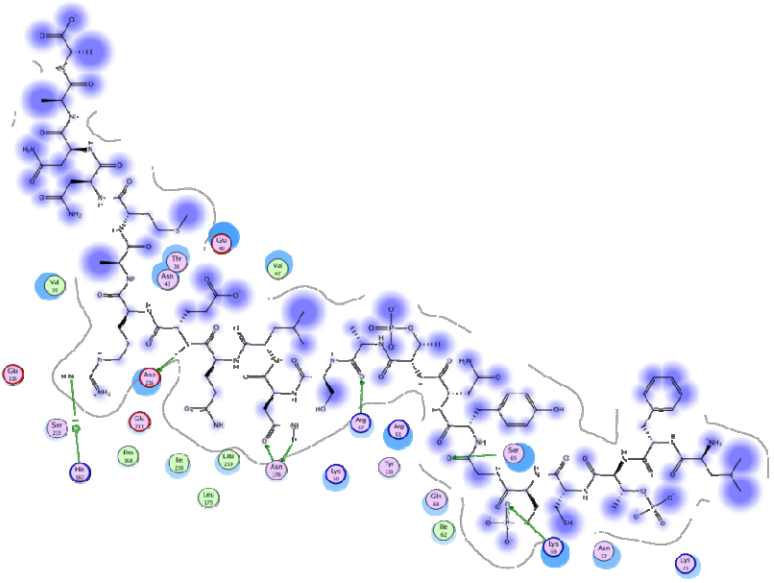

Fig. 4. Ligand interaction plot of solution 7 from the final Pareto front with the ε1 isoform of 14-3-3. As compared to the γ1 isoform (Figure 3) interactions are much weaker, leading only to a free energy of binding of -23 kcal/mol. While the salt bridge of Lys69 to the phosposerine is retained, Glu118 is not able to form an electrostatic interaction with the arginine residue of the ligand anymore. Additional hydrogen bonds such as the one to Asp216 are formed, but they are on average weaker than those in the γ1 complex, resulting in a decrease in binding affinity.

Hence, by analyzing binding interactions we can also rationalize peptide ligand selectivity, leading to an increase of the trustworthiness of the optimization algorithm applied in this work to design isoform-selective ligands for the 14-3-3 protein.

4 Conclusions

By employing evolutionary multiobjective optimization in the form of an SMS-EMOA algorithm we were able to design, de novo, peptide ligands of the γ1 isoform

of the 14-3-3 protein with predicted selectivity over the ε1 isoform. Given the 22-dimensional nature of the search space, this is a practical application of this type of algorithm which will be experimentally validated in the near future.

Acknowledgements. We would like to thank Gerard van Westen (LACDR) for computational support. Andreas Bender thanks Dutch Top Institute Pharma (TI Pharma) for funding. Michael Emmerich thanks LIACS, Leiden University and the Foundation for Science and Technology (FCT), Portugal. Grant: Set-Indicator Based Multiobjective Optimization (SIMO), SFRH/BPD/65330/2009 for financial support. Hernando Sanchez Faddiev thanks Pieter de Knijf and Rudi Westendorp for their support.

Authors' Contributions. Hernando Sanchez Faddiev performed the implementation of the optimization algorithm and performed all computational studies. Andrew Henry supplied support for MOE software and provided with libraries and routines indispensable for the algorithm completion. The work was initiated by Herman Spaink, supervised jointly by Michael Emmerich (multiobjective algorithm) and Andreas Bender with scientific input and support from Herman Spaink and Fons Verbeek and Herman Van Vlijmen.

References

1. Bates, M.: The Origin of Species - by Means of Natural-Selection or the Preservation of Favored Races in the Struggle for Life - Darwin, C. American Anthropologist 61, 176–177 (1959)
2. Back, T., Fogel, D.B., Michalewicz, Z.: Handbook of Evolutionary Computation. IOP Publishing Ltd. (1997)
3. Bender, A., Scheiber, J., Bender, A., Glick, M., Davies, J.W., et al.: Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. Chemmedchem 2, 861–873 (2007)
4. Macchiarulo, A., Nobeli, I., Thornton, J.M.: Ligand selectivity and competition between enzymes in silico. Nature Biotechnology 22, 1039–1045 (2004)
5. Kalyanmoy, D.: Multi-Objective Optimization using Evolutionary Algorithms (2001)
6. van der Horst, E., Peironcely, J.E., IJzerman, A.P., Beukers, M.W., Lane, J.R., et al.: A novel chemogenomics analysis of G protein-coupled receptors (GPCRs) and their ligands: a potential strategy for receptor de-orphanization. Bmc Bioinformatics 11 (2010)
7. Bender, A., Spring, D.R., Galloway, W.R.J.D., Overington, J.P., van Westen, G.J.P., et al.: Chemogenomics Approaches for Receptor Deorphanization and Extensions of the Chemogenomics Concept to Phenotypic Space. Current Topics in Medicinal Chemistry (2010)
8. van Westen, G.J.P., Wegner, J.K., IJzerman, A.P., van Vlijmen, H.W.T., Bender, A.: Proteochemometric Modeling as a Tool for Designing Selective Compounds and Extrapolating to Novel Targets (2010)
9. Fu, H.A., Subramanian, R.R., Masters, S.C.: 14-3-3 proteins: Structure, function, and regulation. Annual Review of Pharmacology and Toxicology 40, 617–647 (2000)
10. Wilker, E., Yaffe, M.B.: 14-3-3 Proteins - a focus on cancer and human disease. Journal of Molecular and Cellular Cardiology 37, 633–642 (2004)

