Identification of novel protein scaffolds for small molecules binding and for catalysis

(Settore scientifico-disciplinare: Anatomia comparata e Citologia - BIO/06)

Dottoranda: Immacolata Luisi

Coordinatore del collegio dei docenti
Chiar.mo prof. Maurizio Fermeglia
Università degli studi di Trieste

Supervisore
Prof. Federico Berti
Università degli studi di Trieste

Tutore
Chiar.mo prof Roberto Marzari
Università degli studi di Trieste

Co-Tutore
Chiar.mo prof. Daniele Sblattero
Università degli studi del Piemonte Orientale

ANNO ACCADEMICO 2010/2011
Table of contents

TABLE OF CONTENTS .................................................................................. I

ABSTRACT ........................................................................................................... V

RIASSUNTO ........................................................................................................ VII

LIST OF ABBREVIATIONS .................................................................................. 1

1 INTRODUCTION .............................................................................................. 2

1.1 PROTEIN SCAFFOLDS ................................................................................. 2
  1.1.1 STRUCTURAL CLASSIFICATION .......................................................... 3
  1.1.1.1 Scaffold formed by α-helices ......................................................... 4
  1.1.1.2 Scaffold formed by β-sheets ...................................................... 6
  1.1.1.3 Irregular scaffolds ........................................................................ 9
  1.1.2 COILED COIL DOMAINS ................................................................. 12
    1.1.2.1 Electrostatic interactions ............................................................... 16
    1.1.2.2 Orientation and oligomerization .............................................. 17
    1.1.3 E-K COILED-COIL ........................................................................ 17

1.2 HUMAN SERUM ALBUMIN ........................................................................ 22
  1.2.1 STRUCTURE AND PROPERTIES ...................................................... 22
  1.2.2 BINDING SITES ................................................................................... 26
    1.2.2.1 Site I ............................................................................................. 28
    1.2.2.2 Site II ........................................................................................... 31
    1.2.2.3 Other binding sites .................................................................... 32
  1.2.3 HSA BINDING INTERACTIONS .......................................................... 32
  1.2.4 ENZYMATIC PROPERTIES ................................................................. 34
  1.2.5 FLEXIBILITY ....................................................................................... 36
  1.2.6 CONFORMATIONAL EQUILIBRIUM .................................................. 37
  1.2.7 HSA RELATED DISEASE .................................................................... 38
  1.2.8 SPECTROSCOPIC FEATURES .............................................................. 39
1.2.9 UV SPECTRAL PROPERTIES .......................................................... 39
1.2.9.1 Circular dichroism ................................................................. 40
1.2.9.2 Fluorescence ........................................................................... 41
1.3 BIOSENSORS TRANSDUCERS ......................................................... 42
1.3.1 MAIN FEATURES ........................................................................ 42
1.3.2 UTILITY OF BIOSENSORS ......................................................... 43
1.3.3 MOLECULAR RECOGNITION ELEMENTS .................................... 44

2 AIM OF THE RESEARCH .................................................................... 47

3 MATERIALS AND METHODS ............................................................... 48

3.1 ABBREVIATIONS ............................................................................. 48
3.2 MATERIALS ...................................................................................... 48
3.3 SOLUTIONS AND BUFFERS ........................................................... 49
3.4 BACTERIAL STRAINS ...................................................................... 50
3.5 OLIGONUCLEOTIDES ................................................................... 51
3.6 EXPERIMENTAL METHODS ........................................................... 52
3.6.1 PCR (POLYMERASE CHAIN REACTION) ..................................... 52
3.6.2 DNA ELECTROPHORESIS ON AGAROSE GEL ......................... 53
3.6.3 DNA PURIFICATION .................................................................. 53
3.6.4 DNA DIGESTION WITH RESTRICTION ENDONUCLEASES .......... 53
3.6.5 DNA LIGATION ........................................................................... 54
3.6.6 PLASMIDIC DNA EXTRACTION ............................................... 54
3.6.7 DNA SEQUENCING .................................................................... 54
3.6.8 PREPARATION OF COMPETENT E. COLI CELLS ..................... 55
3.6.9 BACTERIAL TRANSFORMATION ............................................... 55
3.6.10 PREPARATION OF ELECTROCOMPETENT CELLS ............... 55
3.6.11 BACTERIAL TRANSFORMATION OF ELECTROCOMPETENT CELLS 56
3.6.12 PREPARATION OF BACTERIA FOR -80°C STOCK ................... 56
3.6.13 PREPARATION AND TRANSFORMATION OF PDAN-sCE-KINV RANDOM LIBRARY 56
3.6.14 EXPRESSION OF RECOMBINANT PROTEINS .......................... 58
3.6.15 PRODUCTION AND PEG-PRECIPITATION OF PHAGES .......... 58
3.6.16 LIBRARY SELECTION WITH MAGNETIC BEADS .................. 59
3.6.17 PHAGE ELISA ......................................................................... 59
3.6.18 CELL CULTURE: CELL LINES AND MEDIA ............................. 60
3.6.19 TRANSFECTION ....................................................................... 60
3.6.20 Post-transfection analysis, polyclonal and monoclonal selection and expansion .......................................................... 61
3.6.21 Productivity and specificity ELISA assay ............................................. 62
3.6.22 Cells freezing .................................................................................. 63
3.6.23 HSA100-GST protein production, purification and dialysis .......... 63
3.6.24 Protein electrophoresis on polyacrylamide gel (SDS-PAGE) ......... 64
3.6.25 Coomassie staining ......................................................................... 65
3.6.26 Western blot ................................................................................... 65
3.6.27 Bradford protein assay ................................................................. 65
3.6.28 HSA100-GST ELISA assay ............................................................. 66
3.6.29 LC-MS/MS identification of HSA100-GST .................................. 66
3.6.30 CD spectroscopy ........................................................................... 67
3.6.31 Fluorescence assay ........................................................................ 67
3.6.32 SPR analysis .................................................................................. 68
3.6.33 HPLC analyses ............................................................................. 68
3.6.34 Reduction of diketone 3 ................................................................. 69
3.6.35 Aldol reaction ................................................................................. 69
3.6.36 Preparation and transformation of pGex-HSA100 mutant libraries... 70

4 RESULTS AND DISCUSSION ........................................................................ 72

4.1 Stable scaffold based on EK coiled-coil ............................................. 72
  4.1.1 Single chain E-Kinv design ............................................................. 73
  4.1.2 Library creation by random mutagenesis ....................................... 81
  4.1.3 scE-Kinv random library characterization ...................................... 84
  4.1.4 Selection by phage display system versus small molecules .......... 85
  4.1.5 Selection and validation vs caffeine .............................................. 88
  4.1.6 Measurement of the dissociation constant of the peptide/caffeine complexes ......................................................... 90
  4.1.7 Characterization of bE1 and bE8 sequence ..................................... 91
  4.1.8 Selection and validation vs aldosterone ....................................... 94
  4.1.9 Single chain E_{G7}K_{G7} FC fusion protein .................................... 95

4.2 A binder and catalytic peptide from a natural host: Human Serum Albumin ........................................................................ 100
  4.2.1 HSA100 identification and optimization ...................................... 100
  4.2.2 HSA100 optimization strategy ...................................................... 102
  4.2.3 Structural characterization of HSA100 .......................................... 108
4.2.4 SMALL MOLECULE BINDING ACTIVITY OF HSA100 ............................................. 110
4.2.4.1 ELISA assay ........................................................................................................ 111
4.2.4.2 Fluorescence assay ............................................................................................... 112
4.2.5 CHEMICAL REACTIVITY CONTROL: DIKETONE REDUCTION .......................... 115
4.2.6 CHEMICAL REACTIVITY CONTROL: ALDOLASE ACTIVITY ............................ 116
4.2.7 BIOSENSING APPLICATION: SPR ........................................................................ 117
4.3 TOWARDS A LIBRARY OF HSA100 BASED RECEPTOR ...................................... 123
4.3.1 STRUCTURAL ANALYSIS ....................................................................................... 123
4.3.2 HSA100 SITE-DIRECTED MUTAGENESIS ............................................................ 126
4.3.3 PRODUCTION AND PURIFICATION OF HSA100 MUTANTS ............................... 129
4.3.4 BINDING ACTIVITY OF THE MUTANT CLONES ................................................. 130
4.3.4.1 ELISA assay ........................................................................................................ 131
4.3.4.2 SPR ...................................................................................................................... 132
4.3.4.3 Fluorescence assay ............................................................................................... 133

CONCLUSION .................................................................................................................. 135

REFERENCES ................................................................................................................. 136
Abstract

Protein scaffolds are stable structures capable to recognize and bind, in different conditions, small guest molecules. They are proteins with known conformation which can be used and modified for the construction of variants. Scaffolds include various structural types (α-helices, β-sheets, mixed structures), which differ in stability, tolerance to multiple substitutions, ease of expression and subsequent applications. Short, structurally organized peptides have been also used as artificial catalysts.

The goal of this work is the identification of peptide-based artificial receptors or catalysts. To this purpose, we have considered two different protein motifs to generate new scaffolds: a synthetic E/K coiled-coil domain, and one of the binding sites of the natural protein Human serum albumin. In order to generate stable peptide hosts, we developed peptide libraries to be selected for both properties, adopting two different approaches: RANDOM mutagenesis and SITE-DIRECTED mutagenesis.

In the first approach we have randomly mutated the central aminoacid heptad of the 35 aminoacid peptides E and K, to create a pocket that can binds small molecules. We have generated a $2 \times 10^8$ members library of a new scaffold named single chain E-Kinv. random. Using the phage display selection system, we have identified peptides (bE1, bE8 and G7) that bind targets as xanthines and aldosterone. We have also expressed the binder peptides over antibody constant regions, to obtain a single chain Fc fusion protein that retains the binding capability.

In the second part of this work, the Sudlow binding site of Human albumin has been used to identify a one hundred amino acids peptide, named HSA100, through a structural and computational analysis. The peptide was obtained as GST fusion protein in E. coli in yields as high as 4 mg/ml, by codon usage optimization. This newly cloned polypeptide has been tested for its ability to interact with drugs through biological and chemical assays, and for its ability to catalyze organic reactions. The fusion protein retains its structural integrity and aldolase activity, the ability to direct the stereochemical outcome of a diketone reduction, and its binding capacity to warfarin and efavirenz. GST-HSA100 was used as the starting scaffold for the construction of libraries of
mutant binders and catalysts. We have considered the entry and ligand contact sites in the two hydrophobic pockets. The positions chosen for mutation are seven, recognized as non essential for correct folding by bioinformatics analysis of sequence conservation within the set of all Serum Albumins. Two small mutants libraries were generated, and diversity was effectively introduced in the targeted positions of both the minor and the major hydrophobic pocket. We obtained the production of the mutated clones at concentration of 4 mg/ml and tested their binding activity. As a preliminary test, quantitative affinity data were measured on several, unselected clones carrying mutations in the major hydrophobic pocket.
Riassunto

Gli scaffold proteici sono strutture stabili capaci di riconoscere e legare piccole molecole. Si tratta di strutture proteiche a conformazione nota utilizzabili e modificabili per la costruzione di varianti. Gli scaffold comprendono varie tipologie classificabili in base alla loro struttura (α-eliche, β-foglietti, piccoli scaffold con strutture miste), che differiscono in termini di stabilità, tolleranza verso sostituzioni multiple, facilità di espressione e successive applicazioni. Sono stati utilizzati piccoli peptidi strutturalmente organizzati come catalizzatori artificiali.

Lo scopo finale di questo lavoro è l’identificazione di recettori artificiali o catalizzatori basati su piccoli peptidi. Per fare ciò sono stati considerati due diversi domini proteici per generare un nuovo scaffold: il dominio sintetico E/K coiled-coil e un sito di legame della proteina naturale Human serum albumin.

Per generare strutture stabilì sono state disegnate e generate delle librerie peptidiche per essere selezionate per lo scopo proposto, adottando due approcci diversi: mutagenesi RANDOM e mutagenesi SITO-DIRETTA.

Nel primo approccio è stata mutata in maniera random solo l’eptade centrale di ciascun peptide E e K (ognuno costituito da 5 eptadi ripetute, 35 aminoacidi), in modo da creare una tasca in grado di legare piccole molecole. E’ stata generata una libreria di $2 \times 10^8$ cloni chiamata single chain E-Kinv. random. Mediante la tecnologia del phage display sono stati identificati peptidi random (bE1, bE8 and G7) che legano target quali xantine e aldosterone. Questi peptidi selezionati sono stati inoltre espressi in fusione con le regioni costanti degli anticorpi, per ottenere una proteina di fusione a singola catena che mantiene la sua capacità di legame (scE$_{G7}$-K$_{G7}$ minibody).

Nella seconda parte di questo lavoro di tesi, il sito di legame Sudlow dell’albumina umana è stato utilizzato per identificare un peptide di 100 amminoacidi, detto HSA100, mediante un’analisi strutturale e computazionale. Il peptide è stato ottenuto come proteina di fusione con il GST in E. coli con una resa pari a 4 mg/ml, grazie alla strategia di ottimizzazione dei codoni per l’espressione in E. coli. Questo polipeptide GST-HSA100 è stato clonato e testato per la sua capacità di interagire con i farmaci mediante saggi biologici e chimici, e per la sua capacità di catalizzare reazioni organiche. La proteina di
fusione mantiene la sua integrità strutturale e aldolasica, la capacità stereochemica per la reazione di riduzione del dichetone, e la sua capacità di legame per il warfarin e l'efavirenz. Il GST-HSA100 è stato utilizzato come scaffold di partenza per la costruzione di librerie di binders mutanti e catalizzatori. Per questo sono stati considerati l'ingresso e i siti di contatto con il ligando in due tasche idrofobiche. Le posizioni scelte per la mutazione sono sette, riconosciute come non essenziali per il corretto ripiegamento mediante l'analisi bioinformatica di sequenze conservate all'interno di tutte le siero albumine. Due piccole librerie di mutanti sono state generate e la diversità è stata effettivamente introdotta nelle posizioni identificate sia nella tasca idrofobica minore che in quella maggiore. Abbiamo ottenuto la produzione dei cloni mutanti ad una concentrazione fino a 4 mg/ml ed è stata testata la loro attività di legame. In un test preliminare sono stati misurati i dati quantitativi di affinità su diversi cloni selezionati con le mutazioni presenti nella tasca idrofoba maggiore.
List of abbreviations

AP, alkaline phosphatase
APP, amyloid precursor protein
BSA, bovine serum albumin
CBM, carbohydrate binding module
CD, circular dichroism
CDR, complementary determining region
CMPF, 3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid
CSB, Cystein stabilized β-sheet
CSH, Cystein stabilized α-helix
CTLA-4, Citotoxic T-lymphocyte associated protein 4
dsDNA, double strand DNA
ELISA, enzyme linked immunosorvent assay
Eran, E-coil random
Ewt, E-coil wild type
FA, fatty acid
Fc, fragment crystallizable region
10Fn3, 10th fibronectine type III domain
GCN4, general control non-repressed 4
GFP, green fluorescent protein
GST, glutathione S-transferase
HSA, human serum albumin
Ig, immunoglobulin
Im7, immune innate protein 7
IPTG, isopropyl β-D-1 thiogalactopyranoside
Kinv, K-coil inverted
Kran, K-coil random
Kwt, K-coil wild type
LCFA, long-chain fatty acid
LMWP, low molecular weight proteins
MS, mass spectroscopy
MW, molecular weights
O/N, over night
p8MTCP1, protein mature T-cell proliferation 1
PCA, protein complementation assay
PCR, polymerase chain reaction
PHD, plant homeodomein finger protein
POC, point of care
RSA, rat serum albumin
RT, room temperature
SPA, staphylococcal protein A
SPR, surface plasmon resonance
ssDNA, single strand DNA
TCR, T-cell receptor
TNF, tumor necrosis factor
wt, wild type
WB, western blot

G, glycine (Gly)
P, proline (Pro)
A, alanine (Ala)
V, valine (Val)
L, leucine (Leu)
I, isoleucine (Ile)
M, methionine (Met)
C, cysteine (Cys)
F, phenylalanine (Phe)
Y, tyrosine (Tyr)
W, tryptophan (Trp)
H, histidine (His)
K, lysine (Lys)
R, arginine (Arg)
Q, glutamine (Gln)
N, asparagine (Asn)
E, glutamic acid (Glu)
D, aspartic acid (Asp)
S, serine (Ser)
T, threonine (Thr)
1 Introduction

Chapter 1

1.1 Protein Scaffolds

Scaffolds are small protein structures with stable and known conformation used and modified for the construction of variants able to recognize and bind small molecules. They are often based on engineered antibodies, however, new types of non-immunoglobulin protein and peptide scaffolds are recently emerging with very advantageous characteristics (Schlehuber and Skerra 2005; Hosse et al. 2006). They are similar to antibodies as they show affinity for the molecules against which they were selected, and the possible applications are also similar.

Scaffolds structures may be based on natural proteins elements which have structural function or enzymatic activity (Vita 1997). They may be either optimized or designed de novo (Causton and Sherman 1999), including reduction to minimal dimension, while keeping structure stability, tolerance to multiple substitutions, subsequent applications and easy expression (Nygren and Skerra 2004).

Binders or catalysts can be selected out from a library of mutated scaffolds both in vivo and in vitro and the selection process may be subject to strategies that allow a certain level of evolution, to make possible the isolation of high affinity reagents (Binz et al. 2005).

Protein scaffold are generally formed by α-helices and/or β-sheets connected by loops. The loops are the regions of the scaffold that most lend to modifications such as substitutions and randomization and they are generally more accessible to the ligand (figure 1.1). They are also able to tolerate large variations without significantly alteration of the basic protein structure and stability. The presence of exposed loops often provides the link between the target molecule and structured and constant region of the scaffold. This link may also be mediated by α-helices themselves (Enander et al. 2004) or occur within cavities or pockets (Lee et al. 2006).
Introduction

Fig. 1.1 - Engineering strategies to modify the binding site of protein scaffolds. The binding site has been altered, introducing the diversification with suitable strategies, as the randomization of specific domains (a, b, c, d). A random sequence has been inserted into the scaffold (e), or modified by error-prone PCR (f). Moreover, pre-existing binding sites can be grafted into a novel scaffold (g), or the binding site can be de-novo designed (h) (Binz et al. 2005).

The combinatorial method used to introduce diversity may vary. The modification may also occur at post-translational level. Permissive position on α-helices or β-sheets can be randomized, to make these areas accessible to the target, or one or more contiguous or non-contiguous loops can be randomized either by rationally restricted strategies or by introducing the full amino acid diversity in the sequence space. (Enander et al. 2004).

It is important to pay attention as mutation in amino acid sequence may lead to scaffold misfolding or instability (Enander et al. 2004).

1.1.1 Structural classification

Scaffolds may be classified into three groups:

1. Scaffold formed by α-helices.
2. Scaffold formed mainly by β-sheets.
3. Small scaffold with few secondary structures or irregular architecture of α-helices and β-sheets.

Disulfide bonds may be present that stabilize the structure by linking different regions of the protein to build up the topology of the binding surfaces,
which may be found in shapes ranging from flexible finger to pockets (McLafferty et al. 1993).

1.1.1.1 Scaffold formed by α-helices

α-helices are the most common structural motif in nature but the number of scaffolds based on this architecture and used for phage display libraries is rather limited. Some examples of this kind of scaffolds are listed below:

**Affibodies** (Juraja et al. 2006): are small proteins consisting of an engineered form of one of the five stable domains (Z domain) of the bacterial receptor protein A, located on the surface of *Staphylococci* and able to bind the Fc portion of immunoglobulins (figure 1.2). From the structural point of view, the Z domain is a three helix bundle with small size (6kDa), high solubility, proteolytic and thermal stability and absence of disulfide bridges. 13 residues on the surface of two of the three helices involved in binding to in-kind Fc are commonly engineered.

This scaffold is considered useful for many biotechnological applications in various fields, such as protein microarrays, thanks to the small size and resistance. They were first used to build libraries and select them by affinity (Hogbom et al. 2003).

**Innate immunity proteins (Im7)**: a known member of this group is Im7, an *E. coli* protein that binds the bacterial toxin ColE7, produced by bacteria itself, inactivating it and therefore protecting *E. coli* from the toxin (Juraja et al. 2006). It consists of a hydrophobic core and four α-helices wrapped around it, linked by two first and second type loops (figure 1.3).
\textbf{\textit{\textbf{a2p8}:}} is a polypeptide of 38 amino acids, derived from the N-terminal region of a human protein mature T-cell proliferation 1 (p8\textsuperscript{MTCP1}, a protein of 8 kDa encoded by human oncogene MTCP1) and consists of two \(\alpha\)-helices forming an alpha-hairpin stabilized by two disulfide bridges (Barthe et al. 2000) (figure 1.4).

\textbf{\textit{Ankyrin repeat domains:}} are one of the main groups of \(\alpha\)-helical scaffolds. Ankyrin domains are formed by a base structural unit, consisting of a sequence of 33 amino acid including a turn followed by an antiparallel coiled-coil and a loop that connects the unit to the next repetition. The repeating patterns allow the use of different strategies for the libraries design because the surface of the binding interface may be varied depending on the number of randomized repetitions (Stumpp et al. 2003). This is a typical example of modular motif exploited for protein-protein interactions (figure 1.5). Such scaffolds are characterized by the lack of disulfide bridges, which allows a cytoplasmic expression in \textit{E. coli} in large quantities. The construction of DARPins (Designed Ankyrin Repeat Proteins) libraries has been described, consisting of four repetitions (Amstutz et al. 2005). The most recent application of this type of libraries concerns the selection of an high affinity inhibitor of aminoglycoside phosphotransferase (3')-IIIa (APH), the enzyme responsible for aminoglycoside antibiotic resistance in pathogenic bacteria.

\textbf{\textit{Amyloid precursor protein (APP):}} is an integral membrane protein expressed in many tissues and mostly in the synaptic areas of neurons. The large extracellular region is divided into two domains E1 and E2. E1 contains several subdomains including the growth factor-like domain (GFLD), a metal binding motif (figure 1.6) and a serine protease inhibitor domain. The E2 domain is a coiled-coil dimer and binds proteoglycans in the extracellular matrix. The fully crystallized structure of APP
has not yet been obtained but individual domains including E2 have been successfully crystallized (Kong et al. 2005).

**b$_{562}$ and b$_{5}$ Cytochromes:** are a small four helical bundle serum protein 5nm long and 2.5nm diameter, discovered in the periplasm of *E. coli*. Cytochromes b$_{562}$ and b$_{5}$ and are simple mono-domain proteins that contain heme as the only prosthetic group, and a methionine and an histidine as the axial iron ligands (Mathews 2011) (figure 1.7). Flavocytochrome b$_{2}$ also contains histidine as iron ligand, but also flavin mononucleotide (FMN) as a second prosthetic group, in a separate domain. Cytochromes b$_{5}$ and b$_{562}$ mediate intracellular electron transfer between redox partners, the former inside the endoplasmic reticulum of mammalian cells and the latter in bacteria.

**General control non-repressed 4 (GCN4):** is a yeast transcriptional activator that regulates the expression of a large number of genes including those involved in the biosynthesis of amino acids (Gonzalez et al. 1996). It has a bZIP structure that binds DNA (figure 1.8). As most DNA binding proteins, GCN4 contains a leucine zipper sequence to bind and activate genes (Steinmetz et al. 2007).

1.1.1.2 **Scaffold formed by β-sheets**

Antibodies are the typical binding molecules with β-sheet structure, that are able to specifically recognize their target antigens through their hypervariable loops called *Complementary Determining Regions* (CDR). There are several known antibody scaffolds formed by β-sheets capable of binding molecules. Besides these, there are some non-immunoglobulin scaffolds
formed by β-sheets that can be used to create and select libraries against specific targets. This category includes:

**Statins:** neocarzinostatin (figure 1.9) is a macromolecular chromoprotein enediyne antibiotic complex with anti-tumoral activity secreted by *Streptomyces macromomyceticus*. This protein consists of seven strands arranged to form a β-sheet sandwich, a structure very similar to immunoglobulins. Two of the loops that compose it, forming a deep pocket where in the native state has hosted the chromophore responsible for the anti-tumor activity. The chromophore is a very potent DNA-damaging agent; it is very labile and the role of the apoprotein is to protect it and release it to the target DNA. These loops can be randomized and used as scaffolds to bind other small molecules as testosterone (Heyd et al. 2003).

**Knottins:** are small proteins of 25-35 amino acids structurally similar to the scorpion toxins, with highly conserved disulfide bridges that stabilize the *Knotted topology* (figure 1.10), their typical tertiary structure, a short triple-stranded β-sheet with 3 disulfide bridges. They also present variable loops that tolerate different changes (Souriau et al. 2005). The best known scaffold belonging to this family is Min-23; this scaffold is one of the most achievable short peptides presenting the CSB (cystein stabilized β-sheet) motif (Souriau et al. 2005).

**Lipocalins:** are scaffolds able to bind small and large molecules in a specific way (Schlehuber and Skerra 2005). Also known as anticalins, they are a family of proteins involved in storage and transport of hydrophobic molecules (Weiss and Lowman 2000) (figure 1.11). They have a cone-shaped β-sheet conformation also called "β-barrel" with eight antiparallel
Introduction

strands connected by four hypervariable loops (Skerra 2001).

10th fibronectine type III domain (10 Fn3): is a small β-sheet domain with loops on the surface that can be randomized and used as scaffolds in phage display (Koide et al. 1998; Xu et al. 2002). A subset of these domains employ an Arg-Gly-Asp (RGD) tripeptide motif to mediate contact with cell-surface receptors (integrins). This motif mediates protein-protein interactions in a wide range of biological processes, such as in tissue development, wound healing and metastasis (Carr et al. 1997) (figure 1.12).

Citotoxic T-lymphocyte associated protein 4 (CTLA-4): also known as CD152, is a receptor belonging to the Ig superfamily expressed on CD4 + and CD8 + recently activated lymphocytes (Pier G.B. 2006). It was used for the first time in 1999 as scaffold due to its ability to tolerate loop substitutions without losing the native conformation (Nuttall et al. 1999) (figure 1.13). Much scaffolds can be produced in soluble form in the E. coli periplasm and are functionally expressed as fusion proteins on phage surface for the Display of Functional Polypeptides.

T-cell receptor (TCR): is a transmembrane receptor located on the surface of T lymphocytes, responsible for the recognition of antigens presented by major histocompatibility complex (MHC). Each TCR chain has two Ig-like domains and therefore belongs to the immunoglobulin superfamily (figure 1.14). A decade ago a monomeric class II MHC molecule in which the peptide was covalently linked to β-chain of class II molecule was generated. This type of molecule had a low-binding affinity and did not cause the multimerization of TCR. The requirement of multimerization of TCR led to development of a new class of reagents, chimeric peptides covalently linked to MHC that was dimerized via

---

Fig. 1.12 - Backbone dynamics of homologous fibronectin type III domains from fibronectin and tenascin (Carr et al. 1997).

Fig. 1.13 - Molscript representation of Cα traces: Blue, NMR structure of CTLA-4 V-domain (PDB-1AH1residues 1A-V115); Adapted from Nuttall et al. 1999).

Fig. 1.14 - Bindingin structure TCR-MHCI.
Fc fragment of an immunoglobulin and linked to 3' end of the β-chain of MHC class II molecule (Goldberg and Bona 2011).

**Green Flourescent Protein (GFP):** is composed of 238 amino acid residues (26.9kDa), has a typical β-barrel structure, consisting of a 11 stranded β-sheet with a coaxial helix (figure 1.15). The external surface of the barrier is made by densely packed amino acids and makes the internal cavity unavailable to solvent, thus allowing several internal amino acids to catalyze specific cyclization reactions in the tripeptide Ser 65-Tyr 66-Gly 67, which leads to the formation of the fluorophore (Ormo et al. 1996).

**Carbohydrate binding module (CBM):** many enzymes that digest polysaccharides have a catalytic domain and a separate module that binds carbohydrates (CBM) (figure 1.16). In terms of both total number of families and entries in databases, the dominant fold among CBMs is the β-sandwich (fold family 1) (Richardson 1981). This fold comprises two β-sheets, each consisting from three to six antiparallel β-strands. The most important function of CBMs is to promote the association of the enzyme with the substrate (Boraston et al. 2004).

### 1.1.1.3 Irregular scaffolds

Many scaffolds derived from natural proteins belong to this group:

**Defensins:** are small, arginine rich, cationic proteins found in both vertebrates and invertebrates. They have also been reported in plants. They are active against bacteria, fungi and many enveloped and nonenveloped viruses. They consist of 18-45 amino acids including six (in vertebrates) to eight conserved cysteine residues. Typically they were used to build
a mutant peptide library characterized by α-helix and β-sheet structures (figure 1.17) as insect defensin A, with a secondary structure stabilized by two disulfide bonds, an α-helix, two β-sheets and loops in two regions able to tolerate substitutions and randomizations. This structure was of particular attraction due to its ability to be expressed on a phage, in fusion with gene III (Xu et al. 2002). The library was successfully screened against tumor necrosis factor A, TNF receptor 1 and TNF receptor 2 yielding improved affinity interactors (Xu et al. 2002; Zhao et al. 2004).

**Kunitz inhibitor domains**: are composed of α-helices and β-sheets (about 60 amino acids) stabilized by three disulfide bridges. They are very stable and display a loop which can be randomized without destabilizing the structure: this is an essential feature in order to generate a library. They are also exploited to build and identify new specific protease inhibitors (figure 1.18) (Williams and Baird 2003).

**PDZ domain**: is a common structural domain found in the signaling proteins of bacteria, yeast, plants, viruses and animals. PDZ is an acronym combining the first letters of three proteins — *post synaptic density protein* (PSD95), *Drosophila disc large tumor suppressor* (Dlg1), and *zonula occludens-1 protein* (zo-1) — which were first discovered to share the domain. The PDZ are globular domains consisting of three α-helices and five β-sheets (figure 1.19), containing approximately 80-100 amino acids (Ponting et al. 1997). This scaffold has been described in protein-protein interactions of cellular signaling pathways (Fuh et al. 2000; Laura et al. 2002); it has been engineered for the recognition of the C-terminus of several proteins. Artificial PDZ domains can be isolated and engineered by mutagenesis screen
in vivo and also used to target intracellular proteins to different subcellular compartments (Schneider et al. 1999). One example is the PDZ domain of the protein tyrosine phosphatase PTP-BL that mediate interactions by binding to specific amino acid sequences in target proteins. (Walma et al. 2002).

**Scorpion toxins:** are very stable small protein scaffolds (figure 1.20) due to the presence of disulfide bridges. This motif is the best example of stabilization of α-helix (cystein stabilized helix CSH) and β-sheet (cystein stabilized sheets CSB); and is also called “cysteine-stabilized αβ motif (CS αβ)” (Cornet et al. 1995). The construction of a random library based on a small scaffold derived from charybdotoxin with a randomized β-turn was recently described (Li et al. 2001).

**Plant Homeodomain finger protein (PHD):** was discovered over a decade ago by Schindler and colleagues, who noted a stretch of conserved sequence with regularly spaced cysteines in two plant homeodomain proteins. Homeodomain proteins have been shown to play a major role in the development of various organisms (Schindler et al. 1993). The plant homeodomain finger protein is a small domain with two flexible loops, 1 and 3 (figure 1.21), that tolerate well elongation and mutagenesis events and it has no disulfide bonds. PHD fingers bind specific nuclear protein partners, apparently using their two loops surface and some PHD finger can bind to nucleosomes (Bienz 2006).

**TEM-1 β-lactamase:** this enzyme scaffold (263 amino acids) has a secondary structure formed by several large α-helices and β-sheets (figure 1.22). This type of scaffold was used for the generation of libraries based on phage display in which random inserts are added near the catalytic site; in this
way the link between the scaffold and the target molecule influences the catalytic activity of the enzyme itself. Several experiments were performed using the technique of site-directed mutagenesis at the active site of this enzyme in specific positions. For example, *amber* mutations were introduced at codons 69 or 244 in the *bla* gene by site-directed mutagenesis. In the course of systematic replacements of residues in the active site region of TEM-1 β-lactamase by informational suppression, several mutants obtained at positions 69 and 244 led to clavulanic acid resistance phenotypes (Delaire et al. 1992).

### 1.1.2 Coiled coil domains

α-helices are widespread in nature and are also involved in the folding of an important structural motif: the coiled-coil domain (Lupas and Gruber 2005).

The coiled-coil is an oligomerization domain found in a wide variety of protein, including transcription factors, virus proteins and fibrous proteins (keratin, myosin, fibrinogen). Several surveys of genomic data bases suggest that up to 10% of eukaryotic proteins contain sequences predicted to be coiled-coils (Liu and Rost 2001), confirming that they are the most common interaction and dimerization domain in helical structures.

The coiled coil was theoretically predicted by Pauling and observed experimentally first by Crick in 1952 in the X-ray structure of an α-keratine (Pauling and Corey 1951; Crick 1952). The minimal coil structure is given by two right-handed amphipathic α-helices (figure 1.23) that usually adopt a single left-handed superhelical fold in order to form a coil.

In 1972, the amino acid sequence of a small protein called tropomyosin, with a structure of coiled-coil was described (Sodek et al. 1972). Tropomyosin
has a characteristic motif, known as heptad, formed by 7 amino acid repeat (a-b-c-d-e-f-g), where a and d positions are occupied by hydrophobic residues.

This repetition of type 3-4 or 4-3 (NXXNXXNXXN ...) for the hydrophobic residues is kept along the entire amino acid sequence of tropomyosin. This repetition was then detected in other proteins characterized by α-helix supercoiled structures, such as the "leucine zipper" sequence (figure 1.24) (O'Shea et al. 1993; Whitby and Phillips 2000).

In all coiled coils positions a e d are occupied by hydrophobic residues and are important for interhelical contacts (figure 1.25), which contribute to the largest part of the interaction hydrophobic surface.

![Fig. 1.24 - Structure of a leucine zipper motif.](image)

![Fig. 1.25 - (A) Helical wheel representation of an heterodimeric coiled-coil, viewed from the N-terminus: the fundamental residue for the interaction are highlighted. The hydrophobic residues in positions a, d, a' and d' interact, allowing the formation of the apolar core; charged or polar residues in positions e, g, e' and g' permit the formation of potential ionic or electrostatic interchain interaction. (B) Two different views of a coiled-coil molecular model. Adapted from Arndt et al. 2002.](image)
Typically, polar or charged amino acids occupy the \( e \) and \( g \) positions and help in driving stability and specificity, and increase the solubility of the protein. Furthermore, these \( e/g \) edge positions shield the hydrophobic core from aqueous surroundings. The typical hydrophobic repeat is highly conserved (Arndt et al. 2002; Steinmetz et al. 2007) and several studies have shown that synthetic models of coiled-coil formed by only 4 or 5 repeated heptads with leucine in both positions \( a \) and \( d \) (Zhu et al. 1993; Su et al. 1994), give rise to much more stable and ordered coiled-coil structures than those observed in tropomyosin (Lau et al. 1984; Acharya et al. 2006), because:

1) charged or polar residues inserted in these \( a \) and \( d \) positions have a destabilizing effect on the structure in respect to hydrophobic residues;

2) hydrophobic residues different from leucine, isoleucine, characterized by a bulky side chain, have destabilizing effects (Lau et al. 1984; Moitra et al. 1997);

Moreover, in the leucine zippers, \( \beta \)-branched residues are tolerate only in the \( a \) position, while in tropomyosin they are present in \( a \) and \( d \) positions. The so far observed variability can be explained if we consider the different orientation of these residues to the axis of the dimer formed by two antiparallel helices. The hydrophobic side chain of residue in position \( a \), points outside the hydrophobic interface between the two helices, while the side chain of residue \( b \) is directly projected to the inside interface. The apparent preference for \( \beta \)-branched residues in the \( a \) position probably contributes to stabilizing the hydrophobic core, allowing closer contact between adjacent residues (Zhu et al. 1993; Kohn et al. 1995; Chao et al. 1998).

The model proposed by Crick 40 years ago called "Knob into Hole" (Crick 1952) (figure 1.26) is of particular interest to explain the interactions established between the two hydrophobic surfaces.
The hydrophobic interface consists of two distinct alternated surfaces, which give rise to a tightly packed hydrophobic core. In this structure the Knob is formed by residues a that are inserted inside the Hole present between residues a-g-d of the corresponding heptad on the other helix. The residues d present on the opposite helix, will behave in a similar way to the residues a present in the first helix. In contrast with what is proposed for the leucine zippers, the residues in d position are not interdigitated between them, but arrange themselves side by side with residues in position a (Hu et al. 1990; Acharya et al. 2006).

This conformation leads to a very resistant packaging: the residues a and d are completely buried inside the hydrophobic region, thus justifying the destabilizing effect observed in the presence of small residues such as alanine, bulky residues such as phenylalanine, tyrosine, tryptophan and polar or charged residues (Litowski and Hodges 2002).

Conversely, the presence of polar or bulky residues in positions a and d may induce partial unfolding and the formation of pockets (Litowski and Hodges 2002).
1.1.2.1 Electrostatic interactions

Electrostatic interactions may contribute to the stability of the structure of a protein in a variable manner depending on the micro-environment in which they occur. In fact, electrostatic interactions that are generated on a surface widely exposed to the solvent are generally weak and minimally contribute to the structure stability, while electrostatic interactions created in a hydrophobic environment, may have a strong stabilizing effect (Zhou et al. 1992).

The $e$ and $g$ positions of the coiled-coil that flank the hydrophobic core, may contribute to the hydrophobicity of the core, wrapping on the hydrophobic interface and interacting with the hydrophobic amino acids through methylene groups of the side chains, further protecting the core from the solvent.

The presence of inter-helix ionic interactions has been demonstrated through analysis of the coiled-coil domains, by a preponderance of oppositely charged residues in these positions (figure 1.27).

The inter-helical electrostatic interactions in the coiled-coils may have a stabilizing effect if located in the vicinity of the hydrophobic interfaces ($i$ and $i + 5$ positions of the helices corresponding to residues $e$ and $g$), where the microenvironment surrounding residues is not fully accessible the solvent (Kohn et al. 1995; Acharya et al. 2006).
In addition to inter-helical ionic interactions present in the coiled-coils, intra-helix electrostatic interactions are also possible at positions i, i + 4 that stabilize α-helix. These interactions are capable of altering the stability of the helix itself and the stability of the coiled-coil.

1.1.2.2 Orientation and oligomerization

Coiled-coils are found in nature as dimers, trimers, tetramers, and larger multimeric complex with parallel or antiparallel helices. However, all proteins with coiled-coil conformation, show common sequence patterns (Wagschal et al. 1999) and certainly packing of hydrophobic side chains and electrostatic interactions act synergistically through a very fine control (Monera et al. 1993; Potekhin et al. 2001).

Several studies have been carried out to recognize the structural requirements leading to the preferred oligomerization. However, the preferred arrangement remains unpredictable at the current state of the art and predicting structures is complicated by an enormous diversity of possible topologies. Some softwares are available to predict and identify a coiled-coil structure from a sequence on a statistic base, like Coils, ParCoil, MultiCoil and MarCoil (Berger et al. 1995; Wolf et al. 1997; McDonnell et al. 2006).

With a protein sequence database and residue probabilities, they analyze the sequence and predict the structure with high sensitivity and specificity. In a recent work, some of these softwares has been compared to predict coiled-coil motifs, evaluating the detecting accuracy of everyone (Gruber et al. 2006).

1.1.3 E-K coiled-coil

In 1993 a rational methodology for the design of de novo coiled-coil was reported (O'Shea et al. 1993).

Designed de novo coiled-coil scaffolds have been described for the first time in 1998 for use in the field of biosensors (Tripet et al. 1996; Chao et al. 1998; Litowski and Hodges 2001).

By analyzing the amino acid composition of various existing coiled-coil domains, it was possible to create a map of amino acids frequency in specific positions of the α-helix and the following features have been identified:
the occurrence of helicogenic or neutral amino acids in the Chou-Fasman models;
the presence in A and D positions of non-polar residues, capable of organizing themselves in a very densely packed hydrophobic interface;
the choice of charged amino acids to be placed in E and G positions in order to promote the association of the helices (Arndt et al. 2000);
the insertion of charged amino acids in F position of each helix to facilitate the solubilization of the protein in an aqueous environment;
the charge balance on each α-helix, through the insertion in F position of amino acids with opposite charge to those in the E and G positions;
the definition of an appropriate chain length to allow the production of helices synthetically and so as not to affect the structure or stability of the coiled-coil (Litowski and Hodges 2002; Kwok and Hodges 2004);
the balance between the stability of the hydrophobic core that drives the formation of homodimer and the electrostatic repulsion that prevents the formation of homodimers.

Following these requirements, the research group of Chao and colleagues has synthesized ex-novo a scaffold structure called E/K (figure 1. 28) consisting of two distinct peptides, formed by a five heptad repeats. The sequence of the E peptide is E-V-S-A-L-E-K (g-a-b-c-d-e-f positions) while the K sequence is K-V-S-A-L-K-E. When the two peptides are associated in a dimeric form, they generate a protein of 70 amino acids with a molecular mass of 7682 (Zhu et al. 1993; Kohn et al. 1995).
The hydrophobic core of the E/K coiled-coil consists of valine and leucine in A and D positions; such amino acids are commonly found inside the core of natural coiled-coil proteins. The presence of valine and leucine instead of leucine and isoleucine in the core does not favor the formation of homodimers and multimers in solution but makes the structure less stable.

On the basis of conserved amino acid residues in natural proteins with a similar structure, lysine and glutamic acid were selected at the E and G positions.

The E/K heterodimer is highly stable and its dissociation constant is 64 pM (Ayriss et al. 2007). It is able to maintain its three-dimensional structure even after heating to 85 °C or treatment with urea; neither the combination of these two factors is able to destabilize the molecule. Only 3.9 M guanidinium hydrochloride is able to denature the structure (Chao et al. 1998) (figure 1.29a).

The affinity of the peptide E/K coiled-coil was maintained after incubation in human serum at 37 °C for over 50 hours and the construct is not degraded by proteolysis (figure1.29b).
Fig. 1.29 - a) Denaturation of E/K to the urea (○), guanidinium hydrochloride (GdnHCl) (●) and heat (▲). The graph shows that the denaturation occurs only with GdnHCl (Chao et al. 1998). b) Stability profile of the E/K coiled-coil in human serum at 37 °C; as can be seen from the graph there is a non substantial degradation of samples (Chao et al. 1998).

**E-K coiled-coil in biosensors**

The E/K coiled-coil is used as capture and dimerization domain. The versatility of the E/K heterodimeric coiled-coil allows to explore the possibility of using this protein as an affinity purification tag (Tripet et al. 1996). In this example, one of the two peptides (step 1) is first immobilized on the sensor surface. The second peptide is conjugated to an antibody and injected on the sensor surface (step 2). The formation of the heterodimeric coiled-coil generates a surface of antibodies over the biosensors (figure 1.30).

Fig. 1.30 - A) Use of heterodimeric domain E/K coiled-coil for biosensing use. K-coils are immobilized with thiol bonds onto a glass slide, pre-coated with a gold layer. E-coil-fusion proteins are injected over the slide surface, and, exploiting the E/K dimerization proprieties. B) The surface can be also functionalized with antibodies fused with the E-coil, for antigen-binding biosensing platforms (Chao et al. 1998).
E/K system can also be used for detection of recombinant peptides or proteins and for purification by selective dimerization. The first system can be used in western blot like assays: a protein can be produced by bacteria conjugated with one of the coil compartments (e.g. E) and detected using labeled K-coil. The K-coil may be labeled with biotin, peroxidase, alkaline phosphatase or other labeling systems. In the second system, purification is carried out flowing the cell-extract through a column containing the immobilized K-coil peptide. E-coil fusion proteins dimerize with K-coil in the column and are retained. After washing, the E-protein is eluted; then the tag is removed with chemical or enzymatic processes.

A nanoarray called Nano-Capture K-coil exploits this system. The K helices are currently immobilized and dimerize with the E helices in fusion with any molecule of interest (figure 1.30). In this way these arrays are fully configurable for the application of desired biosensors (Plexera Bioscience Innovations in Antibody Discovery Technology).
Chapter 2

1.2 Human Serum Albumin

Albumin is one of the most studied proteins. Despite this, even after more than 80 years of studies, many of its properties have not been fully clarified. Its complexity has attracted many researches, due to its important role in many physiological and pathological processes.

Albumin is the most abundant protein in the serum of vertebrates (~600 μM, 42 g/L in humans) (Peters 1996) and its main physiological functions are related to its ability to bind a large number of compounds, both endogenous and exogenous. For example, albumin allows to solubilize poorly soluble molecules in an aqueous environment (e.g. fatty acids), or allows to sequester potentially toxic molecules (e.g. bilirubin) (Kragh-Hansen 1990). It can also act as an antioxidant, because it can interact or inhibit the formation of various oxidizing agents (Halliwell 1988). Finally, thanks to its binding properties, Albumin acts as a carrier of metabolites (e.g. L-tryptophan, thyroxine, Ca\(^{2+}\) and Cu\(^{2+}\)) (Peters 1996), and drugs. Another important role of albumin is its contribution to colloid osmotic pressure (80%) which derives from its high concentration and its relatively low weight molecular.

1.2.1 Structure and properties

The primary structure of human serum albumin (HSA) was obtained by classical chemical sequencing methodologies (Behrens 1975; Meloun et al. 1975) before the obtaining DNA sequence (Minghetti et al. 1986). The human albumin gene spans 16,961 nucleotides from the putative “Cap” site to the first poly(A) addition site (figure 1.31).
Fig. 1.31 - Map of the human serum albumin gene. The map shows major restriction endonuclease sites and the location of exons (boxes) and introns (heavy lines). Their lengths in nucleotides are indicated by numbers above and below these elements. The first and the penultimate (14th) exons are partly untranslated as indicated by open boxes. The last (15th) exon is completely untranslated. The gene begins at the Cap site of the first exon and ends at the first polyadenylation site of exon 15 (Minghetti et al. 1986).

It is split into 15 exons by 14 intervening sequences which are symmetrically placed within the three domains of albumin. The 5' region is highly conserved up to position -250 and contains the putative TATA (-32) and CAT (-88) boxes. A consensus 5' splice sequence reads GTAG while the 3' splice sequence is pyrimidine rich and contains AGC at the splice junction. The albumin gene exhibits a high degree of DNA polymorphism and appears to have been recently invaded by Alu repetitive sequences (Minghetti et al. 1986).

These studies revealed that HSA is single polypeptide chain of 585 amino acids with a molecular mass of about 66500. The presence of 17 disulfide bonds give rise to the nine double loops (eight plus half loop). These loops can be further grouped into three homologous domains, each including two long loops separated by a short loop. The three homologous domains are numbered I, II and III from the N-terminal residue. Within each domain, the first two loops (loops 1-2, 4-5 and 7-8) respectively, are grouped into subdomains IA, IIA, and IIIA and loops 3, 6 and 9 are called, respectively, subdomains IB, IIB, and IIIIB (figure 1.32).
The three domains and their subdomains, although homologous, have different binding characteristics.

As to the amino acid composition, a typical feature of HSA is the presence of a single tryptophan residue and a high occurrence of cysteine residues. In addition, in respect to the large majority of proteins HSA has a low content of methionine, glycine and isoleucine, while leucine residues and ionic amino acids such as glutamic acid and lysine are numerous (Peters 1996). The large number of ionizable amino acids leads to an high total charge to HSA, 185 ions per molecule at pH 7, which explains its high solubility. The acidic residues exceed basic ones causing a net charge at pH 7 of -15. This charge is not uniformly distributed in the molecule: it is highest in domain I (-9), decreases in domain II (-8) and almost vanishes in domain III (+2). The first half of the sequence has a net charge of -14 at pH 7.4, while the second of -1. The isoionic point, ie the pH of a solution of completely deionized HSA, is about 5.2 (Hughes and Sinex 1954); in contrast, the isoelectric point, ie the pH where the net charge of the protein is zero, is about 4.7 in 0.15 M NaCl (Longsworth and Jacobsen 1949): probably chloride ions and fatty acids bound to the protein lead to a reduction of this value from the isoionic point (Peters 1996).
HSA is composed only by amino acids without prosthetic groups; it is also one of the few non-glycosylated serum proteins lacking the Asn-X-Ser/Thr sequences. Three of the HSA known variants have mutations which create Asn-X-Ser/Thr sequences but the glycosylation of these sites does not seem to have functional effects. In fact the circulating HSA accumulates glucose by non-enzymatic glycosylation reactions and other molecules, such as Cys or glutathione, through the formation of a covalent bond between the SH-groups of these molecules and the \(^{34}\)Cys residual free by disulfide bridges. The absence of carbohydrate is used as a criterion of purity of preparations of HSA: pure albumin must contain less than 0.05% of carbohydrates (Hughes and Sinex 1954). In addition, the HSA preparations usually contain small amounts of very similar molecules, such as hemin or bilirubin, which causes the typical yellow color of the concentrated solutions of the protein. The 35 Cys residues form 17 disulfide bridges, leaving free only residual \(^{34}\)Cys (figure 1.33). Of such residues, almost half are located in contiguous positions along the chain. Although it has been shown that two adjacent Cys can form S-S bond (Zhang and Snyder 1989), it is generally considered that each Cys form a disulfide bond with Cys residue closer, before and after the Cys-Cys pair. The presence of many disulfide bonds explains both its flexibility, and its particular resistance to external conditions.

![Diagram of HSA structure with disulfide bridges and 34Cys free.](image)

Fig. 1.33 - Human serum albumin structure (PDB 1e78) with disulfide bridges (in black) and 34Cys free. Adapted from Yamasaki et al. 2000.
As to the three-dimensional structure, hydrodynamic studies (Squire et al. 1968) and small-angle neutron scattering (Bendedouch and Chen 1983) converge in attributing to HSA in solution a “cigar” shape (ellipsoid of rotation) with the major axis 140 Å long and the minor one 40 Å long. In the crystal phase, albumin has a tetrahedral shape similar to heart (Carter and Ho 1994), formed by equilateral triangles of 80 Å per side, with an average thickness of 30 Å. This shape difference accounts for the great flexibility of the protein depending on the environment. The X-ray analysis also allowed to determine the secondary structure of HSA, which, in solid form, has 67% α-helix, 10% β-turn and the remaining 23% extended chain (Carter and Ho 1994).

1.2.2 Binding sites

HSA interacts reversibly with a broad spectrum of therapeutic agents through its binding sites. Usually, drugs bind to one or very few high-affinity sites with typical association constants in the range of $10^4$ - $10^6$ M$^{-1}$ (Sudlow et al. 1975). In addition to the primary site(s), several sites of lower or of very low affinity often are found (Kragh-Hansen et al. 2002). However interaction studies are primarily focused on HSA high-affinity binding sites, because the plasma concentration of drugs in normal treatments is generally much lower than that of HSA (~ 600 μM), therefore it can be assumed that only high-affinity sites are occupied.

Binding of chiral molecules to HSA is often enantioselective. In recent years this phenomenon has assumed increasing importance as more than a third of marketed molecules is chiral. It is therefore necessary to investigate whether the enantioselectivity of the binding results in substantial differences in the pharmacokinetics of the drug. In general we must consider that different pharmacokinetic behavior between enantiomers are generally minor if compared with pharmacodynamics ones.

An important study by Sudlow and colleagues (Sudlow et al. 1975), based on competition with fluorescent probes, demonstrated the existence of two major binding sites, named sites I and II, which binds most drugs with high affinity (figure 34). Another crystallographic study (Curry et al. 1998) extended this model identifying seven binding regions (named FA1-7) for long-chain
Introduction

saturated fatty acids; three of the seven sites correspond to sites I and II by Sudlow (FA7 and the complex FA3 -FA4, respectively) (figure 1.34). Subsequent crystallographic studies have shown that these seven sites binds also intermediate-chain fatty acids, long-chain mono and polyunsaturated fatty acids (Curry et al. 1999; Bhattacharya et al. 2000; Petitpas et al. 2001). Finally, further studies have shown that hemin (Wardell et al. 2002; Zunszain et al. 2003), thyroxine (Petitpas et al. 2001) and several drugs (Bhattacharya et al. 2000; Petitpas et al. 2001; Petersen et al. 2002; Ghuman et al. 2005) bind to one or more of these seven binding sites. NMR studies confirmed the localization of the long-chain fatty acid (LCFA) binding sites and also determined the relative affinity. The results obtained indicate that the order of affinity is FA5> FA4> FA2>> FA1, FA3, FA6, FA7 (Simard et al. 2006)

Fig. 1.34 - HSA structure. The six subdomains of HSA are colored as follows. Blue, subdomain IA; cyan, subdomain IB; dark green, subdomain IIA; ight green, subdomain IIB; orange, subdomain IIIA; red, subdomain IIIB. The heme (red) fits the primary cleft in subdomain IB, corresponding to FA1. Sudlow's site I (in subdomain IIA, corresponding to FA7) is occupied by warfarin (purple). Sudlow's site II (in subdomain IIIA, corresponding to FA3-FA4) and FA6 (in subdomain IIB) are occupied by ibuprofen (magenta). Sites FA2 (at the I-IIA interface) and FA5 (in subdomain IIIB) are occupied by myristate (yellow). Atomic coordinates were taken from Protein Data Bank entries 1O9X (23), 1H9Z (20), and 2BXG (12). (Ascenzi et al. 2009)
1.2.2.1 Site I

Dicarboxylic acids and/or bulky heterocyclic compounds with a negative charge located toward the center of the molecule bind with high affinity to site I of HSA (Sudlow et al. 1975). However these features are not sufficient to predict whether a molecule can bind to this site or not, given that even molecules with very different characteristics seem to bind to site I with high affinity (Kragh-Hansen et al. 2002).

The site is also known as warfarin-azapropazone binding site: it is very large and adaptable as it can bind bulky molecules such as bilirubin. It was also shown that more molecules can bind to this site independently (Kragh-Hansen 1990), such as phenylbutazone-indomethacin and azapropazone-indomethacin (figure 1.35) (Ghuman et al. 2005). This phenomena suggests the presence of multiple subsites overlapped or at least very close to each other; secondly, the binding of a molecule able to change the three-dimensional arrangement of the binding site, may create a new structurally different one (Kragh-Hansen et al. 2002).

In a first model, the first binding site was considered as composed at least of two overlapping subsites (warfarin and azapropazone) (Fehske et al. 1982).

Fig. 1.35 - Summary of the ligand binding capacity of HSA as defined by crystallographic studies. Ligands are depicted in space-filling representation; oxygen atoms are coloured red; all other atoms in fatty acids (myristic acid), other endogenous ligands hemin, thyroxin) and drugs are coloured dark-grey, light grey and orange, respectively. From Ghuman et al. 2005.
Recent studies suggest the existence of at least three regions, called la, lb and lc (Yamasaki et al. 1996). According to this model, the la region is located between the regions lb and lc and plays a connection role between these two subsites. On the contrary, the lb and lc regions, do not seem be related to each other, since, by equilibrium dialysis experiments, binding to these regions seems to occur independently. Circular dichroism experiments have shown that specific markers binding of lc site, after lb allosteric change, vary the spatial orientation of specific markers of lb site, without influencing the affinity. This albumin conformational rearrangement has strong similarities to occurring upon the NB transition, from pH 6.0 to pH 9.0. Dansyl-L-asparagine (DNSA) and n-alkyl-p-aminobenzoate (p-ABEs) are used as markers of specific subsites lb and lc, respectively (Yamasaki et al. 2004).