11. Dong, S., Kang, S., Lonial, S., Khoury, H.J., Viallet, J., et al.: Targeting 14-3-3 sensitizes native and mutant BCR-ABL to inhibition with U0126, rapamycin and Bcl-2 inhibitor GX15-070. Leukemia 22, 572–577 (2008)
12. Jayaratnam, S., Khoo, A.K., Basic, D.: Rapidly progressive Alzheimer's disease and elevated 14-3-3 proteins in cerebrospinal fluid. Age Ageing 37, 467–469 (2008)
13. Besser, J., Bagowski, C.P., Salas-Vidal, E., van Hemert, M.J., Bussmann, J., et al.: Expression analysis of the family of 14-3-3 proteins in zebrafish development. Gene. Expr. Patterns 7, 511–520 (2007)
14. Yang, X., Lee, W.H., Sobott, F., Papagrigoriou, E., Robinson, C.V., et al.: Structural basis for protein-protein interactions in the 14-3-3 protein family. Proc. Natl. Acad. Sci. U S A 103, 17237–17242 (2006)
15. Audie, J., Boyd, C.: The Synergistic Use of Computation, Chemistry and Biology to Discover Novel Peptide-Based Drugs: The Time is Right. Current Pharmaceutical Design 16, 567–582 (2010)
16. Belda, I., Madurga, S., Llora, X., Martinell, M., Tarrago, T., et al.: ENPDA: an evolutionary structure-based de novo peptide design algorithm. Journal of Computer-Aided Molecular Design 19, 585–601 (2005)
17. Abe, K., Kobayashi, N., Sode, K., Ikebukuro, K.: Peptide ligand screening of alpha-synuclein aggregation modulators by in silico panning. Bmc Bioinformatics 8 (2007)
18. Zahed, M., Suzuki, T., Suganami, A., Sugiyama, H., Harada, K., et al.: Screening of SMG7-Binding Peptides by Combination of Phage Display and Docking Simulation Analysis. Protein and Peptide Letters 16, 301–305 (2009)
19. Gillet, V.J.: Applications of evolutionary computation in drug design. Applications of Evolutionary Computation in Chemistry 110, 133–152 (2004)
20. Nicolaou, C.A., Apostolakis, J., Pattichis, C.S.: De Novo Drug Design Using Multiobjective Evolutionary Graphs. Journal of Chemical Information and Modeling 49, 295–307 (2009)
21. Keijzer, M.: Genetic and evolutionary computation conference: GECCO 2006, vol. 2. Association for Computing Machinery, New York (2006)
22. Malard, J.M., Heredia-Langner, A., Cannon, W.R., Mooney, R., Baxter, D.J.: Peptide identification via constrained multi-objective optimization: Pareto-based genetic algorithms. Concurrency and Computation-Practice & Experience 17, 1687–1704 (2005)
23. Yagi, Y., Terada, K., Noma, T., Ikebukuro, K., Sode, K.: In: silico panning for a non-competitive peptide inhibitor. Bmc Bioinformatics 8 (2007)
24. Fjell, C.D., Jenssen, H., Cheung, W.A., Hancock, R.E., Cherkasov, A.: Optimization of Antibacterial Peptides by Genetic Algorithms and Cheminformatics. Chem. Biol. Drug. Des. (2010)
25. Li, C., Pazgier, M., Li, C.Q., Yuan, W.R., Liu, M., et al.: Systematic Mutational Analysis of Peptide Inhibition of the p53-MDM2/MDMX Interactions. Journal of Molecular Biology 398, 200–213 (2010)
26. Satoh, J., Yamamura, T., Arima, K.: The 14-3-3 protein epsilon isoform expressed in reactive astrocytes in demyelinating lesions of multiple sclerosis binds to vimentin and glial fibrillary acidic protein in cultured human astrocytes. American Journal of Pathology 165, 577–592 (2004)
27. Roberts, M.R., de Bruxelles, G.L.: Plant 14-3-3 protein families: evidence for isoform-specific functions? Biochemical Society Transactions 30, 373–378 (2002)
28. Beume, N., Naujoks, B., Emmerich, M.: SMS-EMOA: Multiobjective selection based on dominated hypervolume. European Journal of Operational Research 181, 1653–1669 (2007)

29. Emmerich, M., Li, B.V.Y., Bender, A., Sanchez-Faddiev, H., Kruisselbrink, J., et al.: Analyzing molecular landscapes using random walks and information theory. Chemestry Central (2009)
30. Kruisselbrink, J.W., Aleman, A., Emmerich, T.M., IJzerman, A., Bender, A., et al.: Enhancing search space diversity in multi-objective evolutionary drug molecule design using niching. In: Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation, pp. 217–224. ACM, Montreal (2009)
31. Handl, J., Kell, D.B., Knowles, J.: Multiobjective optimization in bioinformatics and computational biology. IEEE-ACM Transactions on Computational Biology and Bioinformatics 4, 279–292 (2007)
32. Paul, A.L., Sehnke, P.C., Ferl, R.J.: Isoform-specific subcellular localization among 14-3-3 proteins in Arabidopsis seems to be driven by client interactions. Molecular Biology of the Cell 16, 1735–1743 (2005)
33. Emmerich, M., Beume, N., Naujoks, B.: An EMO Algorithm Using the Hypervolume Measure as Selection Criterion. In: Coello Coello, C.A., Hernández Aguirre, A., Zitzler, E. (eds.) EMO 2005. LNCS, vol. 3410, pp. 62–76. Springer, Heidelberg (2005)
34. Vilar, S., Cozza, G., Moro, S.: Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery. Current Topics in Medicinal Chemistry 8, 1555–1572 (2008)
35. Vassilev, V.K., Fogarty, T.C., Miller, J.F.: Information Characteristics and the Structure of Landscapes. Evolutionary Computation (2000)