In the crystal structure, the site I of HSA results as a large hydrophobic pocket located in IIA domain, delimited by six helices of IIA subdomain and by a loop of lb subdomain (residues 148-154). The Trp residue, the only one in the HSA primary structure, is located within this pocket. The inner wall of the pocket is formed primarily by hydrophobic residues, except for two clusters of basic amino acids located at the bottom and at the entry of the pocket.

The site can be divided into a large central area with three different compartments: two hydrophobic regions (one on the right and the other on the left) separated by Ile and a third compartment which protrudes from the central area towards the bottom. In the absence of fatty acids, warfarin, phenylbutazone, CMPF (3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid) and oxifenbutazone are placed in the central portion of the pocket (Ghuman et al. 2005), with their planar group between the hydrophobic side chains of Leu and Ala (figure1.36). The position of the rest of the molecule is much variable. These drugs occupy the right rear, while only phenylbutazone and CMPF project their hydrophobic groups into the back left. The front is occupied by the phenyl groups of oxifenbutazone and warfarin.

Tyr plays a central role for the drug binding to the site, in the absence of fatty acids. In fact, in addition to hydrophobic interactions, all drugs are positioned to form a hydrogen bond with Tyr. However, in the presence of myristate, Tyr is positioned so as to play a marginal role for the binding of drugs. The two enantiomers of warfarin adopt very similar conformations when
bound to the protein and make many of the same specific contacts with amino acid side chains at the binding site, thus accounting for the relative lack of stereospecificity of the HSA-warfarin interaction. The conformation of the warfarin binding pocket is significantly altered upon binding of fatty acids, and this can explain the observed enhancement of warfarin binding to HSA at low levels of fatty acid (Petitpas et al. 2001).

Another peculiarity is represented on the orientation of phenylbutazone and oxifenbutazone within the site. The two molecules, despite having only a small structural difference (the oxifenbutazone has a hydroxyl group on one of the aromatic rings) in the absence of fatty acids, bind in positions rotated 180° from each other. However, the presence of myristate, leads oxifenbutazone to bind with the same orientation of phenylbutazone.

Fig. 1.36 - Drug binding to site 1 in HSA-myristate. The detailed binding conformations are shown for (a) azapropazone, (b) indomethacin and (c) phenylbutazone. In each case the drug is shown in a stick representation with a semi-transparent van der Waals surface (magenta). Bound fatty acids are depicted with a yellow semi-transparent van der Waals surface. The methylene tail of a molecule of myristate was observed co-bound with phenylbutazone in site FA9.(d) Top view of the superposition of drugs bound to site 1 in HSA-myristate. In this case the clustering of oxygen atoms is less pronounced. (e) Side view of superposition of drugs shown in (d) along with a semi-transparent surface (blue) depicting the extent of the pocket. (f) Superposition of the structure of HSA-myristate–indomethacin–phenylbutazone (secondary structure coloured by domain with drugs and selected side-chains shown as sticks with grey carbon atoms) with HSA-myristate–indomethacin (drug and side-chains shown as thin sticks with cream carbon atoms) and HSA-myristate–phenylbutazone (blue carbon atoms in side-chains). From Ghuman et al. 2005.
1.2.2.2 Site II

Site II, or indole-benzodiazepine site, binds with high affinity to aromatic carboxylic acids with a negative charge placed at one end of the molecule, away from the hydrophobic core (eg. Nonsteroidal anti-inflammatory drugs, FANS). This site is smaller than the site I, in fact, bulky molecules (such as bilirubin, hemin, hematin, and other porphyrins) are not able to bind to this site. Moreover, this site appears highly stereoselective. A typical example of enantioselectivity of the site II is represented by L-Tryptophan that have an affinity 100 times higher than D-enantiomer (Mc and Oncley 1958).

In addition, replacement of small groups on the ligands of this site may strongly influence the activity. Diazepam, but not its fluorinated analogue flunitrazepam, binds site II with high affinity. α-methylation of tryptophan prevents its binding to site II (Chuang and Otagiri 2001). Thus, although many molecules are capable of binding site II, this site appears to be more restrictive than site I. Nevertheless the site is able to bind different fatty acids with a certain selectivity (FA3 and FA4 sites) (Bhattacharya et al. 2000).

Crystallographic analysis revealed that site II is located within domain IIIA and showed its overall structure is similar to that of site I (He and Carter 1992; Sugio et al. 1999; Ghuman et al. 2005).

410Arg and 411Tyr residues are generally considered important for the binding of molecules, such as ketoprofen, although testing on recombinant HSA (rHSAs) have shown that 410Arg is not essential for the binding of diazepam (Watanabe et al. 2000).

The most important region of binding site II is similar to the central area of the site I. Another subsite, corresponding to the right rear of the site I, is present although it is accessible only after ligand-induced rearrangement of HSA. However, binding pockets similar to the second hydrophobic subsite in site I are lacking. Unlike site I, site II has a unique cluster of polar amino acids (410Arg, 411Tyr, 414Lys, 489Ser) located on one side at the site entry. So, despite several similarities, sites I and II are clearly different by shape, size and polarity, and this accounts for the different binding characteristics (Ghuman et al. 2005).

Site II ligands (diazepam, ibuprofen, diflunisal and indoxil sulfate) are found in the X-ray structures in the central area of the site, with at least one oxygen pointing towards polar cluster.
1.2.2.3 Other binding sites

Despite the classification of Sudlow and co-workers is still very useful, many experimental evidences show that sites I and II are not the only high-affinity binding sites of HSA. For example, probenecid and amitriptyline do not bind any of the two sites (Sudlow et al. 1975), and also the high affinity binding site of digitoxin was not yet located (Kragh-Hansen et al. 2002). Microcalorimetry and NMR studies show that tolbutamide binds to all three sites HSA with a comparable affinity (Jakoby et al. 1995).

Moreover, the binding sites of many basic drugs which bind HSA with high affinity are still unknown (Peters 1996; Kragh-Hansen et al. 2002).

The $^{34}$Cys residue is an important point of interaction for the irreversible binding of some molecules. This residue is located in the IA domain (figure 33) and is the only thiol free from disulfide bonds. Generally, in about half of the sulfhydryl groups of HSA samples, $^{34}$Cys is not free, but is actually present as mixed disulfides with cysteine or glutathione, or is oxidized to sulfite, sulfate, or sulfonate. The remaining sulfhydryl groups are free and accessible, and represent the major source of thiol groups in the blood. This residue is then available to the attack of oxidizing substances, so that the amount of reduced HSA in the blood (mercaptoalbumin) has been proposed as a biomarker of oxidative stress (Peters 1996). This residue also binds irreversibly to various drugs, such as bucillamine derivatives, aurothiomalate, auranofin, D-penicillamine, captopril, ethacrynate, cisplatin but also cysteine and N-acetyl-L-cysteine (Kragh-Hansen et al. 2002)

1.2.3 HSA binding interactions

Binding of a molecule to HSA often changes its ability to interact with other molecules. This is due to the high conformational flexibility of albumin and to the presence of few high-affinity binding sites (Chuang and Otagiri 2006). Alterations in the HSA binding of drugs may lead to a change in the pharmacokinetic properties of these drugs (Kragh-Hansen et al. 2002). This risk is especially relevant if two or more drugs compete for the same HSA-high-affinity binding site. Sometimes free drug concentration increases by less than expected, because a most of the displaced drug rebinds at another site.
(Rahman et al. 1993). This phenomenon was observed with diclofenac in the presence of ibuprofen (Yamasaki et al. 2000).

Simultaneous binding of endogenous compounds such as fatty acids, uremic metabolites, bile salts and bilirubin can modulate the drug binding properties. Long-chain fatty acids (LCFAs) compete with drugs for the binding sites FA1-7, and may also induce allosteric competitions. Fatty acids compete directly with the drug-binding site II, as it corresponds to the sites with the highest affinity for LCFA (FA3-FA4). In contrast, the effect of the presence of LCFAs on the site depends on their concentration:

- LCFAs low excess increases the affinity for various site I-specific ligands such as warfarin, furosemide and bilirubin (positive allosteric competition);
- LCFAs high excess, however, decreases the affinity (negative allosteric competition and direct competition to the site FA7) (Kragh-Hansen et al. 2002).

It was demonstrated that chronic renal failure is associated with the accumulation of uremic toxins in serum, including four organic acids that are able to bind to albumin and compete with many drugs, such as diazepam, warfarin, salicylate, furosemide, L-tryptophan, L-thyroxine and phenytoin. In recent years it has been possible to single out some individual effects of the toxins. In particular, 3-carboxy-4-methil-5-propyl-2-furanpropanoic acid (CMPF), which binds specifically to the site I, competes with drugs that bind to the same site as warfarin, furosemide and bilirubin (Tsutsumi et al. 1999). Similarly indole-3-acetate toxins, indoxyl-sulphate and hippurate, which bind with high affinity to site II, displace drugs that bind to this site. A wide set of drug competes with bilirubin, a toxic and insoluble heme catabolism product, which is sequestered and solubilized by HSA. Several examples are valproate, ceftriaxone, some sulphonamides like sulfisoxazole, oxyphenbutazone, phenylbutazone, glibenclamide, tolbutamide, warfarin, furosemide, and salicylate. Instead, site II ligands generally have little or no effect on the binding of bilirubin. The presence of high levels of free bilirubin can be very dangerous, especially in infants. Therefore, all drugs that should be administered to infants, should be tested for their effect on the binding of HSA-bilirubin.
1.2.4 Enzymatic properties

Although the main role of human serum albumin is transport, the protein shows a variety of enzymatic properties (Kragh-Hansen et al. 2002) of both biological and bioorganic significance.

- Albumin acts as thioesterase because it has a free sulfhydryl group at residual $^{34}$Cys. From a clinical perspective, this property is very important because it allows the degradation of disulfiram by HSA (Agarwal et al. 1986);
- HSA enolase activity was observed on dihydrotestosterone enolization by the amino terminal of the molecule (Drmanovic et al. 1999);
- HSA can catalyze the dehydration of prostaglandin D$_2$ and dehydration, with isomerization of 15-keto-prostaglandin E$_2$ to 15-keto-prostaglandin A$_2$ (PGA$_2$) and 15-keto-prostaglandin B$_2$ (PGB$_2$) by modulating of specific amino acid residues (Fitzpatrick et al. 1984; Yang et al. 2002);
- IIA domain amino acids can degrade sulbenicillin, with the $R$-isomer being degraded much faster than the $S$-isomer (Kragh-Hansen et al. 2002). Thus enzymatic activity of HSA can, like the ligand binding properties, be stereospecific;
- The IIA domain also have esterase properties; a recent study on the crystal by X-ray diffraction (Yang et al. 2007) showed that it catalyzes the transfer of acetyl group from aspirin to $^{199}$Lys (figure 1.37). This lysine residue is also important for rapid trinitrophenylation of HSA and for penicilloylation of the protein;
- The IIIA domain has esterase activity towards various substrates such as p-nitrophenyl acetate and various other esters p-nitrophenyl (Watanabe et al. 2000).
Introduction

The glucuronidase activity is more difficult to locate, because it seems that the different substrates are hydrolyzed by different domains: some, such as diflunisal glucuronide are hydrolyzed at the I site (Williams and Dickinson 1994), others, such as carprofen glucuronide at II site (Georges et al. 2000), and others, such as ketoprofen glucuronide in some other region of the molecule (Dubois-Presle et al. 1995).

The enzymatic activity of HSA can be exploited for practical use. For example a method for the quantification of the protein has been developed, executable with both fluorimetry and spectrophotometry, using coumarin as a substrate (Gurakar and Wolfbeis 1988). In addition, the HSA esterase activity can be used for pharmaceutical purposes, as it allows the activation of prodrugs into drugs. For example medoxomil olmesartan in plasma is converted to olmesartan in few seconds. So the large amount of circulating albumin allows for rapid hydrolysis, although the enzymatic activity of HSA, in itself, is not very high.

Besides such catalytic activities of biological interest, HSA has been widely considered as model system and as a potentially useful tool in bioorganic chemistry and a catalysis.
A broad chemical reactivity is associated with the small molecule binding site located in albumin’s IIA subdomain. A lysine residue is present in this site in both HSA (Lys-199) and BSA (Lys-222). In addition to being a site for covalent interactions with many drugs such as, for example, aspirin and benzylpenicillin, this basic residue, surrounded by a hydrophobic environment, is responsible for albumin’s ability to behave as an enzyme-like catalyst in reactions such as β-eliminations, the decomposition of Meisenheimer adducts and the eliminative ring fission of 5-nitrobenzisoxazole to 2-hydroxy-5-nitrobenzonitrile (Kemp elimination). In the latter reaction, in particular, the catalytic activity of BSA and HSA is similar to that of specifically designed catalytic antibodies.

Due to its low cost, large scale availability, and exceptional ability to bind a wide range of substrates in an asymmetric environment, BSA has found many applications as a stationary phase in chiral chromatography and as a chiral auxiliary and catalyst in organic reactions. Asymmetric oxidations of sulfides, disulfides, thioacetals, and tertiary amines, reduction of aromatic ketones, Diels–Alder cycloadditions, epoxidation and dihydroxylation of alkenes, and the enantioselective hydrolysis of esters have been reported to proceed with varying levels of stereoselectivity in the presence of BSA.

Our research group has recently reported the ability of HSA to directed the borohydride reduction of β-diketones to the formation of anti-1,3-diols, and to catalyze the aldol reaction of aromatic aldehydes with acetone (Benedetti et al. 2011; Berti et al. 2011).

1.2.5 Flexibility

Albumin is characterized by a high conformational flexibility (Weber 1975). The loops that characterize albumin structure allow the protein to expand, contract and flex very rapidly, after the binding of molecules or when it is alone (Peters 1996). The exchange speed of heteroatom-bound protons with water is often used to measure the proteins flexibility using tritium or deuterium as radioactive label (through nuclear magnetic resonance NMR) (Hvidt and Wallevik 1972). At physiological pH, about 750 of 1100 protons potentially exchangeable of bovine albumin (BSA), exchange at a rate that cannot be measured by NMR, even at 0
°C. Two other classes of 280 total protons, exchange at a rate of $10^{-3}$ and $10^{-5}$ sec$^{-1}$. The remaining 70 do not exchange even after 24 hours (Hvidt and Wallevik 1972). This high rate of proton exchange is a characteristic of albumin, even unique among non-enzymatic proteins (Willumsen 1971), and is probably related to its propensity to bind various ligands. Although we assume that albumin has a single form in solution, it is more realistic to think about a molecule constantly changing its shape (Hvidt and Wallevik 1972; Kragh-Hansen 1990).

1.2.6 Conformational equilibrium

Albumin undergoes several conformational transitions depending on pH; four isomers of the normal form (N shape) were identified: extended (E) at pH <3 (Leggio et al. 2008) (figure 1.38), fast (F) at pH 4, basic (B) at pH 8 and aged (A) at pH around 10. The equilibrium between these conformations is reversible.

When albumin is in F shape, it becomes less compact and less soluble and its alpha helix content is reduced. This conformation predominates at pH 4 and its name is due to the fact that it migrates faster than the N form on electrophoresis gel (Rachinsky and Foster 1957). Experiments on HSA fragments suggest that during N-F transition, the domains IIB + III are separated from the domains I + IIA (Carter and Ho 1994). Below pH 3, there is a further molecule relaxation that causes the complete loss of albumin tertiary structure, which becomes a fully extended protein.

Shape B becomes predominant at pH 8 and is also present in physiological pH conditions. For this reason, shape B is believed to have an important physiological role and to be involved in drugs transport. A loss of stiffness is observed especially in N terminal region of the B shape. Some drugs that bind to site I interact with the shape B with higher affinity than shape N, while site II ligands affinity is affected to a less extent (Peters 1996).

Shape A was discovered leaving a HSA solution at pH 9 for 3-4 days (Sogami et al. 1969). This form predominates at pH ≥ 10, and shows a nearly reversible equilibrium with forms B and N. The helix content is not very different from that of form B (Era et al. 1991). Chromophores and fluorophores of HSA shape A appear more exposed to the solvent and the protein is more
susceptible to proteolytic attack. This means that protein assumes a more open and relaxed shape in respect to N shape.

Fig. 1.38 - HSA adopts a heart-shaped structure (left) in solution rather than the cigar shaped ellipsoid (right) (Leggio et al. 2008).

1.2.7 HSA related disease

Hypoalbuminemia is usually caused by a protein reduced synthesis (in certain liver diseases), its greater loss from the body (in cases of extensive burns, and nephrotic syndrome) or states of malnutrition (Kragh-Hansen 1990).

On the contrary hyperalbuminemia is almost always due to a state of dehydration rather than to an increase in protein synthesis.

The analbuminemia is a very rare hereditary disease transmitted by an autosomal recessive trait (Kragh-Hansen 1990; Peters 1996; Caridi et al. 2008), characterized by a very low albumin concentration in plasma. Curiously, in almost half of the analbuminemia cases, patients are completely asymptomatic, while others have only non-severe clinical symptoms (usually edema and hyperlipidemia). Compensatory increase in the levels of other serum proteins is not enough to explain this phenomenon and there are other still unknown mechanisms (Kragh-Hansen 1990; Peters 1996).

Finally, another rare inherited disorder involving albumin is bisalbuminemia, an abnormality characterized by the presence of a certain amount of "normal" serum albumin, known as albumin A (not to be confused with the isomeric form A, see paragraph 1.2.6 ) and a certain amount of its variant. This anomaly has a prevalence of between 0.03 to 0.1% of the general population (that is too low to use the term "polymorphism", which implies a prevalence over 1%) (Peters 1996). Since only one third of the albumin amino acids are ionized, about half of single residue substitutions would not involve changes in the protein net charge, so these variations are not detectable by electrophoresis (Peters 1996), while the use of mass spectrometry can solve
this problem. Among the mutants identified, only one showed functional alteration: variant with increased affinity for thyroxine observed in family disalbuminemic hyperthyroxinemia (FDH). At least one hundred variants have been identified (Peters 1996).

### 1.2.8 Spectroscopic features

Spectroscopic techniques allowed the identification of protein main functional groups and conformational structure by using radiance properties: absorbance, fluorescence, polarization, optical rotation and Raman radiation. The peptide bonds are aromatic amino acids chromophores which are the groups responsible for such properties (Trp, Tyr and Phe). The main techniques used to study the HSA binding features are circular dichroism (CD), fluorescence and UV absorption (Pistolozzi and Bertucci 2008).

### 1.2.9 UV spectral properties

In the far ultraviolet albumin absorption is mainly due to the $n\rightarrow\pi^*$ transitions of the peptide bonds, with a peak near 187 nm (Rosenheck and Doty 1961). The extinction coefficient is very high and much influenced by the presence of hydroxyl or carboxyl groups. The minimum absorbance value in the range of 187-280 nm is about 253 nm.

In the near ultraviolet (240-400 nm) albumins absorption is similar to that of most proteins lacking of prosthetic groups, with a peak at 278.5-279 nm due to $\pi\rightarrow\pi^*$ transitions of aromatic amino acids. However, the absorption is very low due to the presence of a single Trp residue. BSA absorptivity at 280 nm is about 25% higher in respect to HSA due to the presence of two Trp residues, instead of one. Between pH 5.0 and pH 8.0 there is a slight absorption difference, while no differences are observed after changing the ionic strength between 0 and 0.3 M.

The different albumins absorptivity, calculated starting from the amino acid composition and disulfide bridges contribution ($\varepsilon_{280}=134$), agrees with those experimentally determined. For HSA, BSA and RSA (rat serum albumin)
the calculated absorptivities are 0.52, 0.65 and 0.55 (Mach et al. 1992) while experimental ones are 0.531, 0.661, and 0.59 (Peters 1996).

The absorbance measurement at 287 nm, where Tyr absorbance is the maximum, allows to predict the relative position within the molecule. In fact about two thirds of the 18 Tyr residues of HSA are influenced by solvent effects (Steinhardt and Stocker 1973).

Studies on spectral variations in the near ultraviolet show that:
- one third of the Tyr residues are easily accessible and thus are located on the molecular surface of the protein;
- another third becomes accessible under acidic pH conditions, in which the HSA domains separate;
- while the last one third becomes accessible only after the Cys residues reduction, resulting in irreversible opening of the three-dimensional protein structure (Peters 1996).

1.2.9.1 Circular dichroism

Albumin shows high-protein α-helix content, with two negative maxima around 208 and 222 nm, and a positive maximum around at 190 nm. The CD spectrum analysis has allowed to estimate the secondary structure assumed by the protein in solution. An high-energy circular dichroism spectrum of the HAS was analyzed by the use of freeware CDPro (Sreerama and Woody 2000). Data obtained with this technique were in agreement with those determined in the solid state and showed the following secondary structure composition: α-helix 63%, β-sheet 3.8%, β- turn 11.8% and extended chain 21.7%.

The albumin-molecules binding properties can be studied using the technique of CD, according to the induced circular dichroism (ICD) spectral variations by changing different conditions. The binding of a molecule to the protein often stabilizes a particular conformation that leads to chirality, even for non-chiral molecules, for which the analyte-protein binding, generate a different CD signal in respect to amount of individual components spectra (Bertucci et al.). A typical example is diazepam, an achiral drug that exists in two symmetric conformations (M and P) that are in equilibrium in solution. HSA-diazepam binding gives rise to a ICD (induced circular dichroism) signal (Alebic-Kolbah et
al. 1979; Konowal et al. 1979) that was used to study the site location with the highest affinity and to determine the binding constant (Bertucci et al. 1990).

1.2.9.2 Fluorescence

Fluorescence is a process of light emission due to the decay of an excited electronic state generated by absorption of light.

Tryptophan residues allow to the HSA to emit fluorescence, but also tyrosine residues can emit fluorescence. 214 Trp residue fluorescence is one of the properties that can be exploited in the characterization of the drugs-HSA binding (Bocedi et al. 2004; Ascenzi et al. 2005; Bocedi et al. 2005). In fact, this residue is located within the site I (see paragraph 1.2.2.1) and molecules that bind to this site, or that alter its structure in an allosteric manner, modify the microenvironment around amino acid, thus varying the fluorescence signal.

By using an excitation wavelength between 295 and 305 nm, the Tyr residues are not excited, so that the emission around 345 nm is exclusively due to the Trp residue. The contribution of Tyr residues can be estimated using an excitation wavelength less than 295 nm, which is able to excite both Tyr and Trp residues, and then subtracting the contribution of Trp measured at λ > 295 nm (Peters 1996).
Chapter 3

1.3 Biosensors transducers

1.3.1 Main features

Since the first biosensor, consisting of a molecular recognition element and a transducer (Updike and Hicks 1967), was developed by Updike and Hicks and their associated techniques have been studied and developed (Nakamura and Karube 2003) to provide accurate detection of target analytes.

A biosensor (see figure 1.39) in the traditional sense is defined as: bioanalytical device incorporating a biological material or a biomimic (e.g., tissue, microorganisms, organelles, cell receptors, enzymes, antibodies, nucleic acids, etc.), intimately associated or integrated within a physicochemical transducer or transducing microsystem, which may be optical, electrochemical, thermometric, calorimetric, acoustic, piezoelectric or magnetic. The usual aim of a biosensor is to produce either discrete or continuous digital electronic signals, which are proportional to a single analyte or a related group of analytes (Soper et al. 2006). Depending on the measured electrical signal, biosensors can be classified into three principle classes: conductometric, potentiometric and amperometric. Amperometric biosensors are cheaper, faster, and have a higher sensitivity than the potentiometric and conductometric (Dzyadevych et al. 2002).

Fig. 1.39 - Schematic representation of a single element biosensor containing the biorecognition element, transducer and the physical output whose magnitude is related to the concentration of the analyte of interest. Adapted from Soper et al. 2006.
1.3.2 Utility of biosensors

Nanotechnology plays an important role in the development of biosensors, improving the intrinsic features of biodevices: the use of new material, their sub-micron size and their important new physical proprieties at nano-level, has revolutionized the field of chemical and biological analysis, enabling rapid and accurate analysis (Vo-Dinh and Cullum 2000; Vo-Dinh 2002; Jianrong et al. 2004; Jain 2005; Choi et al. 2007).

Biosensor technologies allow high sensitivity and specificity (Ligler et al. 2007) are applied in a wide range of disciplines and for various purposes. The most popular applications of biosensors so far are:

- therapeutic drug monitoring
- metabolomic
- clinical diagnostic
- environmental control (Ramanathan and Simonian 2007)
- monitoring of industrial processes
- food quality control (Babu et al. 2007).

Electrochemical biosensors are the most suitable to be miniaturized; can operate even in turbid systems (unlike optical ones) and have short response times (unlike the bioluminescent ones). The electrochemical biosensors have a lower limit of detection and lower costs in respect to other biosensors types. For this reason, they are the most used in monitoring of industrial bioprocesses (amino acids, yeasts, lactic acid, ethanol, etc.), environment (pesticides, fertilizers, estrogenc, CO, CO₂, etc.), and forensic analysis (cocaine, anthrax, nerve agents, etc.) clinical diagnostics (glucose, alcohol, DNA, hormones, etc.).

Point of care (POC) systems are viewed as integrated systems that can process clinical samples for a number of different types of biomarkers in a variety of settings, such as clinical laboratories, doctors offices and eventually, at home (Soper et al. 2006).

In particular, biosensors may be used to detect and quantify the presence of biomarkers for cancer, to aid clinicians in determining the type of cancer based on a molecular signature or a panel of signatures, and also in
determining the stage of the disease and in selecting the most appropriate therapy.

**1.3.3 Molecular recognition elements**

The ability to recognize the “target” or biomarker is viewed as the critical step in any diagnostic assay.

While antibodies can be used for recognizing such biomarkers, synthetic based recognition elements such as aptamers, peptides, surface-imprinted polymers or molecularly imprinted polymers are being investigated as replacements for antibodies (Soper et al. 2006):

- The most common approach for generation of cell-specific or molecule specific recognition element is represented by the selection of a single, well characterized biomarker and by the generation of antibodies or libraries of antibodies fragment. Most efforts in this field have focused on exploiting monoclonal antibodies as specific ligands (Pietersz and McKenzie 1992; Farah et al. 1998; Glennie and Johnson 2000; Green et al. 2000).

- High-throughput methods are required to screen libraries of potential ligands for cell binding, even when the nature of the cell surface is undetermined. Using phage display libraries, protocols to identify cell-specific binding peptides has been developed (Barry et al. 1996; Brown 2000; Oyama et al. 2003; McGuire et al. 2004).

- Several synthetic hosts have been studied for a variety of targets, such as aptamers and peptides. In contrast with antibodies, these synthetic hosts are stable and can be used in various sensor configuration without losing their specificity. In addition, they can be synthesized by conventional “wet” chemistry techniques and may be easily modified with labels or linkers for immobilization to sensing elements.
**Peptides:** can be chemically synthesized in large quantities, are amenable to derivatization and smaller than antibodies.

- Phage display has been used to identify peptide ligands that target certain cell types by panning on well-characterized cell surface receptors. Furthermore, peptides can display high affinities for cell surface components, making them attractive ligands for cell recognition (Gui et al. 1996; Burg et al. 1999). The peptides are selected in a random way from large, unfocussed, and often preexisting and commercially available phage display libraries, with no design elements.

- Designed synthetic peptides may be also used on the basis of known interactions between single or a few amino acids and targets, with no attempt to build highly organized binding sites, but rather with attention being paid to the presence of peptide motifs known to allow intermolecular self-organization of the sensing peptides over the sensor surface.

- Artificial, miniaturized receptors can be obtained from the reduction of the known sequence of a natural receptor down to a synthesizable and yet stable one. Alternatively, a binding site can be created over a designed, stable peptide scaffold.

- Short peptides can also be used as active elements for the detection of their own receptors. Thus, antimicrobial peptides (AMPs) and cell-penetrating peptides (CPPs) are used for sensing bacterial cells, antigenic peptide sequences for antibody monitoring, and peptide substrates for enzyme detection (Pavan and Berti 2011)

**Aptamers:** can bind to their targets with high affinity and even discriminate between closely related targets. This is due to their adaptive recognition: aptamers, unstructured in solution, fold upon associating with
their targets into molecular architectures in which the ligand becomes an intrinsic part of the associated structure. In respect to antibodies, they have the advantage of overcoming the use of animal systems for the production of the molecule. Aptamers are isolated by *in vitro* methods that are independent of animals and can be generated against any target; the immune response to generate antibodies can fail when the target molecule (a protein) has a structure similar to endogenous proteins and when the antigen consists of toxic compounds that can even kill the animal. Moreover, aptamers can be designed in a signaling architecture, in which the aptamer structure changes upon association, thus producing a signaling event due to the molecular association. The production of aptamers is commonly performed by the systematic evolution by exponential enrichment (SELEX) process (Eaton et al. 1997; Brody et al. 1999; Ulrich et al. 2001; Cerchia et al. 2002).
2 Aim of the research

Molecular recognition plays a key role for both binding and catalysis in biological systems and enzymes, antibodies and receptors are typical examples. The specificity and activity of these macromolecules is due to the presence of binding sites in which they establish multiple interactions with the small molecules of the ligands, antigens or transition state. These molecules are extremely complex from a structural point of view, difficult to purify from natural sources and to synthesize, while the possibility of use as sensing element in biosensors or as catalysts in organic synthesis would be extremely valuable. For these reasons, the identification of peptide based artificial receptors or catalysts are one of the main objectives of biotechnology.

This project is aimed at obtaining new short peptides capable of both binding and catalysis.

The final goal of the project is to develop peptide libraries to be selected for both properties, and two different approaches have been exploited in order to generate stable peptide hosts.

- In the first one, a stable peptide is identified first a highly structured molecule to be used as a scaffold for building up a small receptor.
- In the second approach an existing natural protein binding site is reduced down to a short, yet still stable receptor.

The first approach has been addressed by using a coiled-coil structural motif and by this way several receptors, able to bind target molecules as xanthenes, have been obtained. The Sudlow binding site of Human albumin has been used in the second approach to obtain a one hundred amino acids peptide which retains binding and catalytic properties of albumin, to be used as the starting scaffold for a library of mutated albumin subdomains. Albumin fragment has been tested for its ability to interact with drugs as efavirenz and warfarin and to catalyze reactions as diketone stereoselective reduction and aldol addition.
Materials and methods

3.1 Abbreviations

AP, alkaline phosphatase
APS, ammonium persulfate
BCIP, 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
B-PER, bacterial protein extraction reagent
BSA, bovine serum albumin
CD, circular dichroism
dNTPs, deoxynucleotides
DTT, dithiothreitol
HSA, human serum albumin
HRP, horseradish peroxidase
IPTG, isopropyl β-D-1-thiogalactopyranoside
MW, molecular whieghts
NBT, nitro-blue tetrazolium
O/N, over night
PEG, polyetilenglicole
PPB, periplasmic buffer
RT, room temperature
SDS-PAGE, dodium dodecyl sulphate - polyacrylamide gel electrophoresis
TMB, tetrametilbenzidine

3.2 Materials

HSA essentially free of fatty acids (A3782), (±)-warfarin (A2250) and, unless otherwise stated, all other reagents and solvents were from Sigma – Aldrich. SPR reagents, buffers (except PBS) and the research grade Sensor Chip CM5 research grade were purchased from GE Healthcare Bio-Sciences. Efavirenz and warfarin were synthesized as described below.
Phosphate buffer was used to carry out all CD and fluorescence experiments. PBS buffer was used in SPR analyses.

### 3.3 Solutions and buffers

- **Phosphate buffer saline (PBS)**  
  8 g NaCl, 0.2 g KCl, 1.44 g Na$_2$HPO$_4$, 0.24 g KH$_2$PO$_4$ in 1000 ml H$_2$O, final pH 7.4.

- **Phosphate buffer (PBS no salt)**  
  1.44 g Na$_2$HPO$_4$, 0.24 g KH$_2$PO$_4$ in 1000 ml H$_2$O, final pH 7.4.

- **Phosphate buffer saline Tween 0.1% (PBST)**  
  PBS added with 0.1% Tween 20

- **2XTY liquid broth for bacteria**  
  6.4 g Bacto-tryptone, 4 g Extract Yeast, 2 g NaCl, H$_2$O in 400 ml H$_2$O. If required ampicillin 0.1 mg/ml, kanamycin 0.05 mg/ml, 0.034 mg/ml chloramphenicol.

- **2XTY Agar plates**  
  6.4 g Bacto-tryptone, 4 g Extract Yeast, 2 g NaCl, 6 g Bacto-Agar in 400 ml H$_2$O. If required ampicillin 0.1 mg/ml, kanamycin 0.05 mg/ml, 0.034 mg/ml chloramphenicol.  
  TAE buffer for DNA electrophoresis on agarose gels  
  0.04 M Tris-acetate, 0.001 M EDTA.

- **Loading buffer 6x for DNA samples (agarose gel)**  
  40% glicerol, 60% H$_2$O, 0.10% (w/v) bromophenol blue.

- **SDS Running Buffer for protein electrophoresis on acrylamide gels**  
  25 mM Tris, 250 mM glicin, 0.1% SDS pH 8.3.

- **Loading buffer 2 for protein samples**  
  100mM Tris pH 6.8, 4% SDS, 0.2% bromophenol blu, 2% β-mercaptoethanol, 20% glycerol

- **Running gel for SDS-polyacrylamide gel**  
  12% polyacrylamide mix (29% acrylamide, 1% bisacrylamide), 0.375M Tris pH 8.8, 0.1% SDS, 0.1% APS, 2 μl TEMED (N,N,N',N'-Tetramethylethylenediamine), H$_2$O to 10 ml.
Materials and methods

- **Stacking gel for SDS-polyacrylamide gel**
  5% polyacrylamide mix (29% acrylamide, 1% bisacrylamide), 0.125M Tris pH 6.8, 0.1% SDS, 0.1% APS, 1 μl TEMED (N,N,N’,N’-Tetramethylethylenediamine), H₂O to 4 ml.

- **Loading buffer 2x for protein samples (acrylamide gels)**
  100 mM Tris pH 6.8, 4% SDS, 0.2% bromophenol blue, 2% β-mercaptoethanol, 20% glycerol.

- **AP buffer**
  Tris 1 M pH 9.5, NaCl 100 mM, MgCl₂ 5 mM

- **Lysis buffer**
  20 mM Tris pH 8.00, 500 mM NaCl, 0.1% Triton x100

- **TSS for preparation of competent E. coli cells**
  85% 2xTY, 10% PEG mw 8000, 5% DMSO, 50 mM MgCl₂. Filtration with 0.2 μm filter

- **Glycerol buffer for preparation of electrocompetent E. coli cells**
  10% glycerol in milli-Q H₂O. Autoclave before use.

- **PPB for extraction of periplasm fraction**
  200 mg/ml Sucrose, 1 mM EDTA, 30 mM TrisHCl pH 8.0.

- **Elution buffer**
  50 mM GSH, 100 mM NaCl, in PBS pH 8.0.

3.4 Bacterial strains

The bacterial strain used in this study was:
- *Escherichia coli* DH5αF’ (Gibco BRL), F’/endA1 hsd17 (rK-mK+) supE44 thi-1 recA1 gyrA (Nalr) relA1 (lacZYA-argF) U169 deoR (F80lacD- (lacZ)M15);
### 3.5 Oligonucleotides

All primers were purchased from Tib Molbiol – Roche.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGEX sense</td>
<td>gggctggcaagccacggttgggtg</td>
</tr>
<tr>
<td>pGEX anti</td>
<td>cggggagctgcatgtgctagaggttttcacc</td>
</tr>
<tr>
<td>HSA optimized</td>
<td></td>
</tr>
<tr>
<td>external sense</td>
<td>gctcgccgcagcatgcagcagcaaaaac</td>
</tr>
<tr>
<td>HSA optimized</td>
<td></td>
</tr>
<tr>
<td>mut. Ala215 anti</td>
<td>cttctgcaaaatctcttttcggaaacgcgtgctcagacgtaacgacnnnccgcactttaaatgc</td>
</tr>
<tr>
<td>HSA optimized</td>
<td></td>
</tr>
<tr>
<td>mut. Leu238 sense</td>
<td>gaaagcagattaattgcagaagtttagcgaacctgatgaccgatnnkaccaagttccataccg</td>
</tr>
<tr>
<td>HSA optimized</td>
<td></td>
</tr>
<tr>
<td>anti</td>
<td>atctgcacgatcatctgcac</td>
</tr>
<tr>
<td>HSA optimized</td>
<td></td>
</tr>
<tr>
<td>mut. Leu260-1le264 sense</td>
<td>gcaagatgcgtgagatnnkccaaatattnnkctggaaaccagagatagcattaggc</td>
</tr>
<tr>
<td>HSA optimized</td>
<td></td>
</tr>
<tr>
<td>external anti</td>
<td>agctgctagcttccactctgc</td>
</tr>
<tr>
<td>VLPT1</td>
<td>cgctggagttgtattactgcaagagcggcgcgcagcctggc</td>
</tr>
<tr>
<td>VHPT1</td>
<td>ccaggccccagctgggattttgggattttgctggctta</td>
</tr>
<tr>
<td>VHPT2</td>
<td>tgg tga tgg tga cta cta ccc agg ccc agg cag ggg agg ggg tgg tgg tta tta ttc</td>
</tr>
<tr>
<td>VLPT2</td>
<td>tac cta ttg cct acg gca gcc gct gga ttg tta tta ttc</td>
</tr>
<tr>
<td>Kinv-cys-New</td>
<td></td>
</tr>
<tr>
<td>random1 sense</td>
<td>tg gca agc gtt aaagagaaactcggactttacgtaaa nnk nnk nnk nnk ggaacactggactcggtaag</td>
</tr>
<tr>
<td>Kinv-cys-New</td>
<td></td>
</tr>
<tr>
<td>no NheI anti</td>
<td>cc accgggt acc gcc tcc ggc ttt tac gct ggc gag ctt ttc ctt tac gga cgc aag ttt ctc</td>
</tr>
<tr>
<td>Kinv-cys-New</td>
<td></td>
</tr>
<tr>
<td>random2-sense-BssHII</td>
<td>agcaagcgc gcgcgc actgagccctggcggaga actagt gag aag ctc gca aga agg aag ctc</td>
</tr>
</tbody>
</table>
| Ecys-New-random1 sense      | agt ggc ctg gag aag gag gtt agc gct ctt gag aannnnknknknknknknknknknk gga aannnnknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknknk
3.6 Experimental methods

3.6.1 PCR (Polymerase Chain Reaction)

The proof-reading Phusion Pyrococcus-like DNA polymerase (Finnzymes) was used for construction of coils sequences and libraries.

Reaction mixture:
- Template DNA 0.01-1 ng (plasmidic DNA)
- Sense primer 0.5 pmol/μl
- Antisense primer 0.5 pmol/μl
- Phusion HF Buffer 5x
- dNTPs (Sigma) 0.25 mM
- Polymerase 0.025 units/μl
- H₂O to 50 μl

The following cycles were performed:
1) Denaturation step, 30’ at 98°C.
2) 31 cycles of: denaturation, 15” at 98°C; annealing, 15” at 60°C; elongation, 30” every 1000bp at 72°C.
3) Final elongation step: 10’ at 72°C.

Thermus termophilus DNA polymerase (Biotools) was used to validate the cloning steps.

Reaction mixture:
- Template DNA 0.01-1 ng (plasmidic DNA)
- Sense primer 0.5 pmol/μl
- Antisense primer 0.5 pmol/μl
- Biotools Buffer 10x
- dNTPs (Sigma) 0.25 mM
- MgCl₂ 2 mM
- Polymerase 0.025 units/μl
- H₂O to 20 μl

The following cycles were performed:
1) Denaturation step, 5’ at 94°C.
2) 31 cycles of: denaturation, 45” at 94°C; annealing, 45” at 60°C;
elongation, 1’ every 1000bp at 72°C.

3) Final elongation step: 10’ at 72°C.

3.6.2 DNA electrophoresis on agarose gel

Agarose (Sigma) gels with a concentration of 2-1,5% in TAE buffer were
used to separate PCR products; 0,8% agarose gels were used to separate
plasmidic DNA preparations. 1 μl of ethidium bromide (Sigma) was added to 30
ml of solution. 100 and 1000 base-pairs molecular weights were purchased
from Fermentas.

3.6.3 DNA purification

The GenElute Gel Extraction Kit and GenElute PCR Clean-up kit (Sigma)
were used for purification of DNA agarose gel and reaction mixtures
respectively, following the instruction of the manufacturer.

The NucleoTraP CR (Machery-Nagel) was used for purification of mixtures of
libraries ligations.

3.6.4 DNA digestion with restriction endonucleases

All restrictions endonucleases (BssHII, Nhel, EcoRI, HindIII) were
purchased from New England Biolabs.

Reaction mixture:

- DNA;
- NEB buffer 10x;
- BSA 100x (if necessary);
- Restriction endonuclease, 1 unit/μg of DNA;
- H₂O to 50 μl.

The incubation was performed at the temperature required by the specific
enzyme.
3.6.5 DNA Ligation

Vector and insert DNA were mixed at the experimental-required ratio, as number of molecules (1:3 – 1:5). T4 ligase from New England Biolabs was used:

Reaction Mixture:
- DNA
- T4 ligase buffer (NEB) 10x;
- T4 ligase (NEB), 1 unit;
- H$_2$O to 15 μl;

Incubation O/N at 16°C.

3.6.6 Plasmidic DNA extraction

The GenElute Plasmid Miniprep (Sigma) was used for plasmidic DNA mini-preparations, starting from 1-5 ml of bacteria culture, and following the instruction of the manufacturer.

3.6.7 DNA sequencing

PCR products were purified with Eppendorf Perfectprep Gel Cleanup kit. Reaction mixtures for sequencing was composed as follows:
- 50-100 ng of purified PCR product;
- 1 μl primer (3,2pmol/μl);
- 2 μl Terminator Mix (Applied Biosystems, BigDye Terminator v1.1 Cycle Sequencing Kit);
- 2 μl buffer 5x;
- H$_2$O to 10 μl;

and the sequencing program was:
1) 1’ at 96°;
2) 25 cycles: 15” at 96°; 5” at 50°; 4’ at 60°;

Reactions were afterwards purified with CENTRI SEP Spin Columns, following manufacturer instructions. 5 μl of purified sequences were loaded on
sequencing plates with 10 μl of formamide, denaturated for 2’ at 96° and analyzed with 3100 Genetic Analyzer sequencer (ABI PRISM-HITACHI).

### 3.6.8 Preparation of competent E. coli cells

50 ml of *E. coli* cells were grown at 37°C in 2XTY liquid broth to OD$_{600}$ 0.5. The bacteria were immediately chilled in ice for at least 20’, then centrifuged at 4°C for 5’ at 1200 rpm. The supernatant was discarded and the bacterial pellet resuspended in 5 ml of sterile TSS. The cells were immediately used or stocked at -80°C for two months.

### 3.6.9 Bacterial transformation

A 100 μl vial of competent cells was incubated for 20’ with 5 μl of ligation reaction mixture. Heat shock was applied at 42°C foe 1’15”; after 2’ in ice, the bacteria were resuspended in 1 ml of liquid 2XTY and allowed to grow for 1 hour. No antibiotic was added. Bacteria were then plated on antibiotic-containing agar plates and grown O/N at 28-30°C.

### 3.6.10 Preparation of electrocompetent cells

For the preparation of electrocompetent cells, the following protocol was applied:

- from an O/N culture of bacteria, 200 μl of culture was added to 200 ml of 2XTY liquid broth, and allowed to growth at 37°C to OD$_{600}$ 0.8;
- the whole culture was splitted in 4 50ml-tubes, and chilled for 30’;
- centrifugation at 5000 rpm, at 4°C for 8’;
- the supernatant was discarded, and pellet was resuspended with 5 ml of glycerol buffer;
- glycerol buffer was added to 50 ml and the tubes were chilled for 10’;
- centrifugation at 5000 rpm, at 4°C for 10’;
- the supernatant was discarded, the pellet resuspended with 5 ml of glycerol buffer and the suspensions collected in two 50ml-tubes;
- glycerol buffer was added to 50 ml and the tubes were chilled for 10’;
Materials and methods

3.6.11 Bacterial transformation of electrocompetent cells

The ligation mixture were purified and used to transform electrocompetent bacteria. A Eppendorf 2510 electroporator was used. 1 μl of purified ligation product was added to one aliquot of cells (35-40 μl), and chilled in ice for 30''. The bacteria were transferred into a cuvette (Eppendorf): after the electroporation cycle (at 1800 volts), the bacteria were resuspended with 1ml of 2XTY, allowed to grow at 37°C for 1 hour and plated on antibiotic-containing agar plates and grown at 28-30°C.

3.6.12 Preparation of bacteria for -80°C stock

The bacteria were collected and resuspended with a solution of 80% 2XTY and 20% glycerol.

3.6.13 Preparation and transformation of pDAN-scE-Kinv random library

Each coil E and K was separately constructed and randomized in the central heptad.

Both for K and for E random coil two subsequent cycles were performed. Kinv cys new random1 sense/Kinv cys new no NheI anti primers were used to
construct the inverted K-coil random. The sense and antisense primers were annealed together by their complementary sequences; the amplified fragment was used as template for the second PCR product. By using this product as template, Kinv cys new random2 BsshII sense and Kinv cys new no NheI anti primers were used to produce the full and randomized K inverted.

The E randomized fragments were separately produced, performing two subsequent cycles of PCR amplification, and using this primers: Ecys new random1 sense/Ecys new anti 1 Nhe and Ecys new random2 sense linker/Ecys new anti 2 Nhe PT1.

After purification of final PCR products, the library was created by assembling Kinv randomized fragments with E randomized fragments. The assembly was performed using specific cycles of PCR amplification:

- 9 cycles with Kinv and E randomized fragments mixed at 1:1 ratio (number of molecules): this step allowed the assortment of fragments and the assembly;
- 24 cycles: using external primers (VLPT1/VHPT1), the single chain sequences were re-amplified, and the restriction endonucleases sites were included at the ends (BssHII/NheI).

The scE-Kinv random library (275 μg) and the pDAN vector (1200 μg) were both digested with BssHII and Nhel. After purification, they were mixed in a ligation mixture at a 1:3 ratio. The ligation was performed O/N at 16°C, and purified using NucleoTraP CR Kit (Machery-Nagel) and resuspended in 20 μl water. 15-20 aliquots of electrocompetent cells were transformed, each one with about 1 μl of DNA. After electroporation cells were pooled and grown in 5 ml of 2XTY liquid broth, for 1 hour at 37°C, and then plated onto 625cm²-plates with ampicillin 100μg/ml. Dilution were plated onto a small plate with ampicillin to assess the size of the library.

After 15 hours of incubation at 30°C, the library was inspected by PCR, harvested and stocked in small aliquots at -80°C with 20% glycerol.
3.6.14 Expression of recombinant proteins

Clones were picked from agar plates and grown at 37° in 2 ml of 2XTY. At OD$_{600}$ 0.5, IPTG 0.2 mM was added to induce the production of the recombinant protein, allowing growth O/N at 28°C.

To obtain the total fraction, 100 μl of bacterial culture were centrifuged at 5000 rpm for 5', the supernatant discarded, and the pellet resuspended with 50 μl of PBS.

To obtain the periplasmic fraction, the remaining bacterial broth were centrifuged at 5000 rpm for 5', and the supernatant discarded. The pellet was resuspended in 1:10 of culture volume of PPB, chilled in ice for 1 hour, vortexing every 10'. The mixture was centrifuged at 11000 rpm for 15'. The supernatant containing the periplasmic proteins (periplasmic fraction) was collected.

3.6.15 Production and PEG-precipitation of phages

Small aliquots of the bacterial stock of the library was grown in 10 ml of 2xTY added with ampicilline and 1% glucose, at 37°C to OD$_{600}$ 0.5. Under these conditions, bacteria express the pilus necessary for phage infection. Bacteria were infected with a wild-type helper phage, carrying all the genes necessary for phage replication, at a MOI (multiplicity of infection, that is the ratio viral particles/bacterial cells) of 100. This step was performed at 37°C for 45', without agitation.

Bacteria were then centrifuged at 4000 rpm for 10’, the supernatant discarded and the pellet resuspended in 40 ml of 2xTY broth added with ampicilline (phagemid resistance) and kanamycin (helper phage resistance), to select bacteria carrying the phagemid and the phage genome. Bacteria were allowed to grow O/N to produce recombinant phages.

For collection of recombinant phages, bacteria were centrifuged at 7,000 rpm for 20’. The supernatant, containing soluble phages, was collected.

Precipitation of phages was performed by adding a solution of PEG 20% / NaCl 2.5 M. 1/5 volume of PEG/NaCl was added to the phage-containing supernatant. The solution was incubated on ice for 45’ to allow precipitation,
and then centrifuged at 7,000 rpm for 20’ to pellet phages. The supernatant was discarded and the white pellet resuspended in 1 ml of PBS.

3.6.16 Library selection with magnetic beads

The phages were produced, PEG-precipitated and resuspended in 250 µl of PBS (about 10^{12} phages). Phages were then incubated in a well of ELISA plates (Sarstedt), previously pre-incubated with MPBS (the buffer used for the selection), to remove non-specific phages. The phage solution was moved into another well, still pre-incubated with MPBS: this step was repeated 10 times.

The phage solution was moved in a new well, previously pre-incubated with streptavidin (New England Biolabs) 10 µg/ml, and blocked with MPBS at RT for 10’. The solution was moved into a new well, treated in the same way: this step was repeated 5 times. 250 µl of MPBS and biotinylated caffeine (10 µg for the first round, 50 µg for the following rounds) were added. To the phage solution: the mixture was incubated and gently shaken at RT for 1 hour. Paramagnetic streptavidin-coated beads were added and shaken for 1 hour. After extensive washes with PBS/0.1% Tween (PBST) and PBS (to remove all unbound and unspecific phages), elution was performed incubating beads at RT for 30’ with a 100 ug/ml solution of caffeine without linker nor biotin.

After this selection, 9 ml of DH5α at OD_{600nm} 0.5 were infected with the recovered phages, at 37°C for 45’. Then, bacteria were centrifuged at 2500 rpm for 15’ (15°C), the supernatant discarded, the pellet the resuspended with 1 ml of 2XTY broth and plated, allowing the growth O/N at 37°C. The day after, bacteria were harvested and stocked at -80°C in glycerol. A small aliquot was immediately let to grow in order to produce new phages for a second round of selection. Four passages of selections were carried out, with increasing stringency of washing steps.

3.6.17 Phage ELISA

The phage ELISA assays, by using the selected phage library, were performed as follow:
Materials and methods

- coating: each well was incubated with 100 μl of coating solutions, containing the ligand (table 3.2) O/N at 4°C;
- blocking: after extensive washes with PBS, the wells were blocked with either 140 μl of MPBS or PBS Tween 0.3%, at RT for 1 hour;
- phage incubation: after extensive washes with PBS, 50 μl of page solution and 50 ul of blocking solution were added, at RT for 1 hour;
- extensive washes with PBST and PBS;
- secondary antibody: mouse antiM13 HRP conjugated (Amersham, Pharmacia) was diluted 1:2000 in the blocking solution and added to the wells at RT for 1 hour.
- extensive washes with PBST and PBS;

Immunocomplexes were revealed with the chromogenic substrate TMB (Sigma), and the plate read at OD₄₅₀.

<table>
<thead>
<tr>
<th>Coating solutions</th>
<th>Sarstedt plates</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ 10 μg/mL of streptavidin solution <em>(New England Biolabs)</em>, diluted in PBS. 100 μl each well, O/N at 4°C</td>
<td></td>
</tr>
<tr>
<td>✓ 20 μg/mL of biotinylated ligand, diluted in PBS. 100 μl each well, 1 hour and 30’ a RT.</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2 – Different coating solutions for Sarstedt plates

3.6.18 Cell Cultures: cell lines and media

After bacteria transformation and clones isolation, plasmids are extracted from bacteria (see section 3.1.1) and transfected into eukaryotic cells. For this project, Chinese hamster ovary (CHO-S) cells are employed. These cells are maintained at 37°C with 5% of CO₂ and in CHO-S Serum Free Medium (CHO-S-SFM II; GIBCO, Invitrogen) in addition to Penicillin and Streptomycin mixture (20 mL/L, Lonza) and L-glutamine (2 mM) (Lonza).

3.6.19 Transfection

To transfet CHO-S cells, the FreeStyleTM Max Reagent (Invitrogen) was employed, because of its high transfection efficiency in cells that grow in
Materials and methods

suspension. Moreover, different culture media are used to grow CHO-S cells before transfection, to dilute the FreeStyleTM Max Reagent and the DNA of interest and to select the transfected cells (table 3.3).

<table>
<thead>
<tr>
<th>Media used for transfection</th>
<th>CHO-S Serum Free Medium (CHO-S-SFM II; Gibco, Invitrogen)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media used for selection of transfected cells</td>
<td>Pro CHO5 (Lonza), Penicillin and Streptomycin mixture (20 mL/L, Lonza), L-glutamine (2 mM) (Lonza), hygromycin B (200 µg/mL) (Gibco, Invitrogen)</td>
</tr>
</tbody>
</table>

Table 3.3 – Culture media used for transfection and for the subsequent selection.

The day before transfection, cells are plated in 24 wells plate (1 mL of media) so that the coverage percentage reaches about the 70-80% value, without any cellular aggregate.

The reaction mix is composed by: 200 µL of media, 1 µg or 2 µg of the plasmid of interest and 1 µL or 2 µL of FreeStyleTM Max Reagent respectively. All these components are put together in the same tube and incubated for 20 minutes at room temperature, before their addition to cells.

3.6.20 Post-transfection analysis, polyclonal and monoclonal selection and expansion

Three days after transfection, diluted 1:1 and 1:10 supernatants are analyzed through an ELISA assay (see paragraph 3.6.21) to evaluate proteins production. If positive results are obtained, cells are initially transferred in a 6 wells plate (3 mL) where they grown for 2-3 days as “polyclonal” clones. This definition is due to the different way in which cells are transfected and their subsequently diverse capability to produce the protein of interest. For each expansion step, cells are centrifuged for 5 minutes at 180 x g and resuspended in the desired volume, adding hygromycin B to start the selection of the transfected cells. Polyclonal cells are then used to make monoclonal clones through limiting dilutions (0-1 cell in each 96 wells plate) and to select those that produce the highest amount of the interested proteins. The so obtained monoclonal clones are grown in selection for 2-3 weeks until they cover 50% of the well surface, when they can be tested through an ELISA assay. The best
producing clones are expanded up to a 75 cm$^2$ flask to be frozen. A little amount of 1 of the obtained monoclones cells are maintained and expanded up to a 150 cm$^2$ flask and subsequently to a bioreactor CELLine (BD), which allows a large quantity of protein production. Every 3-4 days cells are centrifuged at 180 x g to collect the supernatants, in which the protein is secreted, and diluted 1:10. In the meantime, the polyclonal clones are expanded up to a 75 cm$^2$ flask to be subsequently frozen. Supernatants, obtained after both monoclonal and polyclonal cells centrifugation, are tested through an ELISA assay and then stored at -20°C.

### 3.6.21 Productivity and specificity ELISA assay

Enzyme-linked immunosorbent assay (ELISA) is a quantitative biochemical technique mainly used in immunology to detect the presence of an antibody or an antigen in a sample. This simple technique is based on the immobilization, through adsorption, of the molecules of interest on a solid support that usually is a polystyrene microtiter plate. For the detection, a primary antibody, able to recognize the molecule of interest adsorbed on the well surface, and a secondary antibody linked to an enzyme through bioconjugation are used. Finally, a chromogenic substrate is added and the evaluation of the proteins’ quantity is obtain through a spectrophotometric measurement (450 nm).

Two different kind of ELISA tests are made: one that evaluates the amount of produced protein (productivity ELISA) and the other one which evaluates the capability of the produced antibodies to bind their specific antigen (specificity ELISA).

The productivity ELISA is characterized by the coating of wells (E.I.A/R.I.A 8 well strip, Costar) O/N at 4°C with 100 µl of transfected CHO-S supernatant, in which there is the protein of interest. The blocking is made with 1500 µl of 2% MPBS for 30 minutes at room temperature and then the α-SV5 primary antibodies, diluted 1:2000 in 2% MPBS, are added and incubated for 1 hour and 30 minutes at room temperature. To eliminate all the unbound primary antibody, three washes with PBS-Tween 0,1% followed by three washes with PBS are made. Subsequently, an α-mouseIgG Fc conjugated with the enzyme
called horseradish peroxidase (HRP) (Jackson) 1:2000 dilution in 2% MPBS is added. To remove all the unbound secondary antibody, three washes with PBS-Tween 0.1% followed by three washes with PBS are made. Finally, 70 µL of 3,3′,5,5′-Tetramethylbenzidine (TMB) (Sigma) are added, to allow the reaction development. The reaction is stopped when 30 µl of H2SO4 (Sigma) 1 M are added and the detection is performed with a spectrophotometric measurement at 450 nm (ELISA reader, Anthos 2020).

For what concerns the specificity ELISA, the coating is made using 100 µL of a specific antigen (10 µg/ml) and the primary antibody is represented by the CHO-S supernatants, where the minibodies are produced and secreted. In this reaction, α-SV5 and α-mouse IgG Fc HRP are respectively considered as the secondary and the tertiary antibodies. All the incubation times are like the productivity assay.

3.6.22 Cells freezing

The clones of interest are centrifuged for 5 minutes at 180 x g at 24°C and resuspended in 1 mL of fetal bovine serum (FBS) (GIBCO, Invitrogen) with 10% of Dimethyl sulfoxide (DMSO). Then, they are stored for 24 hours at -80°C and then moved in liquid nitrogen.

3.6.23 HSA100-GST protein production, purification and dialysis

Bacterial clones were grown in 100 ml of 2XTY liquid broth, added with 100 µg/ml ampicillin, at 37°C to OD600 0.6; the expression of the recombinant protein was induced by adding 0.2 mM IPTG, and grown O/N at 20°-25°C. The next day, the recombinant proteins were extracted as follows:

- the bacterial culture was centrifuged for 20'-30' at 4000 rpm (4°C), the supernatant discarded and the pellet resuspended with 10 ml of lysis buffer/gram of bacteria (20mM Tris pH 8.00, 500mM NaCl, 0.1% Triton x100);
- the mixture was incubated in ice, in gentle shaking, for 15’;
Materials and methods

- the lysozime was added at the mixture, 400 μg/ml of lysate, and the solution was incubated in ice, in gentle shaking, for 30’;
- DNAse (50-100 μg/mL of lysate), PMSF (1 mM) and EDTA (1 mM) were added to lysate, and incubated in ice, in gentle agitation for 45’;
- Sonication (Bandelin Sonoplus) was performed: 3-10 cycles, 1-2 minutes for cycle;
- After centrifugation for 20’ at 11000 rpm, the supernatant was collected and filtered with 0.45 μm filter.

After centrifugation the supernatants were collected, run through a 0.2 μm filter (Millipore) and affinity purified with Glutathione Sepharose 4B (GE Healthcare), following the manufacturer’s instruction. Briefly:

<table>
<thead>
<tr>
<th>Steps</th>
<th>GST purification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding</td>
<td>400 μl of glutathione sepharose was added to the lysate, and incubated for 1 hour in rotation;</td>
</tr>
<tr>
<td>Packing</td>
<td>The resin was packed in a 10 ml-column, and extensively washed with PBS pH 8.0</td>
</tr>
<tr>
<td>Elution</td>
<td>The proteins were eluted with a solution of 50mM GSH (Sigma), 100mM NaCl, in PBS (0.250 ml each elution).</td>
</tr>
</tbody>
</table>

Table 3.4 - Main steps used for affinity chromatography of GST fusion protein.

Fractions were then pooled and dialyzed in PBS O/N at 4°C. The next day proteins were again centrifuged at 14000 rpm (4°C) for 30’. Protein production, degradation, purity and concentration were then checked by SDS-PAGE followed by Coomassie Blue staining and western blotting, carried out using anti flag or anti GST antibodies.

3.6.24 Protein electrophoresis on polyacrylamide gel (SDS-PAGE)

Samples, diluted 1:1 in loading buffer 2x for proteins, were loaded on 12% polyacrylamide gel, composed by a stacking and running gel. The proteins were separated at 18-20 mA.
3.6.25 Coomassie staining

The SDS-polyacrylamide gel was removed from the glass and stained with Coomassie solution, under gentle shaking for 1 hour at 37°C. Then, Coomassie solution must be removed (it can be reused many times) and the gel rinsed with destaining solution on a slow shaker for 6-12 hours.

3.6.26 Western blot

Proteins were separated by SDS-PAGE and transferred onto nitrocellulose membrane (Schleicher-Schuell) by semidry blotting using the Trans-Blot SD Transfer Cell (Biorad). The membrane was treated as follows:

- blocked with MPRB 1 hour at RT (20°C);
- primary antibody: mouse anti-FLAG 1:5000 (Sigma), mouse anti-GST 1:2000 (Sigma); 1 hour at RT;
- two washes with PBST and two washes with PBS;
- secondary antibody: goat anti-mouse AP conjugated (Jackson Immunoresearch), 1 hour at RT or O/N;
- two washes with PBST and two washes with PBS;
- Immunocomplexes were revealed by adding the chromogenic substrate NBT-BCIP (Sigma) in AP buffer, resulting in an insoluble precipitate at positive bands.

3.6.27 Bradford protein assay

The following protocol was used:

add 30 sample volumes of Bradford reagent (Sigma) to protein sample: mix the solution and incubate at RT for 20’;

- make the “blank”, adding the Bradford reagent to the buffer used for the solubilization of proteins;
- make a dilution series of BSA (bovine serum albumin, Sigma) in the range 50 ng - 10 μg;
- cool samples to RT and measure the absorbance at 562 nm;
- plot the average-blank corrected measurement for each standard vs. concentration.
Use the slope of this standard curve to estimate the protein concentration of the unknown samples.

### 3.6.28 HSA100-GST ELISA assay

The ELISA assay with purified proteins was performed as follows:

- Streptavidin (SIGMA) was diluted to 10 μg/ml in PBS and 100 μl were coated in ELISA wells (Costar), O/N at 4°C;
- Biotinilated molecules (efavirenz or warfarin) were diluted to 1 μg/ml in PBS incubating at RT for 1 hour;
- the wells were extensively washed with PBS and 120 μl of blocking solution (MPBS) were added to each well, incubating at RT for 1 hour;
- the blocking solution was discarded; purified HSA100-GST proteins were diluted 1 μg/ml in PBS, and added 100 μl to each well;
- extensive washes were performed with PBST and PBS;
- primary antibody: mouse anti-GST, diluted 1:2000 in MPBS. 100 μl to each well, at RT for 1 hour;
- extensive washes were performed with PBST and PBS;
- secondary antibody: goat anti-mouse HRP conjugated (Jackson Immunoresearch), diluted 1:2000 in MPBS. 100 μl to each well, at RT for 1 hour;
- after extensive washing with PBST and PBS, the immunocomplexes were revealed with the chromogenic substrate TMB (Sigma), and the plate read at OD$_{450}$.

### 3.6.29 LC-MS/MS identification of HSA100-GST

The fusion protein was digested with trypsin (1:10 enzyme/substrate ratio) in 50 mM NH$_4$HCO$_3$ for 16 h at 37 °C. 3 μg of trypsin digest were analyzed by LC-MS/MS using a Capillary HPLC apparatus (Agilent 1200) interfaced to an HCT Ultra IT (Bruker Daltonics). HPLC separations were performed with a Zorbax 300SB-C18 3.5 μm, 150 × 0.5 mm column (Agilent) using a water/acetonitrile gradient (from 2% to 60% acetonitrile in 60 min) and 0.1% formic acid as a modifier. MS/MS parameters were as follows: full MS
scan, m/z 300-1300; number of precursors selected, 3; fragmentation amplitude, 1.00 V with Smart Frag active (from 50 to 200% of the fragmentation amplitude selected); MS/MS scan, m/z 100-2300; active exclusion after accumulation of 2 MS/MS spectra for a period of 2 min; ion charge control (ICC) active allowing the storage of a maximum of 200000 ions in a maximum accumulation time of 200 ms. LC-MS/MS analyses were elaborated with Data Analysis software (v.3.4, Bruker Daltonics). Masses of tryptic peptides were searched by both exporting our analyses as mgf files as input for the MASCOT MS/MS Ion Search option (www.matrixscience.com) and by manual inspection of entire analyses.

3.6.30 CD spectroscopy

Circular Dichroism spectra were obtained on a Jasco J-710 spectropolarimeter under nitrogen flux. Spectra were recorded in a 0.2 cm quartz cuvette at temperature ranging from 5 to 75 °C. Data were collected with a data pitch of 0.2 nm for a single spectra measurement, a scanning speed of 50 nm/min and a band width of 1 nm. Each spectrum was the average of 10 scans. Prior to analysis, each spectrum had the appropriate background spectrum subtracted. A 1.35 x 10^-6 mol L^-1 stock solution of GST-HSA100 was prepared by dissolving 5.14 mg of purified peptide in 100 mL of phosphate buffer. This solution was then diluted to 135 nM polypeptide in a a final 3 mL volum in phosphate buffer to carry out the measurements.

3.6.31 Fluorescence assay

Warfarin and efavirenz stock solutions were prepared in acetonitrile. Steady state fluorescence spectra were recorded at 298° K on a CARY Eclipse (Varian) spectrofluorimeter equipped with a 0.5 cm quartz cuvette. Synchronous fluorescence spectra (SFS) were measured by setting the excitation wavelength in the 240 - 320 nm range, and the emission was recorded at \( \Delta = 60 \) nm in the 300 - 380 nm range. The width on the excitation and emission slits was set to 5.0 nm for all measurements. The concentration of protein (HSA, GST, or GST-HSA100 was kept fixed at 1x10^-6 mol l^-1 in 350 μl phosphate buffer, and the
concentration of warfarin or efavirenz was gradually increased (from 1 μM to 300 μM for warfarin, from 1 μ to 2 mM for efavirenz). After each addition of the ligand, the fluorescence intensity at the maximum emission wavelength – and the drift of such maximum - were measured after equilibrium had been reached (15 min).

3.6.32 SPR analysis

The efavirenz derivative 2b was suspended at 1 mM concentration in 1:10 Borate 8.5 (10 mM disodium tetraborate pH 8.5, 1 M NaCl) with 5% of DMSO. It was then coupled to the carboxylic acid groups of the dextran matrix on the Sensor chip CM5 using the conventional carbodiimide coupling chemistry (EDC/NHS coupling kit) in 1:10 HBS-EP buffer (0.1 M HEPES, 1.5 M NaCl, 30 mM EDTA, 0.5% v/v p20 surfactant) as running buffer. Excess active esters were deactivated with ethanolamine. The coupling reactions were carried out using a flow of 10 μl/min for 7 min, resulting in densities averaging about 400± 50 RU. Interaction analyses were carried out with a constant flow rate of 10 μL/min in PBS (53mM Na₂HPO₄·7H₂O, 12.5 mM KH₂PO₄, 70mM NaCl, pH 7.4) as running buffer with a 5 μM optimized concentration of HSA fatty free and HSA100-GST analytes. After each cycle both surfaces were regenerated with 30% Acetonitrile in NaOH 1mM. The sensorgrams were subtracted with the reference data from a deactivated flow cell and the interaction equilibrium constants (k_on and k_off) were calculated using the BIACORE X100 evaluation software.

3.6.33 HPLC analyses

Analyses of the reduction of diketone 3 and of the aldol addition of acetone to aldehyde 5 were carried out on a HP1100 series instrument equipped with a C18 Phenomenex Luna 5μ 150 x 4.60 mm column using appropriate mixtures of acetonitrile and water as eluant, and a flow rate of 0.5 ml/min, with spectrophotometric detection at 214 nm. Diketone 3 and its anti and syn diols 4 were separated with a 68:32 water/acetonitrile mixture.
Materials and methods

Aldehyde 5 and its aldol product 6 were separated with a 40:60 water/acetonitrile mixture.

### 3.6.34 Reduction of diketone 3

0.5 ml of a 43 μM solution of diketone 3 in acetonitrile was added to 0.5 ml of a 43 μM solution of GST-HSA100 in water and the mixture was kept at room temperature for 30 min. 0.5 ml of a 75 μM solution of NaBH₄ in water was then added at 20 °C and the resulting mixture was stirred at the same temperature for 2 hours. The mixture was acidified with trifluoroacetic acid (150 μl) and 1 ml ethanol was added. The mixture was then filtered over a 0.22 μm filter and analyzed by HPLC. Retention times were 24.7 min for diketone 3, and 10.0 and 8.0 min for trans and cis diol products.

### 3.6.35 Aldol reaction

200 μM of recombinant protein in phosphate buffer (Na₂HPO₄ 0.05 M, NaCl 0.5 M), pH 7.5 (90 μl) was incubated at 37°C for 15 min. 2 μl of an aldehyde 5 mother solution in acetonitrile to final concentration ranging from 0.5 to 2 mM, and 10 μl of acetone were then added. The samples were sealed and kept at 37°C. Initial velocities of the aldol reaction were obtained by measuring the concentrations of aldehyde and aldol products during the first 5% reaction by HPLC. Retention times were 6.3 and 15.1 min for aldol and aldehyde respectively. All the measurements were quadruplicated, and the observed velocities are:

<table>
<thead>
<tr>
<th>[aldehyde] (mM)</th>
<th>V_{aldehyde} (nM/s)</th>
<th>V_{HSA} (nM/s)</th>
<th>V_{HSA100} (nM/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>0.141 ± 0.021</td>
<td>6.45 ± 0.12</td>
<td>4.57 ± 0.39</td>
</tr>
<tr>
<td>0.75</td>
<td>0.242 ± 0.020</td>
<td>9.46 ± 0.17</td>
<td>5.62 ± 0.09</td>
</tr>
<tr>
<td>1.00</td>
<td>0.310 ± 0.006</td>
<td>12.0 ± 2.1</td>
<td>8.01 ± 0.11</td>
</tr>
<tr>
<td>1.50</td>
<td>0.354 ± 0.004</td>
<td>14.4 ± 0.3</td>
<td>12.26 ± 0.25</td>
</tr>
<tr>
<td>1.75</td>
<td>0.438 ± 0.031</td>
<td>14.3 ± 0.2</td>
<td>11.80 ± 0.90</td>
</tr>
<tr>
<td>2.00</td>
<td>0.313 ± 0.045</td>
<td>17.0 ± 1.7</td>
<td>12.81 ± 0.11</td>
</tr>
</tbody>
</table>
3.6.36 Preparation and transformation of pGex-HSA100 mutant libraries

Each HSA100 mutant library was separately constructed and randomized in the predicted positions: Ala 215 – Leu 238 and Leu 260 – Ile 264.

In order to create two libraries two cycles of PCR and one of PCR assembly were performed.

1) pGex-HSA100 Ala 215 – Leu 238 library

Starting from optimized HSA100 (ordered by Invitrogen GeneArt) as template, HSA optimized external sense/HSA optimized mut. Ala215 anti primers were used to insert the first mutation (PCR I). The second PCR was performed by using the same template and HSA optimized mut. Leu238 sense/HSA optimized external anti primers to insert the second mutation (PCR II). The PCR I and II products have an homology region to allow the annealing and the production of the full randomized HSA100.

After purification of final PCR products, the library was created by assembling PCR I and PCR II products. The assembly was performed using specific cycles of PCR amplification:

- 8 cycles with PCR I and PCR II fragments mixed at 1:1 ratio (number of molecules): this step allowed the assortment of fragments and the molecular assembly;
- 24 cycles: using external primers (HSA optimized external sens/HSA optimized external anti), the single chain sequences were re-amplified, and the restriction endonucleases sites were included at the ends (BssHII/Nhel).

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>98°C</td>
<td>3'</td>
</tr>
<tr>
<td>98°C</td>
<td>15&quot;</td>
</tr>
<tr>
<td>65°C</td>
<td>15&quot;</td>
</tr>
<tr>
<td>72°C</td>
<td>20&quot;</td>
</tr>
<tr>
<td>98°C</td>
<td>15&quot;</td>
</tr>
<tr>
<td>65°C</td>
<td>15&quot;</td>
</tr>
<tr>
<td>72°C</td>
<td>20&quot;</td>
</tr>
<tr>
<td>72°C</td>
<td>10&quot;</td>
</tr>
</tbody>
</table>

Table 3.5 - Main steps used for PCR assembly.
2) **pGex-HSA100 Leu 260 – Ile 264 library**

The same protocol was used for the pGex-HSA100 Leu 260 – Ile 264 library construction. HSA optimized mut. Leu 260–Ile 264 sense/HSA optimized external anti primers were used to insert the two mutations in the first PCR (PCR I) product of 133bp. During the second PCR step an amplified product of 216bp was obtained using HSA optimazed external sense and HSA optimized anti, inserting an homology region (18bp) with PCR I product to allow an assembly. After purification of final PCR I and II, the library was created by assembling PCR I and PCR II products. The assembly was performed using the same specific cycles reported in table 3.5.

The HSA100 Ala 215 – Leu 238 and HSA100 Leu 260 – Ile 264 random libraries (550 ng) and the pGEX vector (200 ng) were both digested with BssHII and Nhel. After purification, they were mixed in a ligation mixture at a 1:3 ratio for each. The ligation was performed O/N at 16°C. 4 aliquots of electrocompetent cells were transformed, each one with about 1 μl of DNA. After electroporation cells were pooled and grown in 2 ml of 2XTY liquid broth, for 1 hour at 37°C, and then plated onto 625cm²-plates with ampicillin 100μg/ml.

After 15 hours of incubation at 30°C, the libraries was inspected by PCR, harvested and stocked in small aliquots at -80°C with 20% glycerol.
4 Results and discussion

Chapter 1

4.1 Stable scaffold based on EK coiled-coil

The \textit{de novo} peptide design (see par. 1.1.3) of E and K coils domains is based on the knowledge that natural proteins can dimerize using \(\alpha\)-helical domains. (Chao et al. 1996; Tripet et al. 1996) (see par. 1.1.3). This E/K system was employed in purification systems and for biosensing applications, in virtue of the extremely stable and specific dimerization of E and K peptides. However, some experimental procedures often require extreme pH, redox and ionic strength conditions that could interfere or even disrupt the interactions between the coils and therefore alter the results of the assay. For this reason the development of improved coils with increased stability and specificity is needed.

The physical and chemical properties characterizing this kind of interaction have been deeply characterized using circular dichroism, a technique that allows to resolve the structure of biological macromolecules, in particular the secondary structure of proteins.

In order to validate \textit{in vivo} the data obtained with physical methods and to confirm dimerization capability and effective functionality of the coiled-coil domain a “Protein Complementation Assay” (PCA) was set up in our research laboratory (Dr. T. Gaiotto phd thesis 2009). By using this assay, it was possible to demonstrate the \textit{in vivo} interaction between different peptides containing coiled-coil domains. In particular, it was possible to prove that the E and K coils described by Tripet and Chao (Tripet et al. 1996) dimerize and have functional activity.

The aim of this thesis was to prove whether or not wild type E and K peptides could be considered as protein scaffolds for the generation of small binding sites by mutating the sequence in certain positions. To this purpose, the E and K coils, named “wild type E” (Ewt) and “wild type K” (Kwt), were cloned and validated by sequencing, PCR and western blot assays.

Generally, to use a protein as a scaffold some basic requirements should be considered:

- high stability;
• high interaction affinity;
• small size;
• easy expression in bacteria and eukaryotic cells;
• tolerance to sequence modification;

The tolerance to multiple substitutions and modification is a fundamental feature of scaffolds: the native protein folding must not be altered by the modification or insertion of additional residues, reflecting the elasticity of supporting structures (or frameworks) which can exhibit a novel binding capability. Moreover, these structures should be obtained by rational, or most commonly, combinatorial protein engineering techniques prone to further improvements and modifications (Nygren and Skerra 2004). To do this all parameters have been taken in consideration, according to previous reports (Chao et al. 1996; Tripet et al. 1996; Chao et al. 1998; Krylov et al. 1998; Woolfson 2005; Ayriss et al. 2007). In order to assess the tolerance to modifications of the sequence, some mutations were inserted into wild type E and K sequence by random mutagenesis. We have thus performed a first trial to obtain an E-K based artificial host.

4.1.1 Single chain E-Kinv design

Following previous studies carried out in our research group, the affinity quality of protein scaffolds was ameliorated mutating specific residues into the sequence and selecting new peptides with improved features (Pelletier et al. 1998; Arndt et al. 2000; Arndt et al. 2002; Mason et al. 2006; Mason et al. 2007; Hagemann et al. 2008; Jouaux et al. 2008).

In this work a strategy of RANDOM MUTAGENESIS was adopted to improve the E-K heterodimeric coiled-coil domain and to find the best peptide binders. We designed a single chain E-K as a new protein scaffold, that could be used as biosensor unit in many applications.

The choice of a single chain peptide, rather than of a heterodimer, was dictated by practical reasons, with the aim of generating an easy to handle library.

The overall designed structure is reported in figure 4.1 and contains a 35 amino acids coiled-coil interface (a sequence long enough to ensure stability,
yet still synthesizable). In order to build up the peptide we have taken into account the following points, starting from the structure of the coiled-coil domain proposed by Chao and coworkers (Chao et al. 1996; Tripet et al. 1996):

1) In its original design, the E-K coil is a parallel heterodimer. On planning a single chain equivalent, we need first to connect the two chains with a linker capable to adopt a loop conformation: a linker was thus inserted between the coils: it is composed by eight residues (seven glycines and one threonine), that showed to confer a marked flexibility to the whole molecule and low steric hindrance. This kind of Thr/Gly linker is commonly used in molecular biology technique in order to build up proteins constructs arising from the fusion of different domains (Zoldak et al. 2009).

2) As a consequence, the two helical domains of the peptide are now devoted to form an antiparallel coiled coil, rather than a parallel one. The amino acid sequences of the interface are thus unpaired, and in order to restore the amino acid pairing, the K sequence was inverted from KVSALKE to EKLASVK, to allow the interaction through a parallel-like association resembling that of the original model; conversely, the E sequence was kept unchanged.

3) Two cysteines were inserted at both the amino and carboxy terminals: such residues allow hopefully the formation of a intra-chain disulfide bond, which facilitates the association between coils and stabilizes the scaffold after the modification of the central heptads.
Results and discussion

The E and K coils, connected by a loop, are able to fold into helical domains and spontaneously associate. We have assessed this behavior by both a theoretical and experimental model.

A theoretical model of the coiled-coil was built first by homology modeling. The NMR structure of the C-Jun homodimer, one of the few available in the Protein Data Bank (Pdb id: 1JUN) was chosen as the starting topology (Junius et al. 1996). The C-Jun homodimer is a six heptad parallel homodimeric coiled-coil consisting of a non repeating sequence (table 4.1-entry 1):

<table>
<thead>
<tr>
<th>C-Jun</th>
<th>G B A R E K V K T V K D Q E T A S T A N M E E B O T T A Q L K V M N Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Mix1</td>
<td>G B A R E K V K T V K D Q E T A S T A N M E E B O T T A Q L K V M N Y</td>
</tr>
<tr>
<td>3 Mix2</td>
<td>G B A R E K V K T V K D Q E T A S T A N M E E B O T T A Q L K V M N Y</td>
</tr>
</tbody>
</table>

Table 4.1 - Alignment of the sequences of C-jun with the intermediate and the final E/K model structures.

The initial C-Jun homodimeric structure was relaxed first into a periodic box of TIP water molecules by a 100 ns molecular dynamics run in the NTV ensemble, carried out with the AMBER force field as implemented in Sybyl 8.2 in iTs cornell version (Cornell et al. 1995). The relaxed structure was then gradually mutated to that of the E-K parallel heterodimer, by first replacing all the residues at position a and d with valine and leucine respectively, in both the
Results and discussion

chains (table 4.1, entry 2, mut1). The charged residues at positions g, e, f, where then mutated into E, K, E in one chain and into K, E, K in the other, as in the E-K heterodimer (table 4.1, entry 3, mut2). Finally, also positions b and c were mutated to S and A in both the chains, giving by this way a model of the E-K heterodimer. Full MD relaxation was allowed at each stage, to verify the stability of the mutated models. The final model adopts the typical superhelical conformation of a coiled-coil.

![Image](image)

Fig. 4.2 - Relaxed and optimized model of scE-K in a periodic box of TIP water.

Further mutation to the E-Kinv. model required a more complex approach, and the K chains was temporarily broken at each heptad end, and each heptad was separately inverted and relaxed. Finally, the heptads were joined again and the whole model was submitted to 300 ns MD run. The resulting antiparallel dimer (figure 4.2) resulted stable and no unfolding or separation of the two chains were detected during this very long molecular dynamics simulation. The number of hydrophobic contacts at the coil interface is conserved, and also the buried hydrophobic surface is similar in the E-K and in the antiparallel scE-Kinv. models (table 4.2).

<table>
<thead>
<tr>
<th></th>
<th>E-K</th>
<th>scE-Kinv.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buried hydrophobic surface (Å²)</td>
<td>1290</td>
<td>1253</td>
</tr>
<tr>
<td>Hydrophobic contacts</td>
<td>30</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 4.2 - Hydrophobic interactions in E-K and scE-Kinv.
Results and discussion

Starting from this favourable theoretical prediction, we have synthesized the $E_3Kinv_3$ peptide on a 10 mg scale, as a reduced model of the full scE-Kinv scaffold (figure 4.3).

![Diagram of the E$_3$Kinv$_3$ peptide structure](attachment:image)

Due to its synthetic cost, $E_3Kinv_3$ is shorter than the full scaffold and contains just three heptads of the zipper interface. The absence of final cysteine can influence the real peptide behaviour; however, it can mimic with a certain reliability the behaviour of EKinv and a preliminary study of its folding has been carried out.

Circular dichroism (CD) is one of the most used technique to investigate secondary structure of proteins or peptides, as CD spectra can be readily used to estimate the fraction of $\alpha$-helix, $\beta$-sheet or random-coil conformation.

Protein or peptide secondary structure can be determined by CD spectroscopy in the far-UV spectral region (190-250 nm), where the chromophore is the peptide bond and the CD signal arises when the bond interest with the chiral enviroment of the folded peptide. The lowest energy transition in the peptide chromophore is an $n\rightarrow\pi^*$ transition observed at 210-220 nm, and the highest energy process is the $\pi\rightarrow\pi^*$ transition that gives rise to a band at 120 nm. The ellipticity associated to such transition depends on the local conformation of the peptide chain and $\alpha$-helices, $\beta$-sheets and random-
coils each gives rise to a characteristic shape and magnitude CD spectrum (Kelly et al. 2005) (figure 4.4).

![Image](image.png)

**Fig. 4.4 - CD spectra of various types of secondary structures.** Solid line, α-helix; long dashed line, antiparallel β-sheet; dotted line, type I β-turn; cross dashed line, extended 31-helix or poly (Pro) II helix; short-dashed line, irregular structure.

The circular dichroism spectrum of a freshly prepared 5 μM E₃Kinv₃ peptide solution in water is shown in figure 4.5:

![Image](image.png)

**Fig. 4.5 - CD spectrum of a freshly prepared solution of E₃Kinv₃.**

The previous spectrum is typical for a random-coil conformation due to the presence of a minimum 202 nm, and to the lack of Cotton effect inversion.
along the whole spectral range (Fasman 1996). A fraction of folded peptide can be responsible for the shoulder around 220 nm spectra. A new spectrum was registered after four days (figure 4.6):

![CD spectrum of a solution freshly of E3Kinv3 after four days at room temperature.]

Major changes are observed after this time, and now the spectrum is typical for a folded structure. The α-helix content of this structure has been calculated according to the empirical equation 1 (Ahmad et al. 2005):

\[ f_\alpha = \frac{[\theta]_{222}}{[\theta]_R} - \frac{[\theta]_{222}}{[\theta]_\alpha} \] (equation 1)

which is commonly used for short peptides, where \( f_\alpha \) is the fractional α-helix content, \([\theta]_R\) is the ellipticity at 222 nm for a disordered chain and \([\theta]_\alpha\) is the ellipticity at 222 nm for a 100% α-helix n long peptide. \([\theta]_\alpha\) is obtained from this equation 2 (Fasman 1996):

\[ [\theta]_\alpha = [\theta]_\infty (1-k/n) \] (equation 2)

where \([\theta]_\infty\) is the molar ellipticity per residue for a chain of infinite length with maximal α-helix content and k is an empirically determined factor interpreted as the average number of peptide bonds in the chain that cannot participate in α-helix formation. There is general agreement for an \([\theta]_\infty\) value of 40000 deg dmol\(^{-1}\)cm\(^2\) but the k value remains in dispute. Estimated value of k have ranged from 2.5 to 4.3 (Chen et al. 1974), and even up to 6.3 (Ozdemir et al. 2002); for our work we have used the commonly accepted value of 4
(Fasman 1996); for the $[\theta]_R$ value we have used the estimate 0 deg dmol$^{-1}$cm$^2$. A 40% helix content was obtained from equation 1 for our case. This value is not in agreement with the expected α-helical characteristic of $E_3K_{inv}$ if it were correctly folded. Conversely, the $[\theta]_{222}/[\theta]_{208}$ ratio is consistent with the presence of a single α-helix rather than with a coiled-coil, or, more likely, the coiled-coil is too short to induce the superhelical deformation to the α-helix which are responsible for the enhancement in 208 nm ellipticity that is observed in longer coils (Litowski and Hodges 2002). The folding process of $E_3K_{inv}$ is apparently extremely slow. This point is perhaps remarkable, and would require a more detailed study. Unfortunately, due the small amount of available peptide, we were able to obtain a single very preliminary kinetic run by monitoring at different times the molar ellipticity change at 222 nm (figure 4.7). The resulting sigmoidal behaviour is consistent with a cooperative process with a very long hysteresis followed by an autocatalytic phase.

![Fig. 4.7 - Kinetic folding of $E_3K_{inv}$ peptide](image)

Although very unusual, this behaviour is not unprecedented. The four-helix bundle Rop, also known as Rom, is a homodimer of two helix-loop-helix monomers, each of which is 63 amino acids long and its folding time after guanidinium denaturation, is in the order of days (Munson et al. 1994). The mystery of Rop folding was addressed by many authors, and several theoretical explanations have been proposed. In a recent, general study on the molecular basis of coiled-coil formations, Steinmetz, Kammerer and colleagues have suggested that, in order to allow a correct folding of a coiled-coil, a trigger
sequence must be present (Steinmetz et al. 2007). However, the E/K sequence in the native heterodimer leads to very usual and fast folding kinetics (Salwiczek and Koksch 2009) and the unusual folding of E₃Kinv₃ seems rather connected with the presence of the loop. The presence of a trapping minumun in the protein energy landscape has been proposed for the Rop dimer, and viewed in a series of Rop mutants, as a sin association of the two monomer instead of the correct anti topology, or in other cases as a molten globule (Levy et al. 2005). In an attempt to generate crystals for an X-ray analysis of E₃Kinv₃ structure, we have also submitted the peptide to a GPC analysis, that revealed the existence of three fractions, at molecular weight corresponding to a monomer, a trimer and an octamer respectively. The octamer appears as the predominant fraction. Again, the CD spectrum of a freshly isolated octamer fraction, reveals a largely unfolded population and again the CD slowly changes and after 48 hours the folded structure is observed, thus supporting the possibility that the octameric form could represent the free energy trap that makes slow the folding of E₃Kinv₃.

Single chain E-Kinv was than used to construct a library of receptor by random mutagenesis of the central heptads of both peptides. This randomization was introduced in order to obtain a huge collection of binders with pockets that could be able to bind different ligands.

The randomized library of the scE-Kinv was subsequently cloned into an appropriate vector to perform a selection based on the phage display technique. The peptides are presented on the surface of a filamentous bacteriophage, and the selection is carried out against specific molecules. The main targets considered in this study are caffeine, theophylline and aldosterone.

### 4.1.2 Library creation by random mutagenesis

In this approach, the whole central heptad of both the E and Kinv domains (figure 4.8) were randomly mutagenized (a-b-c-d-e-f-g positions), substituting each residue with one of the 20 natural available amino acids by means of mutated primers. The other four heptads of Kinv and E domains kept the wild type sequence in order to avoid an excessive variability of the whole K and E helices.
Results and discussion

In order to insert random mutations, the use of degenerated random primers was necessary. The modification was introduced by means of degenerated primers. A degenerated triplet is represented as NNN, where N indicates any of the four nucleotides. However, with the NNN triplet three stop codons can be casually inserted (TAA, TAG, TGA), giving a premature arrest of the coil sequence. This drawback was overcome choosing codon NNK as random triplet, where K stands for the nucleotides G or T: NNK allows to encode any of the 20 amino acids, avoiding the stop codons, except for TAG (amber stop). This problem was solved using the E. coli DH5α, a suppressor bacteria strain that can translate TAG in a glutamine residue.

As explained in figure 4.9, the target regions for randomization were the central heptads of both E and Kinv random coils. To this point, it is important the design of correct primers to generate the library by PCR amplification of the fragments. Starting from the sequence of Ewt and Kinv-wt peptides, three subsequent cycles of PCR were performed to insert the central heptad randomized.
During the first cycle of PCR, a mutagenized region was introduced in the Kinv and E fragments. During the second PCR cycle each fragment was amplified and an homology region was inserted between two peptides. The library was finally created by a third amplification reaction, by assembling Kinv randomized and E randomized coils in the single chain format (figure 4.10).

To exploit the advantages deriving from the use of phage display, a single chain E-Kinv random phage library was created using a phagemidic vector. The randomization of the central heptad of each coil allowed to create a collection of single chains able to bind different ligands. In fact, each single chain has a different binding pocket with different ligand affinities (figure 4.11).
Results and discussion

Considering the extension and the available residues for randomization (7 amino acids for each chain), the number of possible scaffolds was $1.63 \times 10^{18}$. However, this variability is too much high to be obtained with this technique, and the actual library will represent just a subset of the theoretical one.

The pool of PCR fragments and the pDAN vector were both digested with \textit{BssHII} and \textit{NheI} restriction enzymes. After purification, fragments and vector were ligated, and the ligation product was used to electroporate \textit{E. coli} DH5\(\alpha\) competent cells. Bacteria were subsequently plated onto six ampicillin plates. After 15 hours of incubation at 37°C, cells were harvested and stocked with 20% glycerol at -80°C.

The size of single chain E-Kinv random phage library (named scE-Kinv random) was $2 \times 10^8$ independent clones. This result did not assure the presence of all possible single chains, the size was nevertheless considered satisfying for this technique.

4.1.3 scE-Kinv random library characterization

The presence of correct fragments in bacteria, cloned into pDAN vector, was assessed performing a PCR with generic primers (VLPT1/VHPT1) on ten randomly picked bacterial clones (figure 4.12). All ten clones contained a full length fragment of the expected size, about 300 bp: the PCR size was due to fragment external position of the annealing primers employed for the screening. The results here reported confirmed the integration of scE-Kinv random fragments into the pDAN vector, and the presence into each bacteria of one plasmid.

![Electrophoresis of PCR-amplified scE-Kinv random DNA of 10 randomly picked clones. Each clone, expressing a different scE-Kinv fragment, was amplified with VLPT1/VHPT1 primers. 100bp DNA marker, ranging from 100 to 1000 bp, is reported in the first line.](image)

Fig. 4.12 - Electrophoresis of PCR-amplified scE-Kinv random DNA of 10 randomly picked clones. Each clone, expressing a different scE-Kinv fragment, was amplified with VLPT1/VHPT1 primers. 100bp DNA marker, ranging from 100 to 1000 bp, is reported in the first line.
To check the real diversity of the clones of the library, the same 20 clones were sequenced, verifying the randomization of selected residues (table 4.3).

<table>
<thead>
<tr>
<th>Clone</th>
<th>Krandom heptad</th>
<th>Erandom heptad</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WCLNQHR</td>
<td>AKHQTGQ</td>
</tr>
<tr>
<td>2</td>
<td>LHRSQGH</td>
<td>VCFVIET</td>
</tr>
<tr>
<td>3</td>
<td>QNIHRSK</td>
<td>LRGVESW</td>
</tr>
<tr>
<td>4</td>
<td>ELHNTMC</td>
<td>IVAAYRE</td>
</tr>
<tr>
<td>5</td>
<td>QDVSKHA</td>
<td>QAVIRNA</td>
</tr>
<tr>
<td>6</td>
<td>DMVTRLN</td>
<td>HQDVMGI</td>
</tr>
<tr>
<td>7</td>
<td>GTAALLR</td>
<td>CVVQMTV</td>
</tr>
<tr>
<td>8</td>
<td>SMGKLGR</td>
<td>EFTPPGS</td>
</tr>
<tr>
<td>9</td>
<td>RMVVOQI</td>
<td>KPTYHNP</td>
</tr>
<tr>
<td>10</td>
<td>RFPPFLF</td>
<td>RDDAFA1</td>
</tr>
<tr>
<td>11</td>
<td>TGGVLLF</td>
<td>SIVYEML</td>
</tr>
<tr>
<td>12</td>
<td>TSSQLRM</td>
<td>QYSFCLA</td>
</tr>
<tr>
<td>13</td>
<td>QQDSQRN</td>
<td>KNQWKR</td>
</tr>
<tr>
<td>14</td>
<td>SERRWVG</td>
<td>AGHDLGR</td>
</tr>
<tr>
<td>15</td>
<td>RRAGGIR</td>
<td>NADQYNA</td>
</tr>
<tr>
<td>16</td>
<td>VHGMSQG</td>
<td>MIVCRF</td>
</tr>
<tr>
<td>17</td>
<td>QGGWNER</td>
<td>FPIEGFY</td>
</tr>
<tr>
<td>18</td>
<td>LVMRHSF</td>
<td>ALARGG</td>
</tr>
<tr>
<td>19</td>
<td>PILCLQH</td>
<td>CNDNIVV</td>
</tr>
<tr>
<td>20</td>
<td>YVDHPQS</td>
<td>WTWTWQR</td>
</tr>
</tbody>
</table>

Table 4.3 – Summary of the randomized-central heptad sequencing of twenty scE-Kinv random clones. The randomization was successful and all 20 clones show diversity in the central random heptad.

4.1.4 Selection by phage display system versus small molecules

Phage Display technology

Phage display is the most well-studied high-throughput screening method to study protein interactions, based on displaying peptides or proteins on the surface of a bacteriophage. Filamentous bacteriophages (F) are a group of related virus that infect only gram-negative bacteria. Thanks to their structure, their replication and packing processes (not limited by DNA size), these phages are useful tools to express recombinant proteins. With molecular methods, the DNA sequences of interest can be embedded in the phage genome with the
correct framework of the genes coding for P3 or P8. The exogenous peptide are displayed on the phage surface in fusion with capsid proteins (Azzazy and Highsmith 2002).

The principles of phage display are that the coat proteins of fF phages might tolerate being fused to foreign polypeptides without losing their function, and that the fused foreign polypeptide can be displayed on the surface of the filamentous phage. The general concept is that a phage encoding a specific fusion protein on its surface, could be isolated for its binding property to a given protein from a collection of billions of phages with different exposed proteins.

In a typical phage display selection, the phage displaying the protein of interest is retained on a surface, coated with an antigen or antibody recognizing, while non-adherent phage are washed away. Bound phages can be removed from the surface, re-infected into bacteria and re-grow for further enrichment and eventually for binding analysis (figure 4.14). The possibility to perform successive rounds of selection permits the isolation of proteins present in very low number in a population of billions different phages. Usually, three or five rounds of panning are sufficient to enrich for binding peptide sequence.

**scE-Kinv random phage library selection approach**

The obtained scE-Kinv random library was then used for subsequent rounds of selection by phage display technology against molecules of our interest to select the best binders versus small molecules.

The ligands of interest must be first immobilized onto a solid phase to select and isolate a single chain-expressing clone. The immobilization was realized using the biotin-streptavidin method.

The molecules used for the selection carry a biotynilated linker (figure 4.13), (phd thesis of Dr. G. Fontanive).
Results and discussion

Fig. 4.13 – Caffeine, theophylline and aldosterone with a biotinylated linker used for phage-display selection.

This method needs a preliminary step for streptavidin immobilization onto a surface, after which, the biotinylated-ligand tightly binds to immobilized streptavidin. Several cycles of selection were performed with an additional pre-incubation step. To avoid the selection of unspecific clones, the library was incubated first over the non-coated surface and binding buffer without the ligand of interest thus removing unspecific clones from the phage population (figure 4.14).

Fig. 4.14 - a) Phage particle displaying the single chain E-Kinv random in fusion with the gene 3 protein (g3p) and b) schematic representation of phage-display selection.
This selection was carried out by use of paramagnetic beads that allow to isolate positive clones during preliminary selections, demonstrating a manageable experimental protocol (see paragraph 3.6.16). The selection method provides pre-coating of magnetic beads with streptavidin, and incubation with biotinylated-caffeine (and then with biotinylated-aldosterone in another selection cycle). The beads were then incubated with the phage mixture and removed from the solution with a magnetic field (figure 4.15).

Three rounds of selections were made in order to enrich the selection for the clones reactive to the ligands. Stringency of washing steps was progressively increased between different cycles (see paragraph 3.6.16).

4.1.5 Selection and validation vs caffeine

After each round of selection, 96 randomly picked *E. coli* colonies were tested in phage ELISA on biotinylated-caffeine in 96 well plates. In the third selection cycle, we identified only six reactive clones of 96 from the library selected against caffeine. Clones that gave a positive signal to our ligands were recovered and sequenced (table 4.4).
Results and discussion

4.4 - Amino acid sequences of the randomized central heptad of six selected clones. Only bE1 and bE8 clones shown a different heptad, thus indicating selection specificity.

<table>
<thead>
<tr>
<th></th>
<th>Kinv chain</th>
<th>E chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>bA2</td>
<td>EKLASVK</td>
<td>FLRRRIR</td>
</tr>
<tr>
<td>bE1</td>
<td>EKLASVK</td>
<td>FLRRRIR</td>
</tr>
<tr>
<td>bF1</td>
<td>EKLASVK</td>
<td>FLRRRIR</td>
</tr>
<tr>
<td>bF8</td>
<td>EKLASVK</td>
<td>FLRRRIR</td>
</tr>
<tr>
<td>bE7</td>
<td>EKLASVK</td>
<td>QSRRLKR</td>
</tr>
<tr>
<td>bE8</td>
<td>EKLASVK</td>
<td>QSRRLKR</td>
</tr>
</tbody>
</table>

The analysis revealed that among the six selected clones, only two are different: bE1 and bE8. In fact four of six selected clones (bF1, bA2, bE7 and bF8) have shown the same sequence, thus indicating the enrichment of libraries during phage display selections and successful amplification of two specific clones. Such sequences were not observed among those clones obtained from the random sequence of the library (table 4.3).

After sequences analysis of the phagemid cDNA inserts, clones were further analyzed by phage ELISA on caffeine and streptavidin as negative control. Given the structural similarity between theophylline and caffeine we have tested the clones also against theophylline which was recognized with full crossreactivity (figure 4.16).

![ELISA assay of the previously isolated clones bE1 and bE8 versus caffeine, theophylline and streptavidin. The negative control was the single chain E-K with the wild type coil sequence. Secondary antibody: anti-M13 HRP conjugated 1:2000.](image-url)
4.1.6 Measurement of the dissociation constant of the peptide/caffeine complexes

After having selected the bE1 and bE8 binders, we have measured the dissociation constant of their caffeine complexes, using an ELISA methodology described by Loomans and colleagues (Loomans et al. 1995). The dissociation constant is measured by coating a series of ELISA plates with different concentrations of caffeine (namely 2.5, 5, 10 and 20 μg/mL) and then by incubating different concentrations of phages expressing peptides bE1 and bE8 over any concentration of immobilized caffeine. A peptide concentration/response plot is obtained at any caffeine concentration (figure 4.17):

Fig. 4.17 - a) The series of curves given by clone bE1 b) the response obtained with clone bE8.

The experimental points are then submitted to a nonlinear regression analysis that yield hyperbolic fittings and OD50 calculated values. A series of Kd values is then obtained by submitting any possible couple of peptide concentrations corresponding to the calculated OD50 values, to equation 1:

\[
K_d = \frac{(n-1)/2}{[A][B]} \quad (\text{equation 1})
\]

where [B] and [A] are the peptide concentrations giving the OD50 signal at two different caffeine concentrations, and n is their ratio. The average of such calculated Kd values gives finally a Kd of 0.70±0.12 nM and of 6.90±0.17 nM for
Results and discussion

bE8 and bE1 caffeine complexes, respectively. A remarkable affinity towards immobilized caffeine is thus measured by the ELISA experiments, and the nanomolar dissociation constants are comparable with these usually observed in antibody-hapten complexes.

With this series of phage-ELISA experiments we have identified two new potential peptides receptors for caffeine and theophylline molecules.

4.1.7 Characterization of bE1 and bE8 sequence

By PCR amplification and sequencing of the DNA strands contained into the phage expressing bE1 and bE8, their aminoacid sequence was obtained (table 4.4). The sequence of bE1 and bE8 at this level are nevertheless similar and very acidic: four arginine residues are found in bE1, three arginine and one lysine residue are found in bE8. The presence of such a high number of arginines is rather surprising. However, an analysis of xanthine binding motives inside enzyme binding sites reveals that strong interactions between arginine and xanthine carbonyls are often found. This is the case of xanthine oxidase, where Arg 880 plays a key role in orientating xanthine and stabilizing its complex with the catalytic molibdenum ion of the enzyme (Pauff et al. 2008). Also in xanthine dehydrogenase from Rhodobacter Capsulata, arginine 310 is essential, and its mutation to methionine leads to a 103 reduction in both the thermodynamic and kinetic parameters for a series of purine substrate (Pauff et al. 2007); arginine is essential also in caffeine inhibition of ionotropic glycine receptors (Duan et al. 2009). With the sequence of bE1 and bE8 in hand, we have tried to build up also a model of the two caffeine complexes. A Predict-Protein (Rost et al. 2004) sequence analysis carried out on the two peptides yields a random-coil secondary structure prediction at the mutated E sequences in both the peptides, while α-helix conformation is confirmed as highly probable along the remaining sequence. This allow to take as a starting hypotesis the possible creation of a small binding pocket onto a largely conserved coiled-coil skeleton. As the starting geometry for the coiled-coil scaffold, we have chosen the homology model described in chapter 4 paragraph 4.1.1. The linker chain, the terminal cysteines and their disulfide bridge were added first, and the topology was submitted again to a relaxation MD cycle as we have previously
described. The sequence were then mutated to obtain the first models of bE1 and bE8. The models were submitted again to a final-multistage relaxation session (Berti et al. 2005), and then docking of caffeine was performed using a GRID-flexible docking protocol (Goodford 1985). We have also docked the linker-modified caffeine and theophylline into the peptide binding site, in order to evaluate also the interactions towards the linker, and its cooperation in recognizing immobilized xanthines. The more stable docking solution obtained after clustering of the docking poses were then optimized again in a periodic box of water molecules to obtain the final models. The final geometry of the bE8-linker-caffeine complex is reported as an example in figure 4.18:

A rather deep caffeine-binding site is observed. Caffeine is buried to a deeper extent at its imidazole system and at methyl in position 7. Carbonyl 2 and partly methyl group 1 and 3 are conversely exposed to the solvent. The site is contigous to a hydrophobic cleft that corresponds to the zipper interface hosting the hydrocarbon chain of the linker. A detailed view of the caffeine interactions is given in figure 4.19a for bE1 and in figure 4.19b for bE8 complexes.
Results and discussion

Fig. 4.19 - a) Detail of bE1 peptide-caffeine interaction and b) bE8 peptide interaction with caffeine.

In the bE1 complex, hydrophobic contacts are established between caffeine and the hydrocarbon chains of Lys 61, Val 65, Lys 68, Leu 13, Leu 17, Arg 20. Phenylalanine 16 is at π-methyl distance to caffeine methyl 1 while Arg 20 is at hydrogen bond distance to carbonyl 2. In the bE8 complex the side chains of Val 61, Leu 62, Leu 20, Lys 61, Lys 21, Ser 17 are at hydrophobic contact distance with caffeine, while the distances between caffeine carbonyl 2, the hydroxy group of Ser 17 and the cationic head of Lys 21 are consistent with an hydrogen bond network.

We have calculated complexation energy values, obtained by adding a solvation energy term (obtained from a discrete water model) to the conformational, non bonded and electrostatic energy terms arising from the AMBER force field (equation 2)

$$\Delta E_{\text{compl}} = \Delta E_{\text{conf}} + \Delta E_{\text{elast}} + \Delta E_{\text{nb}} + \Delta \Delta E_{\text{sol}} \ (\text{equation 2})$$

The complexation energy are reported in table 4.5, relatively to the calculated complexation energy for the bE1/caffeine complex.

<table>
<thead>
<tr>
<th></th>
<th>bE1 $\Delta E_{\text{compl}} \quad \text{Kcal mol}^{-1}$</th>
<th>bE8 $\Delta E_{\text{compl}} \quad \text{Kcal mol}^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caffeine-linker</td>
<td>-44.4</td>
<td>-62.3</td>
</tr>
<tr>
<td>Caffeine</td>
<td>0</td>
<td>8.5</td>
</tr>
<tr>
<td>Theophylline</td>
<td>3.9</td>
<td>7.1</td>
</tr>
</tbody>
</table>

Table 4.5 - Calculated complexation energy for the bE1 and bE8/caffeine complex.
In the lack of a calibration set of experimental data, such complexation energies can not be used to predict binding constants; however, the model predicts a better interaction for the bE1 complexes, with a very large gain in binding energy for the linker-modified target molecules.

### 4.1.8 Selection and validation vs aldosterone

The same scE-Kinv random library was also selected against another molecule, aldosterone, and already from the second round of phage-display selection a positive clone (G7) has been isolated. This clone reacts only with aldosterone and not with caffeine, streptavidin or linker alone. Its sequence is reported in table 4.6.

<table>
<thead>
<tr>
<th>Clone</th>
<th>K_{random inv}</th>
<th>E_{random}</th>
</tr>
</thead>
<tbody>
<tr>
<td>G7</td>
<td>RCFLWGW</td>
<td>FFQYCWV</td>
</tr>
</tbody>
</table>

Table 4.6 - Amino acid sequences of randomized central heptads of G7 clone by sequence analysis.

G7 clone binds specifically aldosterone (figure 4.20) and has been reselected six times during a third round of selection.

![Graph showing reactivity of clones G7 and EwtKwt against aldosterone, caffeine, linker, and streptavidin](image)

**Fig. 4.20** - Reactivity of the isolated clones G7 to aldosterone, caffeine, linker and streptavidin by ELISA. The single chain E-K with wild type coil sequence was considered as negative control. Secondary antibody: anti-M13 HRP conjugated 1:2000.

After identification and validation of clones bE1, bE8 against caffeine and G7 against aldosterone, we wondered whether they might therefore be used as scaffolds instead of antibodies.
Results and discussion

4.1.9 Single chain \( \text{E}_\text{G7} \text{K}_\text{G7} \) Fc fusion protein

G7 clone was thus expressed in the eukaryotic expression vector linked to CH2-CH3 chain of human antibodies. Minibody is a synthetic antibody construct having the CH2-CH3 regions fused to the VL-VH in the form of single chain that dimerizes through the formation of disulfide bridges. For this reason VL-VH have been replaced with the single chain \( \text{E}_\text{G7} \text{-K}_\text{G7} \) and we have observed that the \( \text{E}_\text{G7} \text{-K}_\text{G7} \) minibody retains the binding activity of the phage-dyspalayed peptide (figure 4.21). To produce large quantities of this protein, we have setup a cell culture using a system for cell cultivation in multicameras, where nutrient exchange takes place. After several weeks the supernatants are collected and tested to verify the quantity and functionality.

![Schematic structure of minibody with VL-VH regions substituted by single chain \( \text{E}_\text{G7} \text{-K}_\text{G7} \).](image)

\( \text{E}_\text{G7} \text{K}_\text{G7} \text{-pHygro} \) plasmids were used for the transfection of CHO-S cells. These cells are derived from the ovary of Chinese hamster and they are the most commonly used mammalian hosts for industrial production of recombinant therapeutic proteins.

Three days after transfection, protein production was checked by ELISA, using cells supernatants and the \( \alpha \)-SV5 tag primary antibody and was successfully transfected. The \( \text{E}_\text{G7} \text{K}_\text{G7} \) transfected cells were expanded from a 24 wells plate up to a 75 cm\(^2\) flask in selection (hygromycin) to allow the survival of the transfected cells. Supernatants were collected every 2-3 days and tested by ELISA assay, using the \( \alpha \)-SV5 tag as primary antibodies (figure 4.22).

To producing monoclonal G7, was expanded up to a 150 cm\(^2\) flask. In each expansion step the obtained supernatants were collected, concentrated by
Results and discussion

centrifugal filter units (millipore) and tested with a specificity ELISA assay, performed using the α-SV5 tag primary antibody (figure 4.23). This test confirmed the maintained protein production in all the expansion phases, and the binding specificity to aldosterone.

Fig. 4.22 - a) ELISA test of single chain E$_{G7}$-K$_{G7}$ minibody supernatant collected from a polyclonal and b) monoclonal populations. Aldosterone was used for the coating and the ELISA was performed adding supernatants and antibody anti-SV5 tag. G7 minibody has been succesfully transfected and reacts with aldosterone.

Both the polyclonal and monoclonal supernatant produced by cells containing our selected peptides, in particular E$_{G7}$K$_{G7}$ recognizes very well aldosterone while did not binds caffeine or streptavidin (as linker). A further ELISA test has been made and the results validate our selection system.
Discussion

Coiled-coils have been used in numerous applications, including affinity purification, direct assembly of extracellular receptor domains, creation of miniaturized antibodies, microarray and peptide libraries (Lupas 1996; McFarlane et al. 2009). The small size and defined structure of the coiled-coil make it an extremely used motif for a growing number of de novo design, protein engineering and biotechnological application.

Engineered antibodies are currently the most widespread detecting tool for biotechnological, medical and pharmaceutical applications, because of the deep knowledge of their structure, binding capabilities and specificity of action. However, the emerging field of protein engineering allowed the production of a wide range of scaffolds different than antibodies, with various origins and characteristics, for protein library construction and for grafting experiments of binding sites (Binz et al. 2005; Hosse et al. 2006).

The proprieties of these new scaffolds are quite similar to the ones characterizing antibodies, i.e. specificity and affinity, but some features are improved, like a reduced size, a binding pocket restricted to few residues and a good stability even without disulfide bonds. All these characteristics are suitable to create a customizable nanotechnological device with enhanced spatial resolution, that can be further improved to construct biosensors-based detecting system for any application.

However the possibility to modify the amino acid sequence allows the improvement of binding features. Considering their characteristics, coiled-coils have been used as affinity purification tags, library presentation scaffolds and as bases for the creation of miniantibodies (Chao et al. 1998; Arndt et al. 2001)

The aim of this work was the development and the improvement of novel proteins scaffold for biosensing applications, and to this purpose two methods have been adopted: one based on E/K coiled coil synthetic domain and another based on natural binders Human Serum Albumin. In this chapter (1) we worked with E/K coiled-coil, a de novo designed heterodimeric parallel coiled-coil domain, for biosensor-based affinity and chromatographic applications (Chao et al. 1996). The dimerization propensity and tolerance to sequence modifications of a coiled-coil system allowed us the construction of a new scaffold structure able to bind small molecules, with several analytical implications. This new
structure based on E and K coils has been engineered to generate a single molecule containing both peptides, the single chain E-K (scE-K). As already said, by structure analysis of several natural coiled-coils, some requirements were established to permit the correct folding of the new protein: based on this features we designed a random library E/K in a single chain form starting from E/K wild type. The stability of E/K coiled-coil domain was previously tested with the protein complementation assay and then with circular dichroism. Additional modifications and experiments were performed with the final aim of improving the interaction between the coils, in order to create E/K pairs with increased binding affinities or recognition of new specific ligands.

During this project, in order to create a specific binding site for our target molecules, a further modification at the E/K wilde type scaffold was introduced to create a scE-K random library. The approach was based on the randomization of the central heptad to allow the recognition and the binding of small molecules such as caffeine and aldosterone. The mutation in the only central heptad of both coils allowed the creation of a binding pocket. After design of topology of scE-K random peptide receptor, we have started the development of a library of receptors by phage-display technique. By using this biomolecular technique a large collection of $2 \times 10^8$ random peptides was obtained and selected against immobilized caffeine-biotin and aldosterone-biotin by ELISA-tests in order to identify and characterize the best binders.

The selection of phages was performed using streptavidin-coated magnetic beads (Shlyapnikov et al. 2010); beads were incubated with biotin analyte in the first step and then with the phages. The phages blocked on the beads surface are able to recognize and bind the analyte and can be separated from the other not specific phage through the use of a magnetic device.

Thanks to this technique, the single chain library was expressed in fusion with a surface capsid protein, and after several selection cycles we isolated single chain mutated peptides carrying a binding pocket highly specific for caffeine. We chose this molecule because it is small, easy to handle and of particular interest for industrial processes, or in food and clinical chemistry (Nawrot et al. 2003).

In addition, the most important advantage that we have exploited is that with the aid of phage display one can directly select protein scaffolds able to
bind the analyte of interest bypassing time-consuming 3D modeling of interactions between coils and ligand. Moreover, this technique could be potentially applied to any other protein scaffolds based library like that based on natural receptor (paragraph 4.2). This feature enables the rapid, economic and “on demand” creation of biosensors, maintaining their properties unaltered, without needing to purify the molecules to be immobilized, as the ones isolated and characterized in the present work.
Chapter 2

4.2 A Binder and catalytic peptide from a natural host: Human Serum Albumin

The results reported in this chapter have been submitted to Nature Chemical Biology on 10 February 2012: An albumin-derived peptide scaffold capable of binding and catalysis. Available from Nature Precedings <http://hdl.handle.net/10101/npre.2012.6894.1> (2012).

4.2.1 HSA100 identification and optimization

Due to our interest in the identification of a polypeptide system acting as both binder and catalyst, we reasoned that a convenient starting point would be a “wild-type” sequence that is capable of both recognizing a broad spectrum of small molecules with micromolar affinity and catalyzing a wide range of reactions with moderate proficiency. This polypeptide should ideally present structural features that allow facile, robust detection and binding to the free ligand, avoiding the use of modified targets in the selection and decision steps. Such a peptide could be regarded as an “alternative IgM,” and the possible development of a library of such binders by random or rational mutation strategies could mimic the in vivo antibody affinity maturation process in a single step. Serum albumin is a promising candidate for having such a polypeptide motif, due to its widely reported capacity to bind a variety of small molecules with micromolar affinity (Peters 1996). Broad binding activity and chemical reactivity are associated with the binding site located in the albumin IIA subdomain (Carter and Ho 1994), also known as Sudlow site I.

Based on structural and functional analyses, we have identified a peptide corresponding to a 101-residue stretch of the human serum albumin sequence (A194 to E294) named HSA100 (figure 4.24).
Results and discussion

Fig. 4.24 - Structure of the human serum albumin HSA-100 domain. The full-length protein and HSA-100 domain are shown as white and red ribbons, respectively. The structure was obtained from the Protein Data Bank (1BKE.pdb).

This fragment corresponds to approximately half of subdomain IIa. The binding site inside this subdomain is composed of a continuous sequence of amino acids with no contribution from other far residues. This sequence includes all the residues that contact typical ligands, such as warfarin (Ghuman et al. 2005), as well as all eight cysteine residues that are involved in the formation of the four disulfide bridges (C200-C246, C254-C252, C265-C279, C278-289), which probably stabilize the polypeptide structure. Seventy-four of the amino acids were located in six $\alpha$-helical regions ($\alpha_1$, Q196-F206; $\alpha_2$, R209-R222; $\alpha_3$, A229-H247; $\alpha_4$, L250-S270; $\alpha_5$, L284-G292; $\alpha_6$, L284-G292), and the remaining residues were located in six connecting coil/loop regions. A molecular dynamic simulation of peptide stability was performed at room temperature using the Gromacs Package. The simulation demonstrated that the peptide was
conformationally stable for over 1000 ns simulation (figure 4.25). However, significant unfolding occurred upon the exclusion of C200 or C289 from the designed peptide.

Fig. 4.25 - Molecular dynamic run in the NTV ensemble of HSA100. The simulation was carried out at 300°C after equilibration steps at 100 and 200°C, with the Gromacs force field.

4.2.2 HSA100 optimization strategy

For the in vivo expression of this fragment we decided to use Escherichia coli, that was never used before, and to adopt a technique of sequence optimization to increase the HSA expression level. A preliminary evaluation of protein production was performed by fusing the HSA fragment to either a carrier maltose-binding protein (MBP) or different tag sequences (HIS6 and strep tags). Even though it was possible to demonstrate the HSA100 expression, and we were able to purify, and measure its activity, the protein yield was initially extremely poor (data not shown). Therefore, HSA100 was cloned in the vector of our interest and the HSA100 gene sequence was optimized by using a codon usage optimization strategy for expression in E. coli keeping the HSA amino acid sequence (ordered from Invitrogen GeneArt).

Among the available molecular biology strategies for codon optimization, we have focused on optimization of codon quality and of GC content (figure 4.26a,b).
In particular, regions of very high (> 80 %) or very low (< 30 %) GC content have been avoided where possible, to prolong mRNA half life, and the codon usage was adapted to the codon bias of *E. coli* genes and it has been optimized to 99% (figure 4.26c).

During the optimization process the following cis-acting sequence motifs were avoided wherever possible:

- stop codons
- 5', 3' and internal restriction sites for downstream cloning and processing
- internal TATA-boxes, chi-sites and ribosomal entry sites (to avoid the expression of not full protein)
- AT-rich or GC-rich sequence stretches
- RNA instability motifs
- repeat sequences and RNA secondary structures (to avoid the decrease of cloning efficiency)
- (cryptic) splice donor and acceptor sites in higher eucaryotes
Results and discussion

Fig. 4.26 - c) The histograms show the percentage of sequence codon which fall into a certain quality class. The quality value of the most frequently used codon for a given amino acids in the desired expression system is set to 100, the remaining codons are scaled accordingly (see also Sharp, P. M., Li, W. H., Nucleic Acids Res. 15 (3), 1987).

The optimization was successful: negative cis-acting sites which may negatively influence expression were eliminated. Codon usage was adapted to the bias of *Escherichia coli* resulting in a CODON ADAPTATION INDEX (CAI) value of 0.99 and this should allow high and stable expression rates in *Escherichia coli*. The length of the optimized HSA gene is 335 pb (figure 4.27).

---

1 CAI parameter describes how well the codons match with the codon usage preference of the target organism. Thus, a CAI of 1.0 would be perfect. However, a CAI of > 0.9 is considered as very good (i.e. allowing high expression).
As we said in the previous paragraph, to simplify the production of the HSA100 polypeptide, we decided to clone and express its coding sequence in *E. coli*. To improve the yield, we used the optimization approach described above. The 100aa domain was cloned and expressed in a procariotic expression vector pGEX 4T-1 (GE Healthcare) (figure 4.28); this vector was chosen due to its peculiar characteristics as the presence of an inducible promoter that allow high protein expression and the presence of a tag for recombinant protein (Lim et al. 1994; Rufer et al. 2005).

![Map of the original pGEX-4T1 vector.](image)

The fragment was fused with the glutathione S-transferase (GST) carrier protein (figure 4.29a), generating the GST-HSA100 recombinant protein. The right cloning was obtained as checked by PCR on randomly picked clones using two generic oligonucleotides pGex sense-pGex anti (Table 3.1).

The agarose gel picture in figure 4.29b shows that each clone contains a fragment of the expected size (950pb). Therefore, we were able to demonstrate the integration of HSA100 fragment with the correct length, inside the pGex vector in all the selected clones. The presence of a band with a larger PCR size is due to the annealing of the primers employed for the screening in an external position.
In a second time we chose two different clones and we performed their DNA sequence to verify the presence of the correct amino acids residues in the GST-HSA100 proteins. Figure 4.29c shows the experimentally obtained correct amino acid sequence of GST in fusion with 100 amino acids of HSA (in red) and FLAG tag (8 amino acids).

HSA and HSA related fragments have never been obtained by exploiting bacterial expression systems, and we have developed a protocol for optimal production. The key point in yield optimization was represented by the choice of GST as the fusion protein, as this allowed immediately a significant improvement in comparison to MBP constructs.

We have been optimized the growth conditions for bacteria operating on a set of variables including:

- the nature and amount of the molecular inductor;
- the temperature and the time of culture growth;
- the lysis conditions;
- the nature and amount of added protease inhibitors;
- the sonication time;
Results and discussion

Several conditions have been tested before reaching the optimal protocol. Initially we did not use the molecular inductor by growing the bacterial culture at 37°C O/N. The next day, the recombinant proteins were extracted after centrifugation (20’ at 4000 rpm at 4°C) in order to separate the supernatants from the pellet that contains the recombinant protein. The pellet was lysated by using different conditions varying incubation times and quantities of required reagents. Initially we have not used proteases inhibitors and the lysate was sonicated for 1-3 cycles (1-2’). By using this conditions, the yield of recombinant protein was 1-1.5mg/ml. In order to increase the protein yield, we modified some parameters adding a molecular inductor, decreasing the temperature from 37°C to 25-28°C, increasing lysozime concentration from 200 μg/ml to 400 μg/ml and sonication cycles from 1-3 to 3-10, each of 1-2’ (table 4.7).

We managed to produce 15-20 mg/l of GST-HSA100 fusion protein and obtain up to 4 mg/l of soluble protein.

<table>
<thead>
<tr>
<th>Molecular inductor IPTG</th>
<th>0.2mM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth conditions</td>
<td>25-28°C O/N</td>
</tr>
<tr>
<td>temperature</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Lysis conditions</td>
<td>20mM Tris pH 8.00, 500mM NaCl, 0.1% Triton x100</td>
</tr>
<tr>
<td>Protease inhibitors mix</td>
<td>DNAse (50-100 μg/mL of lysate)</td>
</tr>
<tr>
<td></td>
<td>PMSF (1mM)</td>
</tr>
<tr>
<td></td>
<td>EDTA (1mM)</td>
</tr>
<tr>
<td></td>
<td>Lysozime 400 μg/mL of lysate</td>
</tr>
<tr>
<td>Sonication</td>
<td>3-10 cycles, 1-2 minutes for cycle</td>
</tr>
</tbody>
</table>

Table 4.7 - Main optimized conditions for GST-HSA100 production and purification.

GST-HSA100 protein was produced and purified by affinity chromatography using a glutathione resin (GSH) that allows the binding of GST fusion protein and subsequent elution by competition with glutathione (see paragraph 3.6.23). Production and purity of the protein was assessed by SDS (runs on 12% polyacrylamide gel, stained with blue coomassie) (figure 4.30); after dialysis we were able to estimate GST-HSA100 concentration as high as 4mg/ml by quantitative electrophoresis carried out using BSA standard reference.
Results and discussion

Fig. 4.30 - a) GST-HSA100 fusion protein purification. Lane 1: molecular weight, lane 2-6 show the subsequent elution by release through GSH-agarose resin. b) Lane 1: molecular weight, lanes 2-9 represent a calibration curve with a known concentration protein BSA in 0.25-2 μg range, lanes 10-12 clearly shows a good production and purification of the recombinant protein post-dialysis.

Western blot assay was also performed to confirm the correct peptide expression and production, using specific anti TAG antibody (figure 4.31C). The results are shown in figure 4.31 which reports on B the purified and dialyzed GST-HSA100 polypeptide, and on C the same sample obtained by recognition of antibody anti-FLAG in western blot assay (see paragraph 3.6.26). Both bands show the expected 38 kDa weight and we can conclude that the soluble fraction was purified resulting in highly pure (figure 4.31B), stable, full-length protein (figure 4.31C).

Fig. 4.31 - A) Designed construct bearing the polypeptide GST-HSA100-FLAG tag. B) SDS-PAGE and C) anti-FLAG tag Western blot of purified GST-HSA100 protein.

4.2.3 Structural characterization of HSA100

LC-MS/MS and peptide sequencing analyses were performed to confirm the identity of GST-HSA100. Purified GST-HSA100 was trypsin-digested, and the resulting peptides were separated by RP-HPLC. The identities of the peptides were confirmed by comparison of theoretical and experimental m/z.
values of full-length peptides (MS) and collision-induced dissociation (CID) fragments (MS/MS). The tryptic peptide map referring to the GST-HSA100 sequence is shown in figure 4.32, and the mass spectrometry data are summarized in table 4.8.

![Tryptic map of GST-HSA100](image)

**Fig. 4.32 - Tryptic map of GST-HSA100.** Tryptic peptides obtained by GST-HSA100 digestion were analyzed by LC-MS/MS. Only peptides covering the GST-HSA100 sequence are shown. Their identities were assigned on the basis of molecular mass (red bars) or peptide sequence (blue bars).

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Molecular Mass</th>
<th>Elution (min)</th>
<th>m/z**</th>
<th>m/z 3+</th>
<th>m/z 4+</th>
<th>sequence (MS/MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>196-197</td>
<td>302.17</td>
<td>302.2</td>
<td>3.8</td>
<td>303.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200-205</td>
<td>648.32</td>
<td>648.5</td>
<td>11.4</td>
<td>649.3</td>
<td>325</td>
<td>2</td>
</tr>
<tr>
<td>206-209</td>
<td>507.24</td>
<td>507.3</td>
<td>8.5</td>
<td>508.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200-209</td>
<td>1137.56</td>
<td>1137.6</td>
<td>19.7</td>
<td>1138.5</td>
<td>569</td>
<td>8</td>
</tr>
<tr>
<td>210-212</td>
<td>364.21</td>
<td>364.2</td>
<td>5.1</td>
<td>365.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>213-218</td>
<td>672.37</td>
<td>672.4</td>
<td>19.9</td>
<td>673.4</td>
<td>337</td>
<td>2</td>
</tr>
<tr>
<td>219-222</td>
<td>502.28</td>
<td>502.3</td>
<td>4.5</td>
<td>503.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>223-225</td>
<td>390.22</td>
<td>390.2</td>
<td>10.6</td>
<td>391.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>226-233</td>
<td>879.43</td>
<td>879.4</td>
<td>19.4</td>
<td>880.4</td>
<td>440</td>
<td>7</td>
</tr>
<tr>
<td>234-240</td>
<td>788.46</td>
<td>788.5</td>
<td>20.2</td>
<td>789.5</td>
<td>395</td>
<td>3</td>
</tr>
<tr>
<td>258-262</td>
<td>516.29</td>
<td>516.3</td>
<td>5.9</td>
<td>517.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>263-274</td>
<td>1385.61</td>
<td>1385.6</td>
<td>20.2</td>
<td>1386.7</td>
<td>693</td>
<td>8</td>
</tr>
</tbody>
</table>

**Table 4.8 - GST-HSA100 peptide detected by mass specrometry.**

The circular dichroism spectra of the protein in the near- and far-UV regions shown in figure 4.33 were recorded at neutral pH at temperatures ranging from 25 to 75°C. The far-UV spectrum (figure 4.33a) was deconvoluted using the convex constraint algorithm (Perczel et al. 1992) to reveal a protein composition of 50% α-helix, 30% coil, 16% turn and <5% β-sheet. Thermal
denaturation occurred above 60°C, but at 75°C, the conserved amount of α-helix was still over 40%, with no clear trend towards complete unfolding (figure 4.33a). The presence of disulfide bonds was probed by near-UV CD spectroscopy (figure 4.33b). Many peaks were identified in the fine structure. For example, the shoulder at 267 nm and minima at 276, 284 and 293 nm resemble a set of peaks attributed to the presence of disulfide bridges in the chiral environment of domain II of HSA (Dockal et al. 2000).

4.2.4 Small molecule binding activity of HSA100

(±) Efavirenz (a) and warfarin (b) (scheme 1) were chosen as probes to test the conservation of the binding capacity of GST-HSA100. Warfarin is perhaps the most typical ligand for the subdomain IIa site, and the structure of the HSA complexes formed with both its enantiomers, which have similar affinity, have been fully described (Ghuman et al. 2005). Efavirenz, an inhibitor of HIV reverse transcriptase, is a widely used drug in AIDS therapy and has led to increased interest in the therapeutic drug monitoring of antiretroviral drug. Efavirenz binds to HSA and interacts with the subdomain II site (Bocedi et al. 2004).
4.2.4.1 ELISA assay

Biotinylated analogues of efavirenz (a) and warfarin (b) were prepared (scheme 4.1) in order to setup ELISA tests.

![Scheme 1](image1.png)

The compound a was obtained by alkylating efavirenz with ethyl chloroacetate in the presence of LDA. The linker was then extended by reaction with mono-Boc cadaverine protected, and finally biotin was added by standard amide coupling methodologies. Compound b was obtained in a similar way on the warfarin carboxymethyloxime. The two biotinylated derivates were used to immobilize the ligands on streptavidin-coated microwell plates, and ELISAs were performed after a blocking step with 2% milk in PBS (MPBS). Binding of GST-HSA100 to the immobilized ligands was then revealed with anti-GST antibody (figure 4.34).
Results and discussion

Fig. 4.34 - a) Schematic representation of the ELISA assay: the warfarin and efavirenz were firstly coated, the GST-HSA100 fusion protein was added and the reactivity was assessed. By anti-TAG antibody b) The recombinant protein was assayed for reactivity to GST tag: only GST as negative control at the same protein concentration did not binds efavirenz or warfarin.

The results confirm the reactivity between GST-HSA100 and efavirenz and warfarin respectively. The graph clearly demonstrates that the reactivity is exclusively due to the HSA stretch within HSA100-GST, because GST alone, obtained from the pGex vector, gave no signal. The HSA100-GST-efavirenz and HSA100-GST-warfarin signals show a comparable intensity under the same experimental conditions.

4.2.4.2 Fluorescence assay

The interactions of the two ligands with the peptide fragment were then monitored by fluorescence quenching experiments monitoring tryptophan emission. The emission spectrum of apo-GST-HSA100 is shown in figure 4.35
Results and discussion

(spectrum 1) as a synchronous scan at $\Delta = 60$ nm and shows an overall maximum at 281 nm (i.e., 341 nm emission). Conversely, it is well known that the four tryptophan residues have poor emission in native GST, with a maximum at 335 nm, as they are buried in a hydrophobic environment.

Quenching of GST-HSA100 fluorescence (1 $\mu$M in phosphate buffer, pH 7.4) was tested in the presence of either warfarin or efavirenz. The effect of efavirenz on the synchronous emission spectrum is shown in figure 4.35, and similar results were obtained with warfarin. Quenching occurred and was accompanied by a blueshift of the excitation maximum from 281 to 278 nm. The highest quenching was obtained at 200 $\mu$M efavirenz, which is close to its solubility limit. In the difference spectrum between unquenched and fully quenched GST-HSA100 emission, there is a shoulder corresponding to an emission maximum at 330 nm. The blueshift from 341 nm upon titration with the ligand is consistent with a change to a less polar environment surrounding the active site Trp residue, which results from solvent replacement by efavirenz (or warfarin). The residual unquenched emission may have received a significant contribution from the buried tryptophan residues in GST (figure 4.35b). The titration of fluorescence emission as a function of ligand concentration for HSA, GST-HSA100 and GST is reported in figure 4.35c. Fluorescence quenching was similar in HSA and GST-HSA100 with both ligands, whereas the effect on GST fluorescence was negligible (figure 4.35c). A Stern-Volmer analysis of the quenching data was performed to determine whether the observed quenching was due to binding or collisional phenomena. Stern-Volmer quenching constants were measured in the 4300–24500 l mol$^{-1}$ range (table 4.9). Assuming a 5-ns decay time for tryptophan fluorescence, the apparent bimolecular quenching constants derived from these data were as high as $8 \times 10^{11} - 4.9 \times 10^{12}$ l mol$^{-1}$s$^{-1}$ (table 4.9) and 2-3 orders of magnitude larger than possible for diffusion-limited collisional quenching. These data are consistent with static quenching originating from the association of the fluorophore and quenchers in a bimolecular complex (Lackowicz 2006; Zhang et al. 2010). Thus, the ligand–protein dissociation constants were evaluated by a Hill analysis of the fluorescence data (table 4.9). The $K_D$ values were similar for HSA and GST-HSA100 and were close to the literature values of 3.7–3.5 $\mu$M and 110 $\mu$M for
warfarin (Rich et al. 2001; Twine et al. 2003) and efavirenz binding to native HSA (Bocedi et al. 2004), respectively.

![Chemical structures](image)

Fig. 4.35 - GST-HSA100 small ligand binding. a) Reference albumin IIa site ligands. b) Synchronous (D=60 nm) fluorescence spectra of GST-HSA100 1 mM in 10 mM phosphate buffer, pH 7.4. 1, no quencher added; 2, 10 mM efavirenz; 3, 100 mM efavirenz; 4, 200 mM efavirenz; and 1-4, difference between spectrum 1 and 4. c) Titration of fluorescence emission of HSA, GST-HSA100 and GST by warfarin and efavirenz. ●, warfarin-HSA; ▲, warfarin-HSA100; ■, warfarin-GST; ○, efavirenz-HSA; Δ, efavirenz-HSA100; □, Efavirenz-GST

<table>
<thead>
<tr>
<th></th>
<th>Warfarin</th>
<th>Efavirenz</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HSA</td>
<td>HSA100</td>
</tr>
<tr>
<td>$K_{qH}$ Lmol$^{-1}$</td>
<td>24350</td>
<td>23870</td>
</tr>
<tr>
<td>$k_{qH}$ Lmol$^{-1}$s$^{-1}$</td>
<td>4.97x10$^{10}$</td>
<td>4.77x10$^{10}$</td>
</tr>
<tr>
<td>$K_{dH}$ molL$^{-1}$</td>
<td>2.47x10$^{4}$</td>
<td>2.70x10$^{4}$</td>
</tr>
<tr>
<td>nH</td>
<td>1.07</td>
<td>1.02</td>
</tr>
</tbody>
</table>

Table 4.9 - Binding parameters for warfarin and efavirenz. [a] Stern-Volmer quenching association constant. [b] Bimolecular quenching kinetic constant, assuming t0=5ns for the tryptophan fluorescence decay. [c] Dissociation constant for the protein-ligand complexes from Hill analysis of the quenching data. [d] number of binding sites per molecule of protein from Hill analysis of the quenching data.
4.2.5 Chemical reactivity control: diketone reduction

1,3-diols are important synthetic intermediates which have been stereoselectively synthesized through many different approaches, including biomimetic and organocatalysis methods. We observed that the reduction of diketone (figure 4.36a) and other aryl-alkyl-1,3-dicarbonyl- or 1-hydroxy-3-oxo-derivatives with sodium borohydride in aqueous acetonitrile was highly stereoselective in the presence of stoichiometric amounts of bovine or human albumin, yielding anti 1,3-diols with a diastereoisomeric excess of up to 96% (table 4.10) (Benedetti et al. 2002). The same reaction without albumin yielded syn and anti 1,3-diols in approximately a 1:1 ratio.

We performed the reduction in the presence of GST-HSA100 and verified that the level of stereoselectivity with the shortened polypeptide was as high as with the native HSA (table 4.10). Full conversion of the substrate to diols without loss of selectivity can occur, and the anti diol can be easily recovered from the aqueous medium by simple extraction after denaturing the peptide with ethanol.

![Chemical structures](image)

Fig. 4.36 - Control of reactivity with GST-HSA100. a) diastereoselective reduction of 1,3-diketones; b) albumin-catalyzed aldol addition.
4.2.6 Chemical reactivity control: aldolase activity

The aldol reaction is one of the most important synthetic tools for stereoselectively forming carbon-carbon bonds. This reaction is extremely useful when proficient catalytic and control systems are employed. Many versions of such reactions have been proposed using inorganic catalysts, organocatalysis, and bio- and biomimetic catalysis. We have recently found that human and bovine serum albumins catalyze the aldol addition of acetone to aldehyde (figure 4.36) and other aromatic aldehydes using saturation kinetics. The catalyzed process is three orders of magnitude faster than the uncatalyzed reaction, with Michaelis–Menten constants in the millimolar range (Benedetti et al. 2011).

We have recently reported the aldolase activity of the HSA100 fragment when expressed as an MBP fusion construct (Benedetti et al. 2011), and we have now characterized GST-HSA100 as an aldolase (figure 4.37, table 4.10). The albumin fragment behaved similarly to its parent protein and accelerated the reaction by over 1100-fold. The fragment exhibited multiple turnovers, allowing more than 100 catalytic cycles and full conversion of a 100-fold excess of aldehyde substrate to the aldol.

Fig. 4.37 - Aldolase activity of GST-HSA100. Michaelis–Menten plot of the initial velocities for the aldol addition of acetone to aldehyde 5 catalyzed by HSA (●) and GST-HSA100 (○).
Table 4.10 - Control of reactivity. [a] Observed pseudo first order value; [b] apparent value in 10% aqueous acetone.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Diastereoselective reduction of diketone</th>
<th>Kinetic parameters (at 37°C) for the uncatalyzed, HSA and GST-HSA100-catalyzed aldonol addition of acetone to aldehyde</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%anti</td>
<td>%syn</td>
</tr>
<tr>
<td>BSA</td>
<td>98</td>
<td>2</td>
</tr>
<tr>
<td>HSA</td>
<td>88</td>
<td>12</td>
</tr>
<tr>
<td>GST-HSA100</td>
<td>90</td>
<td>10</td>
</tr>
</tbody>
</table>

### 4.2.7 Biosensing application: SPR

The interaction of efavirenz with both HSA and GST-HSA100 was revealed also by surface plasmon resonance (SPR) by building up an albumin-based on a Biacore instrument. SPR is a tool that allows an accurate study in real-time of biomolecules interactions. It has been to characterize the binding of a wide range of samples such as proteins, nucleic acids, small molecules, lipid vesicles, viruses, bacteria and eukaryotic cells (Jason-Moller et al. 2006).

By using the Biacore system the following parameters may be obtained:

- the affinity of the receptor for its ligand by equilibrium measurement;
- the rate of the association and dissociation processes, and the affinity by kinetic measurements;
- the thermodynamic parameters of the interaction;
- the specificity of the interaction by crossreactivity measurements.

Although SPR is generated by a thin layers of metal conductors, all Biacore sensor chips are covered by thin and uniform layer of gold (figure 4.38). Gold has a wide advantage in respect to common metals, as it is inert under the physiological buffers conditions and may be easily modified with a layer of covalently bound alcane thiol self assembled monolayers, to prevent non-specific proteins adsorption.

Measurements were performed using a Biacore System at the Department of Chemical and Pharmaceutical Sciences (in collaboration with Dr. Silvia Pavan) to study the HSA and GST-HSA100 interaction with Efavirenz (figure 4.39)
Efavirenz was again modified at its carbamate nitrogen with ethyl chloroacetate and then with cadaverin in order to obtain a free amino group at the linker terminal. The sensor surface was then created by using a gold surface covered by a monolayer of thiolated carboxy methyl dextrane. Efavirenz was immobilized via amino coupling, in which the primary amino groups at linker reacts with the free carboxyl groups of the matrix, after pre-activation with 1-ethyl-3-(3-dimethylaminopropyl)-carboimmide (EDC) and N-hydroxysuccinimide (NHS).

The efavirenz modified chip was then mounted inside the biosensor flow chip and binding of HSA (figure 4.39a) or HSA100 (figure 4.39b) was observed by flowing increasing concentrations (0.5 μM to 10 μM for HSA100 and 10 μM to 400 μM for HSA) of protein.

Association and dissociation of both HSA and HSA100 was clearly observed giving rise to signals with similar RU; the equilibrium and the dissociation constant are reported in table 4.11.

<table>
<thead>
<tr>
<th></th>
<th>Efavirenz</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSA</td>
<td>7.1x10^3</td>
</tr>
<tr>
<td>HSA100</td>
<td>5.9x10^3</td>
</tr>
<tr>
<td>Kd [Lmol^-1]</td>
<td>1.9x10^-4</td>
</tr>
<tr>
<td></td>
<td>1.70x10^-4</td>
</tr>
</tbody>
</table>

Table 4.11 - Binding parameters for efavirenz. SPR steady state association equilibrium constants and dissociation constants calculated by equilibrium experiments by plotting the equilibrium RU values vs concentration of protein.
Results and discussion

4.39 - SPR sensogram for binding of a) HSA and b) GST-HSA100 to an efavirenz. Different concentrations of protein has been used to obtain dissociation constant from equilibrium data.

The values are closely similar to there obtained by fluorescence quenching experiments, of 1.89 and 2.61x10^{-4} Mol L^{-1} for HSA and HSA100 respectively. This result confirm the fluorescence data, and moreover offers a suggestion about the way of binding of efavirenz to HSA and HSA100: the carbamate region of the molecule is infact modified in the biotinylated derivative used in SPR, and this modification does not affect binding. This is consistent with shape of site II binding site, which contains two hydrophobic pockets that could interact with efavirenz or with its immobilized analogue in the same way (figure 4.40).

4.40 - Structure of efavirenz coated on surface which interacts with two hydrophobic pockets of II binding site HSA.
Results and discussion

**Discussion**

HSA100 is the shortest functional peptide that has ever been derived from the sequence of HSA. In the past decade, several larger fragments corresponding to more than one domain have been described. Rüker and colleagues prepared five fragments, including domains I, II, I-II, I-Ila and IB-II, to define the structural elements required for the formation of the warfarin binding site (Dockal et al. 1999; Dockal et al. 2000; Dockal et al. 2000).

Similar work was subsequently completed by the East group on fragments corresponding to domains I-II and II-III (Twine et al. 2003). More recently, a recombinant protein corresponding to domain I-II was designed by Fasano (Fanali et al. 2009) to study the allosteric linkage between the warfarin and heme binding sites in the protein. Three recombinant full domains of HSA have also been prepared as potential drug delivery tools (Matsushita et al. 2004). All of the described fragments, including native HSA, have been obtained as secreted proteins by expression in *Pichia pastoris*. To the best of our knowledge, fragments of human albumin have never been successfully expressed in *E. coli*. Using a combined strategy of optimizing DNA sequences and growth conditions, we were able to express and purify soluble, full-length GST-HSA100 protein. Mass spectrometry in combination with SDS-PAGE and anti-FLAG immunoblotting unambiguously demonstrated the correct production of a full-length recombinant soluble GST-HSA100 fusion construct (figure 4.31).

Our CD spectra support the correct folding of the GST-HSA100 protein. The secondary structure composition inferred from the CD data is as expected for GST-HSA100, containing the five helix and six coil/loop region of the GST-HSA100 domain and two GST domains with a βαβαβα folding topology along the first 84 residues followed by an extended α-helical domain of 132 amino acids at the C-terminus (Lim et al. 1994). The thermal stability of the GST-HSA100 protein is also comparable to those of HSA and GST (Rufer et al. 2005; Rezaei Tavirani et al. 2006), and the formation of disulfide bridges is again consistent with a correct folding process. Four disulfide bridges should be present in the GST-HSA100 region of the construct, although none of the four cysteine residues in GST is believed to be involved in disulfide bond formation, even in the predominant homodimer that forms at micromolar concentrations (Lim et al. 1994; Kaplan et al. 1997; Fabrini et al. 2009).
Binding activity is conserved in the GST-HSA100 fragment, and the measured affinities fit well with the hosting capacities of the native protein. The conservation of the binding capacity is consistent with the correct organization of GST-HSA100, thereby enabling the correct formation of the binding site. Based on the fluorescence data, the affinity for free efavirenz in solution is similar to that measured on immobilized efavirenz by the SPR sensor. Efavirenz was immobilized to the sensor surface via its heterocyclic nitrogen, leaving the interaction with the protein and the aromatic/cyclopropyl domain of the molecule exposed. This is consistent with the shape of the IIa binding site, which contains two hydrophobic subsites that host similar aromatic and hydrophobic groups. To design rapid screening methodology for binding by GST-HSA100, we performed the fluorescence-quenching titration in a 384-microwell plate format using a fluorimetric plate reader equipped with 280±15 nm excitation and 350±15 nm emission filters to detect tryptophan fluorescence. Due to the lower sensitivity of this method, microwell screening required a higher concentration of GST-HSA100 to provide a detectable fluorescence signal. However, the results are comparable with fluorimetric titrations in the spectrofluorimeter, enabling a considerably higher throughput and requiring less than a quarter of the volume. This creates the possibility of screening libraries of mutated GST-HSA100 for improved binders.

The control of diketone reduction (figure 4.36, table 4.10) was again consistent with the high degree of structural conservation in GST-HSA100. This reaction requires the recognition of the 1,3-diketone in its enol form, stabilization by ionizable residues H242 and K199 (native protein numbering) in their neutral and protonated forms, respectively, and the correct shape of the binding site around the enolized diketone to control the diastereoselectivity of the products. The initial chemoselectivity occurs in the reduction of the first (aliphatic) carbonyl, which is exposed to the solvent in the albumin-diketone complex, while the aromatic carbonyl is buried inside the albumin binding site. Subsequent diastereoselectivity occurs in the reduction of the second (aromatic) carbonyl (Berti et al. 2011). Thus, the role of albumin is to provide a combination of chemo- and stereocontrol that can be maintained in the shortened polypeptide only if the overall shape of the binding site is substantially conserved.
Experimental evidence supports the hypothesis that the HSA-catalyzed aldol reaction (figure 4.37, table 4.10) occurs through an enamine intermediate via a mechanism similar to that of type 1 aldolases. Lysine 199 reacts with acetone to produce the acetone enamine, which acts as nucleophile toward the aldehyde substrate. The conservation of aldolase activity in the peptide suggests that the peculiar pK of the key lysine residue is not significantly altered in the shortened protein sequence, indicating a conserved local environment. These data demonstrate the possibility of successfully obtaining aldolase peptides with enhanced efficiency and stereoselectivity from mutated libraries of GST-HSA100.

In summary, GST-HSA100 scaffolds can be used in both engineering and display screening technologies and have the potential to overcome the limitations of conventional recombinant antibodies as library templates.
Chapter 3

4.3 Towards a library of HSA100 based receptor

After having reduced human serum albumin binding site II from 300 to 100 amino acids, we have produced the fragment in bacterial system and we have validated it by biological and chemical assays. This small albumin fragment was proven to adopt the correct folding and to conserve the full binding activity for typical ligands. We have therefore attempted the development of a first generation of receptor libraries by site directed mutagenesis. We decided to change specific key positions for drug recognition and then to study possible improvements and/or changes in affinity for the two ligands.

4.3.1 Structural analysis

For this site-directed mutagenesis approach, some consideration based on HSA structure were made in order to identify the amino acids to be mutated. We developed a procedure that randomly mutagenizes specific aminoacids included in the peptide and directly involved in the interaction between protein and small molecules. The HSA100 can be divided into two hydrophobic subsites, one larger and one smaller, and a polar entry.

For this reason we considered the entry and contact sites with the ligand in the two hydrophobic pockets. The positions chosen for mutation are seven (figure 4.42):

- Alanine 215
- Leucine 238
- Leucine 260
- Isoleucine 264
- Arginine 257
- Histidine 242
- Serine 287

The seven mutating positions were recognized as non essential for the correct folding by bioinformatic analysis of sequence conservation within the set
of all Serum Albumins. Using the Uniprot database (The Uniprot Consortium, 2012) we selected an initial group of 93 non-repeated sequences of albumin proteins with chain sizes larger than 600 residues. Sequences were save in FASTA format and used in the plugin Multiseq from VMD (Roberts et al. 2006). The only truly conserved positions in the segment HSA100 are the positions of the cysteins (figure 4.41).
A list of suitable (not conserved) positions can be given (list below). Of this list, Ala 215 seems to be especially suitable for mutation, since it is also in contact with warfarin in the x-ray structure (PDB 1bke).

- Ala 201
- Ser 202
- Ala 215
- Glu 227
- Ala 229
- Ser 232
- Val 235
- Thr 243

The first pair of amino acids is represented by alanine 215 and leucine 238, that are placed inside the minor hydrophobic pocket. The second pair is made by leucine 260 and isoleucine 264, inside the major hydrophobic pocket, and the last 3 amino acids are histidine 242, arginine 257 and serine 287 in the polar site entry. Such amino acids are in close contact with warfarin in the X-ray structure of the warfarin-HSA complex (Curry et al. 1999) and we have decided to build up first small library by mutating two out of the seven amino acids, according to their topological correlation within the binding site. We have inspired by the works of Manfred Reetz on site directed mutagenesis of hydrolytic enzymes (Reetz and 2010; Reetz 2011), in which impressive affinity (and stereoselectivity) maturation were obtained in few steps from the first, slight improvement reached by mutating a first couple of amino acids.

Fig. 4.42 - Tridimensional structure of chosen subdomain of 100 amino acids with indicated positions for mutation by site-directed mutagenesis.
4.3.2 HSA100 site-directed mutagenesis

Site directed mutagenesis has been exploited to design HSA100 variants libraries placing separately the mutation in each site. We developed a molecular strategy for cloning the HSA100 amino acids domain by inserting the mutations in these seven positions. This strategy starts with oligonucleotides design on the basis of HSA100 wild-type optimized DNA template to introduce interesting mutations. The oligonucleotides were designed to insert separately the first two mutation at Ala 215 - Leu 238, the second pair of positions at Leu 260 - Ile 264 and the last three mutations at Hys 242, Arg 257 and Ser 287.

Up to date, we have created by PCR and molecular assembly the first two of the three libraries of HSA100 variants. Two PCR cycles and molecular assembly were performed and such libraries contain just \(2^2\) members. During the first PCR cycle, in order to generate the first library with mutation at Ala 215-Leu 238 positions inside the minor hydrophobic pocket, the Ala 215 position was mutagenized by inserting a NNK triplet into the reverse primer; this triplet encodes for any of the 20 possible amino acids and not for the stop codons. The first mutation into HSA100 peptide was introduced in this way, resulting in a PCR I product. During the second PCR (PCR II) amplification rounds, the second mutation in Leu 238 was introduced by using again in the forward primer the NNK triplet. The combined peptide library with Ala 215-Leu 238 mutations was generated by assembly PCR between PCR I and PCR II by using oligonucleotides HSA optimized external sense and anti (table 3.1) resulting in an amplified product of 351 bp. Polymerase cycling assembly (or PCA, also known as Assembly PCR) is a method for the assembly of large DNA from shorter fragments and allows the amplification of entire gene through an appropriate homology region in the PCR products (figure 4.43a).

The same protocol was used for the second library with mutation at Leu 260-Ile 264 positions inside the major hydrophobic pocket. The first PCR was performed by inserting the two mutations in the forward primer (PCR I) and in the second PCR it was amplified the remaining fragment (PCR II). Also in this case a region of homology between two PCR products (PCR I + PCR II=349 bp) has allowed the creation of a mutant library called HSA100 Leu 260 - Ile 264. The PCR strategy for the generation of the two mutants libraries is depicted in figure 4.43a.
Fig. 4.43a - Schematic representation of libraries construction by PCRI, PCR II and PCR assembly to generate libraries fragments mutated in specific positions.

Fig. 4.43b - PCR products of each libraries construction: lanes 1-2 PCR I of 130bp and its negative control, lanes 3-4 PCR II of 221bp and its negative control for mutant library Ala215-Leu238. Lanes 5-6 PCR I of 133bp and its negative control, lanes 7-8 PCR II of 216bp and its negative control for mutant library Leu260-Ile264; lane 10: molecular weight 100bp.

The PCR assembly products (figure 4.43b) contain proper 5' and 3' restriction sites BssHII and Nhel respectively; these fragment libraries were subsequently purified, digested and ligated with compatible digested vectors pGex 4T-1 (previously cut with the same restriction enzymes BssHII/Nhel) generating pGex-HSA100 Ala 215 - Leu 238 and pGex-HSA100 Leu 260 - Ile 264 plasmids in transformed E. coli DH5α competent cells.

According to the strategies just described, correct primers were designed to introduce codons that translate for random amino acids into the considered positions. In fact the presence of random amino acids in predicted positions was assessed performing a sequencing analysis and PCR with generic primers.
Results and discussion

(HSA optimized external sense/HSA optimized external anti, see table 3.1) on 25 clones randomly picked out of both the libraries.

Fig. 4.44 - Electrophoresis of PCR-amplified HSA100 random DNA of 25 randomly mutants clones of two libraries. Each clone expresses a correct fragment with HSA optimized external sense/HSA optimized external anti primers. 100bp DNA marker, ranging from 100 to 1000bp.

All the tested clones contained a fragment of the expected size (figure 4.44), about 335 bp: the amplified PCR product size was due to external position of the annealing primers employed for the screening. This confirms the integration of HSA100 mutants fragments into the pGex vector, and the presence into each bacteria of one plasmid. To check the actual diversity of the library, several clones were sequenced, verifying the randomization of selected residues. The results are reported in table 4.12 and confirm that effective randomization occured.

<table>
<thead>
<tr>
<th>Library pGex- HSA100 mutations</th>
<th>Ala 215 (wt)</th>
<th>Leu 238 (wt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone 1</td>
<td>Threonine T</td>
<td>Valine V</td>
</tr>
<tr>
<td>Clone 2</td>
<td>Leucine L</td>
<td>Lysine K</td>
</tr>
<tr>
<td>Clone 3</td>
<td>Asparagine N</td>
<td>Lysine K</td>
</tr>
<tr>
<td>Clone 4</td>
<td>Lysine K</td>
<td>Proline P</td>
</tr>
<tr>
<td>Clone 5</td>
<td>Lysine K</td>
<td>Valine V</td>
</tr>
<tr>
<td>Clone 6</td>
<td>Arginine R</td>
<td>Serine S</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Library pGex-HSA 100 mutations</th>
<th>Leu 260 (wt)</th>
<th>Ile 264 (wt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone 1</td>
<td>Leucine L</td>
<td>Valine V</td>
</tr>
<tr>
<td>Clone 2</td>
<td>Valine V</td>
<td>Valine V</td>
</tr>
<tr>
<td>Clone 3</td>
<td>Serine S</td>
<td>Leucine L</td>
</tr>
<tr>
<td>Clone 5</td>
<td>Proline P</td>
<td>Arginine R</td>
</tr>
<tr>
<td>Clone 15</td>
<td>Tryptophan W</td>
<td>Histidine H</td>
</tr>
<tr>
<td>Clone i</td>
<td>Arginine R</td>
<td>Glutamic acid E</td>
</tr>
</tbody>
</table>

Table 4.12 - Sequence analysis of a) six random clones of mutants library pGex-HSA100 Ala215-Leu238 and b) six clones of mutants library of pGex-HSA100 Leu260-Ile264. Tables report mutated amino acids in the selected positions.
4.3.3 Production and purification of HSA100 mutants

The expression of the recombinant proteins in *E. coli* was assessed first by a western blot assay. This test was performed by using picked random clones from the two HSA100 variant libraries. Protein expression was induced in bacteria with molecular inductor 0.2 mM IPTG (isopropyl β-D-1 thiogalactopyranoside) at 25°C over night. Starting from total bacterial extracts of each random clones, peptides expression, production and translocation were analyzed by SDS-PAGE gel, and western blot was carried out by using antibodies towards the GST tag (figure 4.45).

![Western blot assay to check the size, the expression and production levels of random selected bacteria HSA100. The SDS-PAGE was loaded with total extract fractions of DH5α induced for the expression of the recombinant proteins GST-HSA100. The arrows indicate the predicted molecular weights of the fusion proteins, (38kDa) of all clones except for clones 7, 4 and 16 which do not produce the protein. Primary antibody: mouse anti-GST 1:2000. Secondary antibody: goat anti-mouse AP conjugated 1:2000.](image)

Recombinant proteins were clearly detected in almost every clone, except for clone 7 in the first library and for clones 4 and 16 in the second library, that showed a lower expression level. Two bands were observed in several clones: the upper band is the correct fusion-protein product, showing the predicted molecular weight (about 38 kDa), whereas the lower band may be related to a GST alone (26 kDa).

After having obtained such evidence of correct bacterial expression in most clones, we have attempted a large scale production of the twenty clones that resulted fully expressed. This experiment was performed in order to check first the technical ability of obtaining useful amounts of the mutant proteins from the libraries, and to verify the effect of mutations on the yield.
No attempts were made at this stage to setup a selection strategy, and twenty clones were just randomly selected without trying to fish out the best binders from the small libraries.

The selection of soluble bacterial expression products requires in fact dedicated strategies, and we prefer to setup such methodologies after having the evidence that the libraries are effective in producing peptides.

The pGex-HSA100 mutant proteins were produced and purified with the glutathione purification system, starting from 150 ml of bacterial culture induced O/N with IPTG. Each recombinant protein was produced in the same way following the protocol reported in 3.6.23.

Purified proteins were quantified by Coomassie gel staining and their concentration were in the 0.2 mg/ml and 4 mg/ml range (figure 4.46).

![Image](image_url)

**Fig. 4.46** - A) This picture reportes the procedure of production-purification by affinity chromatography of four mutated clones: 1-2-3-5 of one of the two libraries. The same procedure was adopted for all analyzed mutant clones. I-II-III-IV-V= the first five elutions of each purification, Fl= flowthrought, R= resin, mw= molecular weight. B) 0.1 or 1 μl of each purified and dialized protein on 12% polyacrilamide gels.

### 4.3.4 Binding activity of the mutant clones

Despite the fact that no selection pressure was introduced in order to identify powerful binders among the twenty clones, we have nevertheless decided to perform binding tests on the peptides, in order to check, in this preliminary analysis, whether the choice of randomized sites would lead to changes in the affinity.

The tests were performed as described for HSA 100 wild type, by ELISA, SPR and fluorimetric techniques.
### 4.3.4.1 ELISA assay

The interaction between Efavirenz and Warfarin with random mutant clones were observed first by ELISA assay with the same protocol used for HSA100-GST wild type (see par. 4.2.5.1). The HSA100-GST recombinant mutants proteins were used at the same 1 μg/ml concentration (figure 4.47).

Both the sets of peptides show reactivity towards efavirenz and warfarin. Mutations in the minor pocket (Ala 215 – Leu 238) seem to induce major variations on efavirenz binding rather than on warfarin; all the best clones 3, 4 and 5 contain a lysine residue inside the pocket.

Mutations in the major pocket (Leu 260 – Ile 264) affect the ELISA outcome in a similar way on both warfarin and efavirenz. An arginine is found in clone 5, while tryptophane replaces leucine 260 and hystidine replaces isoleucine 264 in clone 15.
Rather surprisingly, proline is also found in two cases, namely in clone 4 of the minor pocket library and in clone 5 of the major pocket one.

### 4.3.4.2 SPR

A SPR-based screening was set up for efavirenz.

The drug was immobilized on a gold chip as described for the wild type HSA100 binding assay, and all the clones were tested by flowing 5 μM solutions of each peptide in the sensor cell.

Binding and dissociation are again observed in most clones. Clone 5, used as positive control (because is the best clones of library Leu 260 – Ile 264), gave the highest signal in the screening library, while the other clones of the minor pocket (figure 4.48a) seem to give lower signals than wild type HSA100. Conversely, most of the major pocket (figure 4.48b) mutants perform better than wild type HSA100, and clone 5 gives the highest signal, together with clone 1. Clone 5 was already seen in the ELISA test, as an arginine containing peptide, while clone 1 contains just a I264V mutation.

Fig. 4.48 - SPR sensorgrams for binding of HSA100 mutants clones of a) Ala215-Leu238 and b) Leu260-Ile264 libraries to efavirenz. Efavirenz was firstly coated and the HSA100 mutants proteins (5μM) were added.
4.3.4.3 Fluorescence assay

The binding activity was also confirmed by fluorescence quenching assay. This have been done in order to obtain quantitative data on binding. Both ELISA and SPR results should be regarded as qualitative experiments, since the signals might be affected by errors in the peptide concentrations that should be perfectly equal to allow a quantitative comparison. On the contrary, $K_d$ measurements carried out by fluorescence quenching, via titration of the peptide receptor with an excess ligand give results that are not affected by errors in the concentration of peptide, provided that the latter is kept well below the expected $K_d$ values.

The experiments were by the way carried out on the major pocket library, as the SPR data showed that most of its clones were more active than wild type. The titrations with efavirenz and warfarin are reported in figure 4.49 and the $K_d$ values obtained from Hill analysis are reported in table 4.13.

![Graph a](image1.png)

**HSA100<sub>260</sub>Leu<sub>264</sub>Ile mutants clones with Efavirenz**

![Graph b](image2.png)

**HSA100<sub>260</sub>Leu<sub>264</sub>Ile mutants clones with Warfarin**

Fig. 4.49 - HSA100 mutants small ligand binding. Titration of fluorescence emission of GST-HSA100 wt and of mutant clones of a) Ala215-Leu238 and b) Leu260-Ile264 libraries by warfarin and efavirenz.
Results and discussion

Table 4.13 - Dissociation constants calculated by fluorescence quenching.

<table>
<thead>
<tr>
<th>Clone</th>
<th>Warfarin (Kd)</th>
<th>Efavirenz (Kd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>91.8 μM</td>
<td>233.9 μM</td>
</tr>
<tr>
<td>1</td>
<td>88.3 μM</td>
<td>113.0 μM</td>
</tr>
<tr>
<td>2</td>
<td>83.2 μM</td>
<td>134.9 μM</td>
</tr>
<tr>
<td>3</td>
<td>85.9 μM</td>
<td>105.0 μM</td>
</tr>
<tr>
<td>5</td>
<td>83.6 μM</td>
<td>103.5 μM</td>
</tr>
<tr>
<td>15</td>
<td>72.1 μM</td>
<td>77.4 μM</td>
</tr>
<tr>
<td>i</td>
<td>58.5 μM</td>
<td>281.8 μM</td>
</tr>
</tbody>
</table>

Discussion

Bioinformatic analysis of the conserved positions in serum albumin, and structural analysis of the drug-protein interactions within subdomain II binding site allowed to identify suitable positions for site-directed mutagenesis.

To the best of our knowledge, mutagenesis on serum albumin was sever attempted neither on the whole protein, nor on its subdomain fragment.

The two small libraries were obtained, and diversity was effectively introduced in the decided positions of both the minor and the major hydrophobic pocket.

We were also able to obtain the production of the mutated clones at concentration of 4 mg/ml.

Preliminary binding tests carried out on a set of twenty unselected clones showed that binding was conserved in most peptide. Quantitative data were obtained on the major pocket clones (table 4.13): mutations at leucine 260 and isoleucine 264 seem to affect binding of efavirenz in a more signifivant way, while the affinity for warfarin is not affected. This might be consistent with a recognition of efavirenz occurring mostly in the major pocket of HSA. On the other side, warfarin interacts with both the hydrophobic pocket of albumin with two aromatic rings, and with the polar entry of the site with its ionizable 1,3 dicarboxyl moiety: thus interaction at multiple sites make less important the major pocket for binding.

Clone 15 recognizes efavirenz and warfarin with equal affinity, and the affinity for efavirenz is enhanced three times with respect to wild type HSA100: an extra tryptophan and hystidine are present in its binding site.
Conclusion

The detection of small molecules with classical biochemical systems is still a challenge. The identification and optimization of novel protein scaffolds for this purpose is an emerging field of research. Binding and catalysis are two important points of molecular recognition, which plays a key role for both in biological systems and enzymes. Peptide scaffolds that bind small molecules can be applied as innovative biosensors devices, for example in therapeutic drug monitoring, while libraries of peptide catalysts would be highly valuable tools in organic synthesis.

The aim of this project was to obtain new short peptides capable of both binding and catalysis.

We have been able to generate a protein scaffold based on ex novo E/K coiled-coil domains and to optimize its properties. This has been performed starting from a library of single chain E-Kinv peptide, designing a random mutagenized small site, and selecting the library byphage-display. The selected peptides appear to be stable and bind with nanomolar affinity caffeine and aldosterone, thus indicating that the library can be used as a source of peptide hosts to be used in biosensors field.

Moreover, Sudlow binding site of Human albumin was identify as suitably organized to serve a stand-alone domain, and reduced from 300 to 100 amino acids. We have demonstrated the structural and functional integrity of this isolated domain. In fact, peptide HSA100 is stable, folds correctly and retains binding and catalytic properties of albumin. We have carried out a preliminary experiments in order to explore the generation of mutated albumin fragments by site directed mutagenesis, and we have created two small HSA100 mutants libraries in key positions for drug recognition. Even without attempting any affinity selection strategy, we have found binders in a set of randomly selected members of the library, with improved affinity towards efavirenz. This results opens the way to the development, after setting up proper selection protocols, of HSA100-based sensing systems for antiretroviral drug monitoring.
References


References


References


References


References


Acknowledgments

We are grateful to:

- Commissariato del Governo nella Regione Friuli – Venezia Giulia (Fondo Trieste Grant 597/08 “piccoli peptidi per lo sviluppo di biosensori”) for having funded our research project and Regione Piemonte “piattaforma ImmOnc”;
- Prof. R. Marzari (Dept. of life sciences, Trieste) and prof. D. Sblattero (Dept. of medical science, Novara), my mentors;
- Prof. F. Berti for all supports, Dr. Silvia Pavan and Dr. Giampaolo Fontanive (Dept. of chemical and pharmaceutical sciences, Trieste);
- Prof. A. Tossi (Dept. of life sciences, Trieste)
- Prof A. Laio and Dr. Rolando Hong (SISSA, Trieste)
- Dr. Adriano Savoini, Biosensor Technologies srl (Trieste)
- My research group Dr. T. Gaiotto, Dr. S. Boscolo, Dr. S. Capolla and my biologists colleagues Dr. L. Marson and G Ceddia.
- My family and Diego…

I dedicate this work to the memory of Marta Licciulli, my friend and my teacher in the laboratory and in life, molecular biologist (1976-2012).